

DOCUMENTI ISTAT

L'utilizzo dei dati INPS per la stima trimestrale del numero dei dipendenti, le retribuzioni, il costo del lavoro e le ore lavorate

Problemi ed errori di tipo non-campionario

Rapporto intermedio all'Eurostat

*C. Baldi, E. Cimino, F. Rapiti, P. Minicucci, D. Tuzi,
(Servizio statistiche congiunturali su occupazione e sui redditi)*

*R. Succi
(Ufficio valutazione e controllo strategico)*

Abstract

The Italian Statistical Office is currently experimenting the use of Social Security data to produce short term and annual official economic statistics regarding labour market. All firms with employees have to fill in the monthly form termed DM10 to declare compulsory contributions to the Social Security Institute (INPS). The DM10 contains an ample range of information concerning the level and the structure of employment and wages. These data can be captured and processed to obtain a considerable number of statistical annual and short term indicators covering a population of firms of all size classes, without increasing the burden on firms with new surveys and at a relatively low cost.

Relying on INPS data to produce official statistic information is not an easy task. Among the quality problems that must be evaluated and taken into account the main are the following: 1) the administrative unit (*posizione contributiva*) is not a statistical one; 2) there are some differences between the concept and the definition underlying the administrative variable and the statistical concept that should be captured (e.g. the one described in EUROSTAT regulations); 3) the units are classified in economic activity sectors for contribution purposes and, hence, can be misclassified according the statistical scope; 4) there can be differences between the population of firms that INPS data cover and the coverage required; 5) for short term statistics only a non random sample of units is timely available to produce indicators within a reasonable time-lag from the reference period.

In this paper we report the state of advancement of a project aimed at the production of the quarterly labour cost index in Italy (LCI). The 5 issues before mentioned are addressed and most of the problems that they can imply for the LCI are evaluated along with the proposal of some solutions.

Sintesi

Presso l'Istituto Nazionale di Statistica è attualmente in corso di sperimentazione l'uso di dati amministrativi di origine previdenziale al fine di produrre statistiche ufficiali, congiunturali ed annuali, inerenti il mercato del lavoro. Tutte le imprese con occupati sono mensilmente tenute a presentare all'Istituto Nazionale di Previdenza Sociale (INPS) un modulo denominato DM10 per la dichiarazione dei contributi obbligatori dovuti. Il DM10 contiene un ampio insieme di informazioni riguardanti il livello e la struttura di occupazione e retribuzioni e i dati che se ne possono ricavare, adeguatamente trattati, consentono di ottenere un considerevole numero di indicatori statistici su una popolazione di imprese in grado di coprire tutte le classi dimensionali. In tal senso, le imprese verrebbero alleggerite del carico statistico, consentendo anche di limitare l'onere finanziario conseguente l'alternativo ricorso a nuove indagini.

Il ricorso ai dati INPS per produrre informazioni statistiche ufficiali, tuttavia, non è privo di complessità. Tra i problemi di qualità che devono essere valutati e presi in considerazione, è necessario evidenziare i seguenti aspetti: 1) l'unità amministrativa (*posizione contributiva*) non ha natura statistica; 2) vi sono alcune differenze di fondo sottostanti concetto e definizione della variabile amministrativa e della variabile statistica da analizzare (ad esempio quella descritta nei regolamenti EUROSTAT); 3) le unità di origine sono classificate in settori di attività economica con scopi contributivi e, dunque, possono risultare non correttamente classificate secondo gli scopi statistici; 4) possono esserci differenze tra la popolazione di imprese coperta dai dati INPS e quella richiesta; 5) per le statistiche congiunturali è disponibile solo un campione non casuale di unità per poter produrre indicatori con un ragionevole ritardo temporale dal periodo di riferimento.

In questo documento si descrive lo stato di avanzamento di un progetto finalizzato all'uso dei dati INPS per produrre un indice trimestrale del costo del lavoro in Italia (LCI). I vari errori e problemi di carattere non campionario citati in precedenza vengono approfonditamente analizzati e per alcune problematiche vengono proposte soluzioni operative.

INDICE

Sintesi	
1 La realizzazione dell'indice europeo di costo del lavoro in Italia attraverso i dati INPS: obiettivi e problemi metodologici	
1.1 Premessa	5
1.2 L'uso dei dati amministrativi nella statistica ufficiale e i recenti orientamenti comunitari	6
1.3 Elementi di forza e debolezza dell'uso dei dati INPS come fonte statistica	7
1.4 Una prima ipotesi di struttura della rilevazione trimestrale sui dati INPS	9
1.5 I problemi metodologici e di qualità dei dati	13
1.6 I problemi di copertura della fonte INPS	14
1.7 Il contenuto delle variabili presenti sul DM10: una prima riflessione	15
2 Identificazione dei codici presenti sul dm10 INPS e la ricostruzione delle variabili obiettivo	17
2.1 Alcune note sulla contribuzione INPS	17
2.3 La riorganizzazione dei codici	20
2.3.1 Individuazione dei codici occupazione	21
2.3.2 Analisi degli altri codici presenti nei quadri B-C	21
2.3.3 Analisi dei codici presenti sul quadro D	22
2.3.4 Ricostruzione delle aliquote contributive	22
2.4 Esempio puntuale di calcolo del costo del lavoro su base DM 10/2	23
3 La qualità dei dati INPS: un confronto con l'indagine Istat LPS per l'anno 1995	27
3.1 Introduzione	27
3.2 Specifiche e limiti dell'analisi	27
3.3 Occupati dipendenti	30
3.3.1 Definizioni a confronto	30
3.3.2 Una valutazione della distorsione dell'occupazione dipendente	30
3.4 Retribuzione lorda	35
3.4.1 Definizione a confronto	35
3.4.2 Una valutazione della distorsione della retribuzione lorda	36
3.5 Costo del lavoro	39
3.5.1 Definizioni a confronto	39
3.5.2 Un esercizio di calcolo del costo del lavoro INPS	40
3.5.3 Una valutazione della distorsione del costo del lavoro	42
3.6 Conclusioni	43
4 Alcune valutazioni sulla qualità della codifica dell'attività economica negli archivi amministrativi	45
4.1 La codifica del settore di attività economica negli archivi INPS	45
4.2 Alcune considerazioni generali	46
4.3 Analisi nell'universo DM10 di dicembre 1996	47
4.4 Analisi nel campione DM10 di gennaio 1999	48
4.5 Conclusioni	49
5 La ricostruzione delle imprese a partire dalle posizioni contributive degli archivi INPS	51
5.1 Introduzione	51
5.2 La situazione informativa	51
5.3 Dalla posizione contributiva all'impresa: le chiavi di aggancio	53
5.4 La procedura di ricostruzione delle imprese	55
5.4.1 Una descrizione generale della metodologia	55
5.5 Un esercizio di ricostruzione dell'impresa al mese di dicembre 1996	57
5.6 Alcuni risultati	58
5.7 Validazione dei risultati	60
5.7.1 Confronto delle imprese ricostruite da fonte INPS con le imprese dell'archivio ASIA	61
5.8 Conclusioni	66
6 La stima del tempo di lavoro delle dichiarazioni contributive INPS. Alcune ipotesi preliminari e problemi aperti	71
6.1 Introduzione	71
6.2 Alcune linee per una stima delle ore sulla base dei dati INPS	72
6.3 Il calcolo delle ore retribuite	72
6.3.1 L'informazione del DM10	72
6.3.2 L'uso delle informazioni contrattuali	73
6.4 Il calcolo delle Ore Effettivamente Lavorate	74
6.5 Il problema dell'underreporting delle giornate retribuite	76
6.6 Conclusioni	77

7	Le caratteristiche strutturali dello “pseudo-campione” e dell’universo dei DM10 INPS.....	83
7.1	Introduzione.....	83
7.2	Gli arrivi dei DM10 su supporto magnetico	84
7.3	Il trattamento preliminare dei dati.....	85
7.3.1	Campioni.....	85
7.3.2	Attribuzione del codice di attività economica e misclassificazione	85
7.3.3	L’Universo INPS	86
7.4	Confronti tra le strutture.....	87
7.5	La permanenza delle posizioni nei campioni	90
7.6	Conclusioni	91
8	Conclusioni provvisorie: le principali caratteristiche della rilevazione e i problemi aperti	93
8.1	Le difficoltà nella progettazione del disegno della rilevazione	93
8.2	Sintesi delle principali caratteristiche della rilevazione	94
	Riferimenti bibliografici	97

SINTESI*

Il presente documento riprende integralmente il Rapporto intermedio inviato all'Eurostat (nel quadro del contratto Labour Cost Index) che illustra lo stato di avanzamento dei lavori di progettazione e realizzazione dell'indice del costo del lavoro (*Labour cost index* - LCI) in Italia sulla base dei dati amministrativi INPS raccolti trimestralmente. Il Rapporto documenta anche come tali dati verranno utilizzati per la stima delle variabili relative al numero di occupati, retribuzioni e ore lavorate di tutti i moduli del Regolamento del Consiglio relativo alle statistiche congiunturali sulle imprese (R. 1165/98).

Il rapporto è composto di otto capitoli:

- Nel primo capitolo vengono presentati gli obiettivi del progetto. Si evidenziano, inoltre, i punti di forza e debolezza dei dati amministrativi INPS nonché i problemi metodologici (in particolare, quelli riguardanti gli errori non campionari) che si incontrano nella realizzazione di una rilevazione basata su tali dati.
- Nel secondo capitolo viene illustrata in dettaglio la principale fonte d'informazione su cui si basa la rilevazione: le dichiarazioni contributive mensili (i moduli DM10). Vengono descritti, inoltre, i criteri seguiti per la riorganizzazione dei codici, riportati sul modello, in partizioni funzionali alla ricostruzione ed al calcolo delle variabili occupazione dipendente e retribuzione lorda da un lato ed oneri contributivi dall'altro; a tal fine di particolare importanza risulta la distinzione tra codici "tipo occupato" e codici tipo contribuzione".
- Nel terzo capitolo attraverso un microdata matching si passa a valutare la qualità dei dati INPS con riferimento al problema del contenuto delle variabili amministrative ed il loro rapporto con le variabili statistiche. Le variabili ricavabili dal modulo DM10 dell'INPS (occupazione, retribuzione lorda, costo del lavoro) vengono considerate come una possibile misurazione di ignote variabili vere. Per approssimare queste ultime si utilizza una fonte interna Istat. Vengono presentate analisi univariate e multivariate per misurare la grandezza della distorsione e per valutarne la natura sistematica o casuale. I risultati ottenuti evidenziano come la maggior parte dei problemi legati al contenuto delle variabili INPS, risulta nella pratica di entità limitata.
- Nel quarto capitolo si valuta la qualità delle classificazioni delle attività economiche presenti negli archivi INPS. In particolare, si studiano le caratteristiche del CSC (codice statistico contributivo) e del codice ATECO91-INPS recentemente inserito direttamente dall'Istituto di previdenza nei suoi archivi amministrativi in confronto con l'ATECO91 presente nell'archivio ASIA considerato come benchmark. Dall'analisi emerge con chiarezza che l'ATECO91-INPS attribuito direttamente dall'Istituto di previdenza, che corrisponde, con un certo grado di sistematicità, al CSC, è una variabile affetta complessivamente da pochi errori. La conclusione a cui si perviene consiste nella possibilità di fare un uso diretto dell'ATECO91-INPS nella

*Il presente Rapporto è stato realizzato nell'ambito del progetto interarea Istat "Utilizzo dei dati INPS per la realizzazione di statistiche correnti su occupazione e retribuzioni". Il Rapporto, frutto di un proficuo lavoro collettivo di tutti i partecipanti al progetto, è stato coordinato da Fabio Rapiti. I singoli capitoli sono stati redatti da:

cap. 1 Fabio Rapiti;

cap. 2 Pierluigi Minicucci;

cap. 3 par. 3.1, 3.2, 3.3, 3.4, 3.6 Eleonora Cimino, par. 3.5 Pierluigi Minicucci ed Eleonora Cimino;

cap. 4 par. 4.2, 4.3, 4.4, 4.5 Donatella Tuzi, par. 4.1 Fabio Rapiti;

cap. 5 par. 5.1, 5.2, 5.4 Raffaella Succi, 5.5, 5.6, 5.7 e 5.8 Donatella Tuzi, par. 5.3 Eleonora Cimino;

cap. 6 par. 6.2, 6.3, 6.4, 6.5 Ciro Baldi, par. 6.1, 6.6 Fabio Rapiti;

cap. 7 Ciro Baldi;

cap. 8 Fabio Rapiti.

Si ringrazia il personale dell'INPS per la preziosa e continua opera di collaborazione.

costruzione degli indici piuttosto che utilizzare la trascodifica del CSC in ATEC91.

- Nel quinto capitolo si descrive la procedura di ricostruzione dell'impresa come unità di analisi, a partire dalle informazioni sulle "posizioni contributive" (l'unità elementare di rilevazione dei dati INPS non sempre coincidente con la definizione di impresa) e quelle contenute nell'archivio anagrafico delle dichiarazioni contributive dell'INPS. La procedura si basa sull'individuazione di alcune chiavi identificative che consentono di raccordare le singole posizioni contributive alle relative imprese e prevede l'attribuzione, per ogni impresa ricostruita, di variabili anagrafiche e strutturali. La qualità dei risultati viene valutata confrontando le variabili chiave dell'impresa ricostruita con le imprese dell'archivio ASIA.
- Nel sesto capitolo si discute dei problemi di stima delle ore effettivamente lavorate a partire dai dati del DM10. Tali dati non sono sufficienti ed hanno problemi di qualità ai fini della stima delle ore effettivamente lavorate. Viene, quindi, proposto un sistema di fonti che correggano ed integrino tali dati e viene presentato un metodo di stima delle ore retribuite, basato sul supporto delle informazioni di tipo contrattuale. L'analisi di un tale metodo ha però bisogno di ulteriori approfondimenti soprattutto per verificarne la fattibilità date le fonti di dati disponibili.
- Nel settimo capitolo si descrivono le caratteristiche strutturali dello pseudo-campione di DM10 disponibile entro tre mesi dal periodo di riferimento e si effettua una prima valutazione della sua idoneità a rappresentare l'universo dei dati. Tale operazione è la premessa per la messa a punto della metodologia di stima congiunturale che viene presentata nel Rapporto Finale.
- Nel capitolo ottavo alla luce dei problemi e delle soluzioni proposte nei capitoli precedenti si presentano in sintesi le principali caratteristiche che avrà la rilevazione trimestrale basata sui dati DM10 per la realizzazione del LCI.

1 LA REALIZZAZIONE DELL'INDICE EUROPEO DEL COSTO DEL LAVORO IN ITALIA ATTRAVERSO I DATI INPS: OBIETTIVI E PROBLEMI METODOLOGICI

1.1 Premessa

Il presente rapporto intermedio illustra lo stato di avanzamento dei lavori della progettazione e realizzazione dell'indice di costo del lavoro in Italia sulla base dei dati amministrativi INPS. Esso illustra anche come tali dati verranno utilizzati per stimare le variabili relative al numero di occupati, retribuzioni e ore lavorate di tutti i moduli del Regolamento del Consiglio relativo alle statistiche congiunturali sulle imprese (R. 1165/98) e Short Term Statistics

L'indice trimestrale europeo del costo del lavoro (*Labour cost index* - LCI) dovrebbe rappresentare nel quadro del Sistema Statistico Europeo (SSE) la principale informazione per il monitoraggio congiunturale delle retribuzioni e del costo del lavoro. Le caratteristiche metodologiche del LCI concordate in sede europea sono le seguenti. L'indice dovrebbe includere tutte le componenti misurabili congiunturalmente del costo del lavoro: 1. Salari e Stipendi Lordi; 2. Contributi per la previdenza sociale a carico del datore di lavoro. Nel costo non vengono inclusi i costi per la formazione, i costi di assunzione e le spese per la mensa. L'unità di lavoro preferibile si riferisce alle Ore lavorate (incluse quelle di straordinario) ma in mancanza di tale dato diversi paesi utilizzano il numero di occupati equivalenti a tempo pieno. Nell'occupazione sono compresi tutti i dipendenti inclusi: i lavoratori part time o su base occasionale, gli apprendisti, chi è assunto durante il periodo di riferimento, chi smette di lavorare durante il periodo di riferimento. I lavoratori forniti da imprese di intermediazione vanno inclusi nella sezione K della classificazione NACE rev.1

La base dell'indice è 1996=100. La periodicità deve essere trimestrale mentre le informazioni vanno trasmesse all'EUROSTAT entro tre mesi dalla fine del trimestre di riferimento.

L'indice dovrebbe essere rappresentativo delle imprese di tutte le dimensioni appartenenti alle seguenti sezioni della classificazione NACE rev. 1 C,D,E,F;G;H;I;J;K. Gli indici dovrebbero essere prodotti secondo il seguente dettaglio: C+D+E, D, F, G+H+I, J+K, TOTAL.

Dal punto di vista metodologico l'importanza di giungere in breve tempo ad indicatori confrontabili a livello comunitario ha indotto il "Gruppo di lavoro sulle statistiche sulle retribuzioni e il costo del lavoro" e l'Eurostat a preferire un approccio graduale e di "armonizzazione dell'output e non dell'input": i singoli paesi producono gli indici a partire dalle loro fonti statistiche e amministrative nazionali e soltanto ex post e in itinere avviene una "relativa" armonizzazione.

In Italia attualmente la produzione di indici sul costo medio del lavoro è limitata alle imprese con oltre 500 addetti dell'industria in senso stretto e dei servizi sulla base della rilevazione mensile sull'occupazione, le ore lavorate, le retribuzioni delle grandi imprese. Gli elevati costi e gli enormi problemi che avrebbe posto la rilevazione congiunturale diretta del costo del lavoro presso le imprese con meno di 500 addetti ha indotto l'Istat a progettare e sperimentare una rilevazione basata sui dati provenienti dagli archivi dell'Istituto Nazionale di Previdenza Sociale (INPS). I dati derivano dalle dichiarazioni contributive mensili, cioè i moduli DM10, utilizzati dalle imprese per pagare i contributi sociali relativi ai loro lavoratori dipendenti. L'elemento centrale dell'utilizzo statistico dei DM10 è la possibilità di raccogliere l'intera popolazione di riferimento delle imprese con almeno un dipendente. Queste informazioni raccolte dall'INPS e trasferite all'Istat vengono sottoposte a complessi processi di linkage, normalizzazione, controllo, correzione, stima e integrazione con altre fonti statistiche.

L'avvio a regime della rilevazione basata sui dati INPS fornirà risultati a partire dal primo trimestre 2001. Poiché dalla primavera del 1999 è iniziata la pubblicazione del LCI da parte dell'Eurostat attraverso comunicati stampa per quanto riguarda gli indici italiani si è concordato di fornire fino al primo trimestre del 2001 indici costruiti utilizzando le informazioni presenti nella rilevazione sulle

grandi imprese (Istat 1999). Quando i dati provenienti dalla nuova rilevazione saranno disponibili sostituiranno l'intera serie a partire dal 1996.

Il Regolamento relativo alle statistiche congiunturali (R. 1165/98) è stato approvato nel maggio del 1998 ma i tempi di messa a regime sono piuttosto lunghi. E' previsto, infatti, che tutti i paesi possano avere su singole variabili delle deroghe per un periodo massimo di 5 anni, quindi, fino alla primavera del 2003. Questo è il caso delle variabili "numero di persone occupate", "ore di lavoro" e retribuzioni lorde". Inoltre, sul "numero di persone occupate" il Regolamento prevede che tale variabile possa essere approssimata (per un ulteriore periodo di 5 anni) con il numero di dipendenti. Attualmente l'Istat utilizza i dati della rilevazione sulle grandi imprese per soddisfare alcune esigenze del Regolamento congiunturale ad eccezione delle variabili e dei settori su cui è stata richiesta la deroga: tutto il modulo B (variabili B210, B220, B230); alcuni settori del modulo A e del modulo C.

Per le particolari difficoltà di rilevazione l'Istat intende avvalersi della possibilità di approssimare la variabile numero di occupati (A210, B210, C210, D210) con il numero di dipendenti totale della rilevazione (variabile A211, B211, C211, D211).

Analogamente a quello che avverrà per il LCI anche per le variabili del Regolamento congiunturale quando i dati provenienti dalla nuova rilevazione basata sulla fonte INPS saranno disponibili sostituiranno l'intera serie presumibilmente a partire dal 1996.

1.2 L'uso dei dati amministrativi nella statistica ufficiale e i recenti orientamenti comunitari

In Italia numerose istituzioni raccolgono dati come basi informative delle rispettive procedure amministrative e quindi per fini operativi e non statistici. L'utilizzo di questi dati a fini statistici è iniziato soltanto di recente. Ancora oggi numerosi problemi di natura amministrativa, legislativa, e tecnologica limitano l'estensione dell'uso di queste fonti. Anche per queste ragioni gli specifici problemi di ordine metodologico che riguardano l'uso di tali dati sono ancora in discussione.

A parte i problemi suddetti per molti anni l'uso dei dati amministrativi da parte della statistica ufficiale è stato particolarmente limitato per i forti dubbi che si avevano sulla qualità degli archivi amministrativi di riferimento. Su questo atteggiamento ha pesato una sorta di pregiudizio a favore di strumenti d'indagine tradizionali in cui lo statistico ha il controllo, o almeno presume di averlo, su tutte le fasi delle rilevazioni: dalla progettazione del questionario e definizione dei concetti, delle variabili da rilevare, delle classificazioni fino al trattamento finale dei dati.

Anche l'Eurostat non sempre ha favorito la diffusione di fonti amministrative nella produzione di statistiche e ha preferito la rilevazione diretta con strumenti tradizionali. Questo perché l'uso di fonti amministrative, fortemente caratterizzate a livello amministrativo nazionale, avrebbe condotto ad input piuttosto eterogenei e non avrebbe consentito una facile e adeguata armonizzazione e comparabilità dei risultati.

Negli anni più recenti il tema dell'uso dei dati amministrativi si è imposto sull'agenda di tutti gli Istituti di statistica europei e dell'Eurostat. La ragione di ciò può essere sintetizzato nei seguenti punti:

1. I continui avanzamenti nelle tecnologie dell'informazione hanno condotto ad una più facile gestione e manipolazione di grandi masse di dati e hanno aperto nuove possibilità all'abbinamento di diversi archivi e basi dati amministrative e statistiche;
2. L'Eurostat e tutti Istituti di statistica nazionali stanno cercando di soddisfare sempre di più la crescente domanda di informazione statistica di tipo economico sulle imprese e allo stesso tempo tentano di ridurre l'onere statistico sulle imprese;
3. Una proporzione sempre più rilevante di occupazione e reddito è attribuibile a piccole e medie imprese per le quali le indagini statistiche campionarie non risultano essere né facili da condurre né sempre particolarmente efficienti per ottenere risultati ad un certo livello di disaggregazione; questo problema in Italia è particolarmente evidente.

Il nuovo orientamento comunitario emerge chiaramente anche nei nuovi regolamenti che stabiliscono i concetti, le definizioni, le classificazioni e i criteri di raccolta e di trattamento dei dati sulle imprese. Se in passato si preferiva orientarsi verso una "armonizzazione dell'input", oggi si è scelto di andare verso una "armonizzazione dell'output". Vi è, quindi, maggiore flessibilità nelle procedure di raccolta dei dati e anche un allargamento della possibilità di stima delle variabili richieste. Nei nuovi regolamenti sulle statistiche strutturali e congiunturali sulle imprese ciò è chiarito in modo inequivocabile. Nell'articolo 4 del regolamento del Consiglio relativo alle statistiche congiunturali si legge:

“Gli stati membri possono acquisire i dati necessari utilizzando una combinazione delle diverse fonti di seguito specificate, applicando il principio della semplificazione amministrativa: a) indagini obbligatorie (...); b) altre fonti appropriate, compresi i dati amministrativi; c) procedure appropriate di stima statistica” (Eurostat, 1998).

Il regolamento, nell'articolo 10 specifica, inoltre, alcuni requisiti in termini di qualità che le statistiche prodotte dagli stati membri devono avere per garantirne l'affidabilità e la comparabilità internazionale. Al comma 4 di tale articolo si aggiunge però che:

“La valutazione della qualità deve essere effettuata confrontando i benefici derivanti dalla disponibilità dei dati con il costo della raccolta e l'onere per le imprese, specialmente quelle di minori dimensioni” (Eurostat 1998).

In definitiva questi recenti orientamenti dell'UE consentono a tutti gli Istituti Nazionali di Statistica una maggiore flessibilità rispetto al passato in termini di combinazione di fonti nella raccolta e nel trattamento dei dati.

In Italia il vero salto di qualità nell'uso dei dati amministrativi è stato fatto molto recentemente da quando è stato realizzato l'archivio ASIA (archivio statistico delle imprese attive). La costruzione di ASIA implica, infatti, l'utilizzo di sette fonti principali di cui soltanto una è una indagine statistica Istat (SIRIO-NAI). Le altre sei fonti sono di natura amministrativa: il registro delle Camere di Commercio, Industria, Artigianato e Agricoltura (CCIAA), il registro IVA del Ministero delle Finanze, l'archivio INAIL (Istituto Nazionale per la prevenzione degli Infortuni sul Lavoro), l'archivio dell'ENEL, il registro SEAT e anche l'archivio delle imprese con lavoratori dipendenti dell'INPS (Istituto Nazionale di Previdenza Sociale).

I buoni risultati della realizzazione dell'archivio ASIA hanno indotto l'Istat a fare ulteriori passi avanti nell'uso dei dati amministrativi per le statistiche sulle imprese e sul lavoro. D'altronde la statistica ufficiale in Italia di fronte alla duplice necessità di, incrementare il flusso di offerta di informazioni e contemporaneamente di ridurre la pressione statistica sulle imprese, si trova, in sostanza di fronte a scelte obbligate. In pratica per diversi fenomeni oggi non vi sono alternative all'uso dei dati amministrativi. Certamente l'utilizzo di tali fonti deve essere effettuato *cum grano salis*. È corretto parlare di integrazione e complementarietà fra fonti amministrative e fonti d'indagine piuttosto che di sostituzione di dati. È però opportuno specificare anche i compiti che si attribuiscono ad ogni fonte: la fonte amministrativa consente di avere la massa dei dati cioè l'universo della popolazione di riferimento, mentre l'indagine diretta, su piccoli campioni della popolazione, permette di trattare in modo opportuno i dati amministrativi in modo da garantirne una elevata qualità in termini di accuratezza e confrontabilità. In pratica la complementarietà e integrazione di cui si parla, è fortemente asimmetrica, con le indagini statistiche che servono sostanzialmente da supporto qualitativo alla fonte amministrativa.

1.3 Elementi di forza e debolezza dell'uso dei dati INPS come fonte statistica

I vantaggi dell'uso dei dati INPS per la produzione di statistiche sulle retribuzioni e il costo del lavoro sono molti:

1. costi di raccolta dei dati estremamente bassi per l'Istat (soltanto costi di accesso ai dati);
2. nessun ulteriore onere statistico sulle imprese (sono previste infatti eventualmente soltanto delle

piccole indagini supplementari);

3. gli archivi INPS includono (a parte poche eccezioni) tutte le imprese che impiegano lavoro dipendente nel settore privato dell'economia. Questo significa che tutte le piccole imprese (più di un milione) possono essere coperte ed è possibile realizzare statistiche per piccoli gruppi, come piccole aree o settori di attività economiche a 4-5 cifre senza problemi di precisione delle stime.

In ogni caso la produzione di statistiche sulla base dei dati INPS implica anche svantaggi notevoli:

1. problemi di natura tecnologica; costi di elaborazione elevati per l'enorme quantitativo di record;
2. problemi di natura amministrativa in particolare la consistenza e la continuità delle informazioni nel corso del tempo;
3. i problemi di qualità tipici dei dati amministrativi.

Con lo sviluppo di tecnologie dell'informazione migliori e più economiche il primo problema può essere facilmente risolto attraverso una opportuna pianificazione delle attività e delle procedure necessarie al trasferimento e trattamento dei dati. Il secondo problema è invece potenzialmente molto pericoloso. Prima di impegnarsi ad iniziare la complessa operazione di estrazione, trasmissione ed elaborazione dei dati amministrativi INPS è molto importante avere la garanzia che l'amministrazione, le procedure, e la modulistica che sono utilizzate per raccogliere i dati siano stabili. Ogni cambiamento nelle procedure amministrative e nella modulistica deve essere accuratamente monitorato altrimenti si mette in pericolo l'intero sistema di rilevazione. Da questo punto di vista le relazioni fra Istat e INPS sono eccellenti. Negli ultimi dieci anni tra le due istituzioni c'è stato un continuo sviluppo della cooperazione nel campo statistico nel quadro del SISTAN (Sistema Statistico Nazionale). Recentemente (29.09.98) è stata firmata una convenzione bilaterale che facilita e semplifica le relazioni amministrative e l'accesso ai dati. Inoltre nella recente legge per il censimento intermedio (DL n. 681/1996) l'Istat viene esplicitamente autorizzato a collaborare alle modifiche della modulistica di tutte le istituzioni che rientrano nel SISTAN, nel quadro della semplificazione amministrativa e ai connessi cambiamenti nelle procedure.

1.4 Una prima ipotesi di struttura della rilevazione trimestrale sui dati INPS

Come è stato accennato in precedenza attualmente l'Istat raccoglie informazioni mensili sull'occupazione, le retribuzioni, le ore lavorate e costo del lavoro limitatamente alle imprese con oltre 500 addetti dell'industria in senso stretto e dei servizi sulla base della rilevazione mensile sull'occupazione, le ore lavorate, le retribuzioni delle grandi imprese. La rilevazione basata sui dati INPS non implicherebbe l'eliminazione di tale indagine ma soltanto il complemento per la copertura delle imprese sotto i 500 addetti.

La maggiore difficoltà metodologica nella progettazione della rilevazione riguarda l'utilizzo dei dati amministrativi a scopi congiunturali sulla base non della popolazione di riferimento ma di partizioni particolari della popolazione. Questo pone complessi problemi di tipo tecnico-amministrativo in relazione con il fornitore di dati (INPS) e complessi e, relativamente nuovi, aspetti metodologici di tipo non campionario e problemi di tipo campionario. In particolare per quanto riguarda la struttura statistica della rilevazione congiunturale l'obiettivo è quello di giungere ad una metodologia che consenta di ottenere le stime migliori con i costi più bassi tenendo conto degli sviluppi e del possibile aggiornamento dell'archivio ASIA e dei vincoli tecnologici INPS e Istat.

La progettazione di una rilevazione congiunturale è stata resa possibile da quando l'INPS ha accentuato l'informatizzazione del sistema di compilazione e trasmissione dei DM10. Le imprese con dipendenti ogni mese devono presentare il modulo DM10 entro il 20 del mese successivo a quello in cui si riferisce la denuncia presso le competenti sedi INPS territoriali. I moduli pervengono in varie forme (cartacea, floppy-disk o altro supporto magnetico, telematica) alle sedi dell'INPS dove vengono inseriti in un archivio informatizzato locale e poi vengono scaricati in modo

scaglionato nel tempo presso l'archivio centrale. Attualmente circa un decimo dei moduli viene inviato in forma telematica o su supporto magnetico. Ciò significa che circa un milione di moduli DM10 devono essere digitati o passati al lettore ottico. Nell'archivio centrale pervengono, quindi, molto rapidamente i dati che riguardano i DM10 inviati telematicamente o disponibili su supporto magnetico, più lentamente invece arrivano quelli inviati su carta. Poichè il ritardo di produzione degli indicatori non può essere superiore a 90 giorni l'Istat ha suggerito all'INPS di realizzare un sistema che consenta di prelevare i dati mensili che pervengono su supporto magnetico immediatamente appena si rendono disponibili presso le sedi periferiche dell'INPS, prima cioè che essi passino attraverso le procedure di controllo e correzione effettuate presso queste ultime. I dati a cui si fa riferimento sono circa il 15% del totale e risultano immediatamente disponibili presso le sedi locali con un ritardo non superiore ai 45 giorni.

La metodologia statistica che si ipotizza di impiegare consiste in una stima aggiornata al tempo corrente (o stima anticipata) utilizzando sia l'informazione disponibile negli archivi di fonte INPS (un archivio anagrafico e uno pseudo campione di DM10 al tempo corrente, un universo di DM10 al tempo t-12), sia di fonte Istat e in particolare l'archivio ASIA. Le stime verrebbero calcolate sulla base di un modello statistico di tipo regressivo che lega le variabili d'interesse (gli occupati, la retribuzione e il costo del lavoro) a variabili ausiliarie, di cui si può conoscere la distribuzione nell'universo delle imprese dagli archivi INPS. Per la stima del mese t si utilizza un "presudo-campione" non rappresentativo di dichiarazioni contributive DM10 che è costituito dall'insieme dei modelli che sono compilati dalle imprese su supporti magnetici (circa 160.000) riferito al mese t e l'universo dei DM10 INPS, disponibile sempre al tempo t ma riferito a t-12. In particolare le fonti necessarie alla stima sono:

1. una consistente partizione rilevata mensilmente della popolazione di riferimento (attualmente circa 160.000 DM10) che, ovviamente, non è un campione selezionato casualmente, ma corrisponde al sottoinsieme dei DM10 che vengono trasmessi dalle imprese all'INPS attraverso un supporto informatico e riportano informazioni relative al mese precedente rispetto a quello corrente;
2. l'universo dei modelli DM10 con informazioni che presentano un ritardo temporale di circa 11-12 mesi rispetto a quello corrente;
3. l'archivio anagrafico delle posizioni contributive dell'INPS ovvero l'universo delle informazioni anagrafiche delle imprese aggiornato trimestralmente, con tutte le posizioni contributive formalmente aperte all'INPS.
4. l'archivio statistico ASIA che contiene l'universo delle imprese.

Per controllare più direttamente tutto il processo di produzione delle informazioni statistiche si preferisce utilizzare soltanto dati riferiti ai DM10 grezzi (ma abbinati dall'INPS con alcune informazioni di base come i codici fiscali, ecc. assenti nei DM10) piuttosto che dati già elaborati dall'Istituto di previdenza. Nella prima fase la trasmissione dei file da parte dell'INPS viene realizzata attraverso supporto magnetici (cassette); in futuro gran parte delle trasmissioni dovrebbe essere effettuata tramite un collegamento telematico diretto fra i due istituti.

I dati trasmessi dall'INPS devono essere sottoposti ad un complesso trattamento di cui si riporta schematicamente una sintesi:

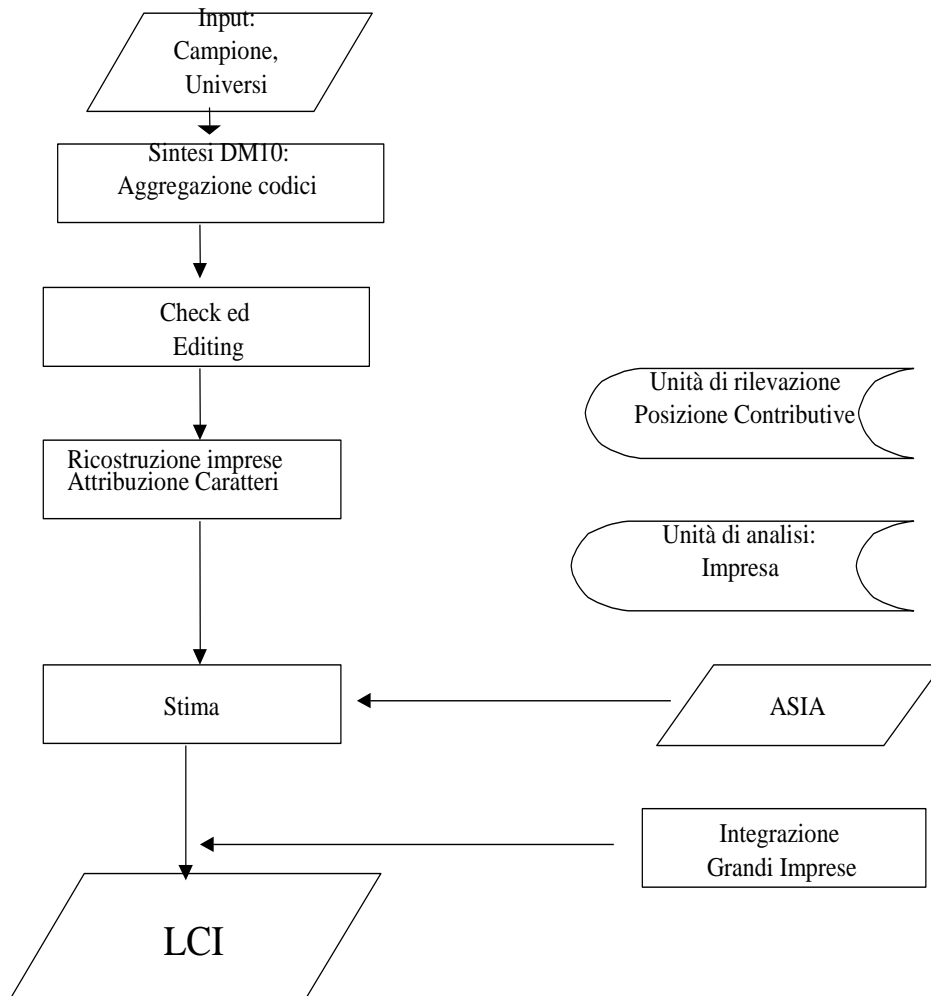
1. trascodifica ed aggregazione delle variabili elementari presenti sui dati grezzi. In questa fase oltre ad un primo stadio di controllo e correzione dei dati grezzi, si deve provvedere alla riaggregazione di alcuni codici tipo occupati e tipo contribuzione presenti nel DM10 in base alla omogeneità contributiva; questa fase viene discussa in modo approfondito nel capitolo 2;
2. controllo e correzione dei dati attraverso una procedura che utilizza editing selettivo, microediting e macroediting;
3. ricostruzione dell'impresa e trattamento delle anomalie delle unità di rilevazione ed eventuale

assegnazione di caratteri corretti sulla base dell'archivio ASIA; la procedura di ricostruzione viene descritta in modo approfondito nel capitolo 5;

4. eventuale stima e correzione delle variabili mancanti attraverso informazioni derivanti direttamente dai dati INPS e da un sistema di indagini supplementari di controllo. Alcune delle quali possono essere quelle attualmente già portate avanti dall'Istat in particolare la rilevazione mensile sulle grandi imprese (più di 500 addetti) e molte altre (SCI, Indagine quadriennale sul costo del lavoro, ESES, RTFL, ecc.). Eventualmente per avere dei benchmark dell'andamento congiunturale delle piccole e medie imprese potrebbero essere realizzate altre piccole indagini supplementari utilizzando tecniche diverse (postali, CATI, ecc.); una prima valutazione della qualità dei dati viene presentata nel capitolo 3;
5. stima congiunturale trimestrale dei parametri d'interesse utilizzando tutte le informazioni viste sopra, attraverso due passi distinti e tra loro sequenziali: nel primo passo si calcolano le stime aggiornate al tempo corrente dei parametri d'interesse utilizzando unicamente la fonte INPS, ed ipotizzando che quest'ultima riproduca in modo fedele la struttura della popolazione delle imprese con dipendenti. Le stime sono calcolate sulla base di un modello statistico di tipo regressivo che lega le variabili d'interesse (gli occupati e la retribuzione) a variabili ausiliarie di cui si può conoscere la distribuzione nell'universo delle imprese riportato negli archivi INPS; nel secondo passo si correggono le stime sopra definite utilizzando un modello statistico esplicativo delle differenze esistenti tra l'universo delle imprese con dipendenti, costruito in base agli archivi INPS e lo stesso universo definito però in base ad ASIA.
6. integrazione finale con i dati della Rilevazione mensile sulle grandi imprese.

Figura 1 - Il modulo DM10 INPS

Schema sintetico dei diversi passi della rilevazione



1.5 I problemi metodologici e di qualità dei dati

I dati dell'INPS hanno tutti i tipici problemi di qualità dei dati amministrativi. Il programma di lavoro fa riferimento ad uno schema di problemi e possibili errori campionari e non campionari di cui va compresa l'entità prima di procedere alla produzione del prototipo e della messa a regime del processo. Nella tavola sono rappresentati i principali tipi di errori nei dati derivanti da indagini statistiche con quelli di generici dati amministrativi e con quelli specifici della fonte INPS.

Indagine campionaria	Uso dei dati individuali di fonte amministrativa	Uso dei dati individuali di fonte INPS
1. Errori campionari	Sostanzialmente stessi tipi di errore dell'indagine campionaria	Valutazione rappresentatività del subset e riporto all'universo
2. Errori non campionari		
2.1 Errori nella base campionaria	Errori nell'archivio che contiene i record	Errori nell'archivio che contiene i record
2.1.1 Errori di sopra copertura: la lista contiene unità non necessarie all'indagine	Uguale all'indagine campionaria	Uguale all'indagine campionaria
2.1.2 Errori di sotto copertura: la lista non contiene unità necessarie all'indagine o le unità non contengono informazioni che sono necessarie all'indagine	I dati amministrativi non contengono tutta l'informazione desiderata, è quindi necessario introdurre un questionario supplementare	I dati INPS non coprono alcune aziende (FF.SS., l'editoria, ecc.) e/o settori (settori M, N, O) che hanno altri istituti di previdenza
2.1.3 Errori di classificazione delle unità incluse nel campione	La classificazione delle unità (per numero di addetti, per settore di attività economica) usata dalla fonte amministrativa non coincide con la classificazione statistica	<ul style="list-style-type: none"> La classificazione per attività economica dell'INPS (CSC) non è precisa, anche se l'INPS sta cercando di utilizzare il codice ATECO91 La classificazione delle qualifiche non è sempre corretta
2.1.4 Errori di definizione: errori derivanti dalla definizione della unità inclusa nel campione, che è diversa dalla definizione desiderata	La definizione dell'unità amministrativa (rilevata) e la definizione dell'unità di analisi sono differenti	<ul style="list-style-type: none"> L'unità di rilevazione è la posizione contributiva (diversa da impresa, unità locale, unità funzionale). Le informazioni prodotte non possono essere per unità funzionale come richiesto da molti regolamenti.
2.2 Errori di elaborazione: errori derivanti dalle procedure di trattamento dei dati - di codifica - di immissione - di editing - di ponderazione	Errori che vengono fatti al momento della inclusione dei dati nel computer. Inoltre, errori derivanti dalle procedure specifiche (es. procedure per il trasferimento dei file dalla PA all'istituto nazionale di statistica)	Stessi tipi di errore che per dati amministrativi.
2.3 Errori derivanti dalla non misurazione di alcune variabili, le cosiddette mancate risposte	Le mancate risposte sono in generale esigue. Invece può succedere che non sono disponibili gli archivi desiderati per varie ragioni	Utilizzando i dati grezzi è poco probabile che si verifichino DM10 mancanti, ma ci possono essere casi di ritardi.
2.4 Errori di misurazione errori dell'impresa al momento della compilazione del questionario	Errori dell'impresa al momento della compilazione del modulo della pubblica amministrazione, questi errori possono essere sistematici e non sistematici	Come nella generica fonte amministrativa. Esempio: inclusione dei lavoratori intermedi nella classe degli operai.
2.5 Errori del modello di misurazione: errori che derivano dall'avere assunto certe ipotesi	Errori che derivano dalla differenza tra lo scopo della pubblica amministrazione e lo scopo dell'indagine statistica	Le differenze fra le definizioni INPS e quelle utilizzate dall'Istat sono ampie per tutte le variabili

Fonte: Nostra integrazione sulla base della tabella presente in EUROSTAT(1998)

Mentre i problemi di tipo campionario verranno trattati nel Report finale in questo rapporto intermedio restringiamo la nostra attenzione sugli errori non campionari. Con riferimento ai dati INPS i problemi più importanti riguardano:

1. la definizione dell'unità di rilevazione (posizione contributiva);
2. I problemi di sottocopertura e sovracopertura;
3. la classificazione dell'attività economica.
4. i contenuti delle variabili (occupazione, retribuzione, oneri sociali)
5. l'assenza (o la presenza solo in casi particolari) di variabili rilevanti come ore lavorate;

Dei suddetti problemi si discute nei paragrafi e capitoli successivi.

1.6 I problemi di copertura della fonte INPS

La copertura settoriale dell'archivio dei moduli DM10/2 è molto ampia e riguarda tutte le imprese con lavoratori dipendenti che non hanno copertura previdenziale e assistenziale alternativa a quella INPS. Nel settore privato soltanto in alcuni settori esistono gestioni sostitutive a quella dell'INPS. Nella tavola che segue sono elencati gli enti gestori delle prestazioni pensionistiche alternative all'INPS.

Enti ed imprese con prestazioni pensionistiche alternative e/o sostitutive all'INPS

Nome Ente	presenti anche nell'INPS	nel campo di osservazione	n. dipendenti (assicurati)
FS (ferrovie dello stato)	NO	SI	125.848 (31.12.95)
IPOST (istituto postelegrafonici)	NO (SI solo contratti a tempo determinato)	SI	189.016 (31.12.95)
ANAS (confluita nella CTRS INPDAP dal 1996)	NO, (Si solo i nuovi assunti dal 1997)	SI	9.261 (31.12.96)
INPGI (Giornalisti)	SI - assicurati per fondi minori	SI	11.019 (31.12.96)
INPDAI (dirigenti di aziende industriali)	SI - assicurati per fondi minori	SI	118.267 (31.12.96)
Fondo nazionale previdenza impiegati imprese di spedizione e agenzie marittime IPSMA)	SI - assicurati per fondi minori	SI	40.356 (31.12.96)
ENPALS (lavoratori dello spettacolo)	in parte assicurati per fondi minori	NO	155.500 (31.12.96)
Altri Enti di minore dimensione (*)	NO	SI	300 circa
ENPAIA (ente naz. prev. Addetti impiegati agricoli)	NO	NO	31.611 (31.12.96)
ENAV, Ente nazionale di assistenza al volo	NO, Si solo i nuovi assunti dal 1997, il resto Tesoro	SI	3.100 (31.12.97)
Fondo previdenza personale INA	?	SI	681 (31.12.96)
INPDAP-fondi vari	SI soltanto per i lavoratori a tempo parziale fino al 1991	NO	
INPDAP-INS (asilo e scuole elementari parificate)	NO	NO	21.700 (31.12.96)

(*) Ordine Mauriziano, Torino; Istituto Triestino Per Interventi Sociali; Acega - Azienda Comunale Elettricità Gas Acqua, Trieste. Questi enti hanno in totale poche centinaia di dipendenti.

Escludendo i settori fuori dal campo di osservazione per gli altri enti si verificano due casi:

1. imprese in cui tutto il personale dipendente non ha copertura previdenziale INPS (Ferrovie dello stato, altri enti minori, ecc.);
2. imprese coperte parzialmente dall'INPS (es ENAV, con il 10% coperti dall'INPS, quelli assunti dal 1997 e il resto suddiviso fra Tesoro e INPDAP).

Per quanto riguarda i lavoratori del punto 1. si può fare ricorso semplicemente alla rilevazione sulle grandi imprese. Nel caso di imprese coperte parzialmente è opportuno recuperare le informazioni da altra fonte oltre quella INPS e stare ben attenti a non fare doppi conteggi. Nel caso degli enti minori vista la scarsità del numero dei casi è possibile trascurare questi dipendenti senza problemi.

1.7 Il contenuto delle variabili presenti sul DM10: una prima riflessione

Da un punto di vista generale la differenza tra le variabili amministrative e i concetti statistici può sorgere principalmente per due cause diverse:

1. le definizioni sono formalmente differenti.
2. le definizioni formali sono le stesse ma ci sono incentivi per le imprese a dare informazioni scorrette.

Il problema delle definizioni delle variabili amministrative non è al centro dell'attenzione nei documenti Eurostat (vedi ad es. Eurostat 1998) né delle discussioni (vedi ad esempio i paper presentati nel Seminario Eurostat di Gennaio 1998b). L'assunzione generale è che *"this problem is not different from the one faced by survey statisticians because we can generally expect to get from the enterprises only the information that is readily available, that is more or less the one required by some administration"* (Eurostat 1998b, p.2).

In altri termini il problema definitorio è inquadrato come un problema di qualità ed a questo proposito L'Eurostat Quality Report suggerisce l'uso di modelli per correggere i dati. In particolare quando c'è disponibilità di micro dati sono suggeriti modelli econometrici.

Mettiamo ora a confronto le diverse definizioni iniziando da quella degli occupati alle dipendenze. Gli occupati alle dipendenze secondo le definizioni in uso presso le indagini congiunturali e strutturali dell'Istat sono costituiti dal personale in forza alle imprese legato ad esse da un contratto di lavoro diretto, in base al quale percepisce una retribuzione. Nelle indagini congiunturali (grandi imprese) viene rilevato il numero degli occupati alle dipendenze presenti in libro paga alla fine del mese di riferimento. Nelle indagini annuali (SCI, PIC, ecc.) viene rilevata la media annua. Nell'archivio ASIA aggiornato annualmente è presente sia la media annua che il dato al 31 dicembre del totale dei dipendenti. Frequentemente nelle indagini si distinguono le seguenti posizioni nella professione: dirigenti, impiegati (inclusi gli intermedi e i quadri), operai, apprendisti. Il numero dei dipendenti nel modello DM10 è pari a quanti hanno percepito, nel mese di riferimento, compensi soggetti a contribuzione obbligatoria ed hanno almeno una giornata retribuita. I dipendenti vengono suddivisi per gruppi omogenei ai fini contributivi che corrispondono alle posizioni nella professione: operai, impiegati, dirigenti, apprendisti. Nel calcolo dei dipendenti non vengono compresi i lavoratori presenti in libro paga, ma assenti per tutto il mese senza diritto a retribuzione o che hanno percepito solo importi non imponibili. Nel caso in cui i lavoratori seppure assenti nel mese, beneficiano di integrazioni a carico del datore di lavoro (per malattia, maternità, cassa integrazione, ecc) hanno delle giornate retribuite e quindi risultano presenti.

Rispetto alle definizioni standard, quindi, l'occupazione dipendente rilevabile dai moduli DM10 ha alcuni problemi:

1. non viene rilevata la presenza dei lavoratori dipendenti rispetto al libro paga, ma rispetto alle giornate retribuite; i lavoratori in malattia, maternità, cassa integrazione a zero ore non risultano presenti tranne nel caso di integrazioni a carico dell'azienda;
2. essendo la misura non effettuata in un momento nel tempo, ma nell'arco di tutto il mese, maggiore è la variabilità a breve dell'occupazione (elevati movimenti occupazionali in entrata e in uscita) più ampia è la sovrastima dell'occupazione rispetto a rilevazioni puntuali.

In pratica il numero di occupati rilevabile sui moduli DM10 a livello mensile comporta delle distorsioni non irrilevanti rispetto a rilevazioni puntuali sulla base degli iscritti a libro paga. Occorre verificare se, come appare a prima vista probabile, rilevando il numero medio di occupati a livello trimestrale tali distorsioni si attenuino.

Retribuzioni lorde. Le retribuzioni rilevate nelle indagini congiunturali e strutturali sono dette lorde perchè in genere comprendono i contributi sociali e le imposte sul reddito a carico del lavoratore dipendente. Il concetto di retribuzione a cui si riferisce l'INPS è invece quello di retribuzione imponibile ai fini del calcolo dei contributi di previdenza ed assistenza in base alle leggi vigenti. L'art. 12 della legge n. 153 del 30 aprile 1969 definisce retribuzione imponibile "tutto ciò che il lavoratore riceve dal datore di lavoro, in denaro o in natura, al lordo di qualsiasi ritenuta, in dipendenza del rapporto di lavoro". Seguendo Revelli (1995) la retribuzione imponibile include:

- remunerazione del lavoro ordinario e straordinario
- integrazioni a carico del datore di lavoro delle indennità per malattie e infortuni erogate dagli enti assicuratori
- somme corrisposte dal datore di lavoro in aggiunta alla CIG
- diarie ed indennità di trasferta e trasferimento, rimborsi forfettari
- indennità una tantum, premi di anzianità e simili;
- compensi per ferie e festività non godute;
- generi in natura ceduti gratuitamente o a prezzo simbolico
- indennità sostitutiva di mancia e percentuali;
- indennità sostitutiva del preavviso, somme corrisposte a seguito di reintegrazione nel posto di lavoro
- somme dovute dal datore di lavoro per il periodo di congedo matrimoniale;
- retribuzioni corrispondenti ai permessi sindacali, alle giornate concesse ai lavoratori studenti, assegno mensile corrisposto al personale assente per servizio di leva;
- trattenute fiscali e previdenziali a carico del lavoratore.

L'analisi delle numerosissime voci retributive che in genere vengono considerate nelle definizioni Istat non coincidono esattamente con tutte le voci presenti nel monte retributivo INPS. Per un confronto approfondito delle voci incluse ed escluse nelle definizioni INPS e in quelle utilizzate presso l'Istat (che derivano sostanzialmente dai regolamenti comunitari, dal SEC, ecc) si vedano Revelli 1995 e Alemanni 1996.

Uno dei problemi più rilevanti che emerge è quello delle integrazioni a carico delle imprese (ad esempio in caso di malattia, infortuni o CIG). In presenza di eventi del genere non soltanto fra gli occupati risultano retribuiti lavoratori assenti dal processo produttivo ma nella retribuzione imponibile sono presenti tali integrazioni che se in termini contributivi sono retribuzione a tutti gli effetti in base alle definizioni che devono essere utilizzate per la stima sono in realtà oneri sociali a carico della azienda e quindi rientrano nel costo del lavoro ma non nella retribuzione lorda.

In ogni caso sulla base del confronto delle definizioni e delle prime sperimentazioni e verifiche effettuate presso l'Istat si può affermare che:

1. in teoria la rilevanza dello scostamento nel livello fra le due retribuzioni lorde si può ritenere verosimilmente estremamente limitato;
2. in pratica numerose imprese nel rispondere alle indagini dirette trascrivono nei moduli Istat il monte retributivo già segnalato sul DM10 dell'INPS;
3. anche in presenza di eventuali scostamenti nei livelli non necessariamente tali differenze si manifestano in diversi andamenti nelle variazioni delle retribuzioni.

Pertanto sulla base delle analisi effettuate, il dato INPS sulle retribuzioni lorde sembra quello che più direttamente possa essere utilizzato ai fini statistici congiunturali.

Per un approfondimento dei problemi definitivi e una valutazione accurata della qualità delle variabili INPS si veda il capitolo 5.

2 IDENTIFICAZIONE DEI CODICI PRESENTI SUL DM10 INPS E LA RICOSTRUZIONE DELLE VARIABILI OBIETTIVO

Questo capitolo si propone, nell'ambito del trattamento preliminare dei modelli DM10/2 INPS, l'obiettivo della riorganizzazione ed aggregazione dei codici presenti in tale modello. In particolare si ha lo scopo di fornire una partizione dei codici di interesse funzionale alla ricostruzione e calcolo delle seguenti variabili:

- Occupazione dipendente;
- Retribuzione lorda;
- Costo del lavoro;

2.1 Alcune note sulla contribuzione INPS

Ricordiamo che all'INPS devono essere versati i contributi obbligatori¹ (compresi quelli di solidarietà) per finanziare:

- il fondo pensione lavoratori dipendenti;
- il fondo pensioni per speciali categorie di lavoratori;
- le prestazioni economiche assistenziali (malattia, maternità, CIG, disoccupazione, mobilità, TBC, richiamo alle armi ed ogni altra forma di previdenza a carattere temporaneo diversa dalla pensione);
- il fondo garanzia TFR;
- la GESCAL (esclusi i dirigenti dell'industria);
- l'ENAOLI (l'erogazione delle prestazioni sono trasferite ai comuni);
- il fondo rimpatrio lavoratori extracomunitari.

I contributi, compresi quelli a carico dei lavoratori dipendenti, sono versati mensilmente dal datore di lavoro utilizzando il modello DM 10/2. Con la stessa procedura gli interessati conguagliano le somme a proprio credito.

La procedura in questione deve essere usata dai datori di lavoro appartenenti alle seguenti categorie:

- industria;
- artigianato;
- commercio;
- proprietari di fabbricati;
- professionisti ed artisti;
- credito, assicurazione e servizi tributari appaltati;
- agricoltura (con esclusione degli operai soggetti all'iscrizione negli elenchi nominativi dei lavoratori agricoli, i cui contributi sono riscossi dallo SCAU).

L'obbligo riguarda anche le amministrazioni pubbliche.

Inoltre l'INPS per alcune categorie di lavoratori gestisce fondi speciali di previdenza sostitutivi di quello generale, quali:

- Fondo per i dipendenti dall'Ente nazionale per l'energia elettrica e dalle aziende elettriche private;
- Fondo per il personale addetto ai pubblici servizi di trasporto;
- Fondo per il personale addetto ai pubblici servizi di telefonia;
- Fondo per il personale di volo dipendente dalle aziende di navigazione aerea;
- Fondo per il clero secolare e per i ministri di culto delle confessioni religiose diverse dalla cattolica.

Sempre gestiti dall'INPS vi sono poi dei fondi di previdenza integrativa quali:

¹ Nel periodo di riferimento 1995-99, taluni degli istituti contributivi richiamati hanno subito diverse modificazioni. Per esempio dal 01/01/1999 sono stati aboliti i cosiddetti oneri impropri (TBC, GESCAL, ENAOLI).

- Fondo per i lavoratori di miniere, cave e torbiere;
- Fondo per gli impiegati dipendenti dai concessionari del servizio di riscossione tributi;
- Fondo per il personale dipendente dalle aziende private del gas.

E' importante ricordare che vi sono, però, dei regimi previdenziali sostitutivi di quello generale dell'INPS gestiti da altri Enti quali la Cassa nazionale del notariato, le varie Casse nazionali di previdenza ed assistenza dei professionisti, il Fondo previdenziale spedizionieri doganali, l'ENPAIA per gli impiegati e i dirigenti dell'agricoltura, l'INPDAl per i dirigenti delle aziende industriali, l'ENASARCO per i rappresentanti di commercio, l'INPGI per i giornalisti, l'ENPALS per i lavoratori dello spettacolo ed i lavoratori sportivi.

Spetta, infine, all'INAIL la gestione assicurativa relativa agli infortuni ed alle malattie professionali. 2.2 Il modello DM/10

Il modello DM10 (Fig.1) contiene tutte le informazioni di carattere amministrativo necessarie per calcolare i debiti e i crediti contributivi dell'impresa relativamente all'INPS.

In questo modello sono pertanto presenti informazioni molto dettagliate ma non immediatamente utilizzabili per la costruzione delle variabili di nostro interesse che ricordiamo sono:

- occupazione
- retribuzione
- costo del lavoro

Fig. 1 – Esempio di dichiarazione contributiva DM10/2

A								
Matricola	CSC	C. ISTAT	Cod. Fiscale	Cod. Autorizzazioni	Sede INPS			
3406947599	10615	28622	276400108		3400			
Data Compilazione			N. dip. occupati	Lav. Tempo Det.				
9504			17					
B				C		D		
Dati retributivi				Somme a debito del datore di lavoro		Somme a credito del datore di lavoro		
Codice	N. dip.	N. gior.	Retr.					
10	7	140	16.456.000	7.216.000		35	290.000	
11	2	52	4.963.000	2.066.000		52	50.000	
Y000	1	86	1.135.000	473.000		L037	20.000	
1560	1	20	1.828.000	824.000		S197	456.000	
20	3	15	4.871.000	328.000		L172	166.000	
W000	2	52	4.154.000	1.871.000		L156	1.637.000	
S141	2	10		45.000				
W001				2.300.000				
						Totale B		2.619.000
						Totale A - B		12.504.000
Totale A				15.123.000		Totale B - A		

Dalla figura 1 è evidente come il modello DM10 sia articolato in più quadri, ognuno caratterizzato da una specifica finalità espositiva.

- Nel **quadro A** sono riportate le informazioni di tipo anagrafico relative cioè:

- all'identificazione dell'azienda (matricola aziendale, codice statistico contributivo², codice ISTAT, codice fiscale o partita iva);
- alla data di compilazione e sede di competenza INPS;
- ad alcune particolarità contributive aggiuntive a quelle implicite nel CSC individuate dai codice autorizzazione (nel campo "codice autorizzazione" del quadro A, possono essere presenti più codici autorizzazione. Tali codici alfanumerici, un numero ed una lettera, sono assegnati al momento dell'inquadramento o con variazione successiva, per l'individuazione di particolarità contributive più specifiche di quelle rappresentate dal CSC)

- Nel **quadro B** sono esposte, desunte dai libri paga e matricola dell'azienda, relativamente ad ogni codice ammissibile le diverse informazioni, ove richieste, relative al periodo di paga a cui si riferisce la denuncia rispetto a :

- numero dipendenti
- giornate retribuite
- importo delle retribuzioni imponibili

Il quadro B è infatti utilizzato tanto per l'esposizione dei codici cosiddetti "Occupazione" che di "altri codici" finalizzati ad evidenziare le più diverse particolarità:

contribuzione separata, fondi, contribuzione addizionale, aggiuntiva, ecc.

Ritornando alle variabili prima menzionate, queste sono così definite:

per **occupazione** si intende il numero di dipendenti che nel mese di riferimento hanno almeno una giornata retribuita soggetta a contribuzione INPS. In tal senso non vengono compresi quei lavoratori che pur presenti nel libro paga sono assenti per tutto il mese senza diritto a retribuzione o che hanno percepito solo importi non imponibili (indennità erogata per conto degli enti di previdenza: malattia, infortunio, ecc..).

Sono invece compresi quei lavoratori che pur assenti per l'intero mese beneficiano di integrazioni a carico del datore di lavoro (retribuzioni soggette quindi a contribuzione).

I dipendenti sono suddivisi per gruppi omogenei ai fini contributivi: operai, impiegati, apprendisti, dirigenti.

Sono presenti ulteriori distinzioni finalizzate ad evidenziare lavoratori a domicilio, part-time, con contratto di formazione lavoro e d altre particolarità.

Con **giornata retribuita** si intende una giornata in cui si è lavorato minimo un'ora e per la quale il datore di lavoro eroga una retribuzione soggetta a contribuzione obbligatoria. Nel caso di lavoratori part-time sono riportate le ore per cui il datore di lavoro eroga una retribuzione.

Per retribuzione si intende quella imponibile ai fini contributivi. Tale retribuzione è quella sulla quale devono essere calcolati i contributi previdenziali e assistenziali come individuati dall'articolo 12 della L.153/69 e da altre norme che interpretano la portata (salvo siano previsti salari medi o convenzionali).

In particolare, la retribuzione ai fini contributivi è individuata in "tutto ciò che il lavoratore riceve dal datore di lavoro in denaro o in natura al lordo di qualsiasi ritenuta in dipendenza del rapporto di lavoro".

La **retribuzione imponibile** ai fini contributivi comprende pertanto, oltre agli elementi costituenti il diretto corrispettivo della prestazione del lavoratore, tutto ciò che quest'ultimo percepisce dal datore di lavoro in relazione al rapporto instaurato ed in dipendenza di esso salvo i casi tassativamente esclusi dalla legge stessa.

² Il codice statistico contributivo individua il settore di appartenenza del datore di lavoro e determina il comportamento contributivo: esso è formato da cinque cifre delle quali la prima indica il settore, la seconda e la terza la classe, la quarta e la quinta la categoria.

Poiché la norma fa riferimento a quanto viene erogato dal datore di lavoro a proprio carico restano ovviamente esclusi dalla disciplina in esame - e sono quindi esenti da contribuzione - le somme corrisposte dal datore di lavoro per conto di terzi, quali ad esempio, le prestazioni anticipate dal datore di lavoro per conto di gestioni previdenziali e mutualistiche (assegni familiari, indennità in caso di malattie, di infortunio, di gravidanza, assegno per congedo matrimoniale, compenso ai donatori di sangue, ecc.) e le somme percepite dal lavoratore a titolo di mancia.

Secondo l'ormai costante giurisprudenza della Corte di Cassazione nella retribuzione imponibile ai fini assicurativi e previdenziali debbono essere comprese tutte le erogazioni corrisposte dai datori di lavoro a favore (diretto o indiretto, attuale, o secondo alcune pronunce, potenziale) dei lavoratori ed in costanza del rapporto di lavoro. Sono escluse soltanto le somme erogate per uno dei titoli tassativamente indicati dall'art. 12 della L. n. 153/1969, la cui elencazione deve ritenersi non suscettibile di interpretazione analogica, ma anche di interpretazione estensiva. Fra gli elementi imponibili rientra anche l'equivalente della retribuzione corrisposta in natura.

- Nel **quadro C** sono riportate, in corrispondenza delle indicazioni apposte nel quadro B le somme a debito del datore di lavoro.

E' importante ricordare che nel modello DM 10/2 viene riportato, per ciascuna categoria di lavoratore, l'importo complessivo degli oneri sociali (somme a debito del datore di lavoro), ovvero sia la parte a carico del datore di lavoro che quella a carico del dipendente.

Per predisporre i dati al fine di poter effettuare il calcolo del costo del lavoro è pertanto necessario riuscire a ripartire tale importo nelle sue componenti dato che nella retribuzione imponibile quadro B precedentemente definita è comunque compresa la quota a carico del lavoratore.

Tale operazione non è però di immediata e semplice esecuzione.

La maggior parte dei contributi deriva infatti, dall'applicazione di aliquote percentuali che, incidendo sulla retribuzione imponibile determinano i corrispettivi importi. Tuttavia, sia tra i diversi rami di attività economica che all'interno degli stessi, si evidenziano caratteristiche impositive diverse.

Tale eterogeneità contributiva è funzione sia delle specifiche caratteristiche del tipo di attività svolta, sia della dimensione aziendale. Ulteriore differenziazione contributiva è prevista con riferimento alle categorie di lavoratori: apprendisti, operai, impiegati, viaggiatori, dirigenti. Infine esistono specifiche agevolazioni previste per particolari forme di assunzione (CFL, apprendisti trasformati, assunti dalle liste di mobilità, ...).

- Nel **quadro D** sono esposti gli importi a credito del datore di lavoro. Tali importi derivano dalle prestazioni anticipate per conto dell'INPS (CIG, CIGS, malattia) oppure dalle agevolazioni contributive previste a norma di legge (sgravi, fiscalizzazione) e di altri importi che il datore di lavoro deve recuperare per titoli diversi.

Tali poste sono indicate nelle somme a credito del datore di lavoro, quadro d del modello, dove figurano gli importi posti a conguaglio nel mese essi, quindi, non sono necessariamente riferiti ai contributi versati nel mese stesso.

2.3 La riorganizzazione dei codici

Al fine di calcolare il dato relativo alle variabili di nostro interesse:

- numero dei dipendenti
- retribuzioni lorde
- contributi a carico del lavoratore e gli oneri a carico del datore

è indispensabile quindi operare a diversi livelli come sinteticamente illustrato in figura 2.

2.3.1 Individuazione dei codici occupazione

Per la ricostruzione del numero di dipendenti, il primo passo effettuato è stato quello relativo all'analisi dei codici presenti nei quadri B-C del modello.

Tali codici, individuano tanto gruppi omogenei di lavoratori ai fini contributivi (operai, impiegati, dirigenti, apprendisti, lavoratori assunti con CFL, ecc.) che molteplici particolarità espositive.

Si è proceduto quindi all'individuazione dei **codici "occupazione"**, esposti nel quadro B-C del DM10.

Codici occupazione, sono quei codici che individuano senza duplicazione il numero dei dipendenti, e a cui sono associate le rispettive retribuzioni imponibili e contribuzioni lorde, utilizzabili come base di calcolo per la determinazione degli aggregati di nostro interesse.

Sommando i dipendenti associati a tali codici, è possibile quindi ricostruire correttamente il numero totale dei dipendenti per DM10. Inoltre, tali codici, correttamente aggregati, permettono come detto l'individuazione di tipologie di lavoratori omogenee rispetto alla qualifica occupazionale (operai, impiegati, dirigenti e apprendisti) o al tempo di lavoro (full-time, part-time), funzionale all'analisi statistica della qualità del dato.

In particolare, vedi, sono stati individuati tre gruppi di "codici occupazione":

codici prestampati sul modello (0010, 0011, ecc.)

codici relativi ai soli *minatori in sotterraneo* (k100, k200, ecc.)

codici basati su codici tipo contribuzione.

In particolare, tali codici sono costituiti da 3 sotto campi:

- il primo sotto campo, costituito da 1 digit, individua la **qualifica**: 1 = operaio, 2 = impiegato, 3 = dirigente, ...)
- il secondo sotto campo, costituito da 2 digit, individua il **tipo contribuzione**, cioè la **particolarità contributiva** e può assumere valori compresi tra 00 e 99 e T4, T5 e T6 (00=nessuna particolarità contributiva, ..., T4=contributi CUAFF settore credito, ecc.)
- il terzo sotto campo, costituito da 1 digit, individua la **particolarità della retribuzione imponibile** (0=nessuna particolarità, P= part-time < 78 ore, ecc.)

2.3.2 Analisi degli altri codici presenti nei quadri B-C

Sempre con riferimento ai quadri B-C, sono stati analizzati gli "altri codici", vedi, cioè quelli diversi da quelli "occupazione".

Gli "altri codici" rappresentano:

1. delle specificazioni espositive dovute per particolari prestazioni contributive e riconducibili ad appropriati codice tipo occupazione, duplicandone così gli occupati, precedentemente esposti;
2. ulteriori grandezze, come per esempio straordinari, quota sindacale, quote associative, contribuzioni addizionali, rettifiche e restituzioni sgravi, restituzioni anticipazioni erranee, conguagli arretrati e molte altre.

Sono stati individuati circa 380 codici ammissibili (diversi da quelli tipo occupazione).

Questi codici sono stati analizzati rispetto alle diverse finalità contributive-espositive, dando luogo alla seguente bipartizione:

1. Codici da utilizzare nell'algoritmo relativo al calcolo del costo del lavoro (conguagli, restituzioni sgravi, restituzioni anticipazioni, ...)
2. Codici da non considerare per il computo del costo del lavoro (quota sindacale, quota associativa, ...)

2.3.3 Analisi dei codici presenti sul quadro D

Con riferimento al quadro D del modello, dove sono esposti gli importi a credito del datore di lavoro derivati da prestazioni anticipate per conto dell'INPS, oppure dalle agevolazioni contributive previste a norma di legge e di altri importi da recuperare, è stata effettuata una analoga operazione di analisi e riaggregazione sui codici ammissibili (circa 220).

Tali poste sono indicate nelle somme a credito del datore di lavoro dove figurano gli importi posti a conguaglio nel mese. Essi quindi, non sono necessariamente riferiti ai contributi versati nel mese stesso.

Anche in questo caso, si è prodotto una bipartizione dell'insieme dei codici ammissibili nel quadro D:

1. Codici che evidenziano crediti dovuti per la fiscalizzazione degli oneri sociali, sgravi contributivi e agevolazioni previste a norma di legge. La fiscalizzazione agisce attraverso la riduzione di talune aliquote contributive, relative principalmente all'assistenza sanitaria, differenziate per settore ed area geografica. Lo sgravio, che riguarda le aziende localizzate nel Mezzogiorno, consiste in una riduzione degli oneri contributivi dovuti dal datore di lavoro per le assicurazioni obbligatorie; esso si applica soprattutto sui contributi previsti per previdenza invalidità e vecchiaia. Le agevolazioni, infine, determinano un conguaglio della quota dei contributi non dovuti per un certo intervallo di tempo, relativamente a particolari modalità di assunzioni (apprendisti trasformati, giovani con diploma di qualifica, formazione e lavoro, lavoratori con contratto di reinserimento, ecc...).

Gli importi relativi a tali codici vanno quindi sottratti dalla contribuzione lorda per ottenere quella netta, quella cioè effettivamente versata dal datore di lavoro;

2. Codici che non devono essere considerati nel calcolo del costo del lavoro in quanto funzionali a dar evidenza di una posta di giro (indennità di malattia, maternità assegni familiari,..). Tali codici sono utilizzati, cioè per l'esposizione di importi a credito del datore di lavoro, derivanti da prestazioni anticipate per conto dell'INPS, che non figurano nella retribuzione imponibile e posti a conguaglio.

2.3.4 Ricostruzione delle aliquote contributive

Per calcolare il costo del lavoro in base ai dati presenti sul modello DM 10/2, è necessario avere una completa conoscenza delle aliquote contributive.

Infatti, una volta evidenziate le diverse tipologie di lavoratori e le relative retribuzioni imponibili ai fini contributivi, è necessario scomporre le corrispondenti somme a debito nelle due componenti:

- quella a carico del lavoratore (già compresa nella retribuzione imponibile);

- quella a carico del datore di lavoro.

Si è pertanto proceduto a ricostruire la serie storica delle aliquote contributive in vigore dal 1995 al 1999.

Tali aliquote, sono funzione delle seguenti grandezze:

- il settore contributivo di appartenenza dell'azienda (industria, commercio, artigianato...)
- il tempo (nel corso del periodo considerato sono intervenute diverse modifiche legislative volte a modificare la struttura contributiva stante in essere)
- la dimensione aziendale (<15 dipendenti, 15-50 dipendenti,...)
- la categoria di appartenenza del lavoratore (operai, impiegati, dirigenti, apprendisti)
- l'applicazione di particolari forme di assunzione (individuabili dai codici occupazione: per esempio il C.F.L. e risolvibili conseguentemente con il corretto utilizzo dei codici del quadro D)

A tal fine è stato predisposto un set di tabelle, (una per ogni settore contributivo di interesse), che con riferimento ai primi quattro punti permettono l'individuazione di tali aliquote (sia complessive che disaggregate per tipologia di forma assicurativa: fondo pensioni lavoratori dipendenti, fondo malattia, fondo mobilità, ...).

2.4 Esempio puntuale di calcolo del costo del lavoro su base DM 10/2

Con riferimento alla metodologia illustrata nei paragrafi precedenti è di seguito riportato un breve esempio applicativo.

Sulla base del DM 10/2 riportato (tabella allegata a fine sezione) illustriamo le diverse fasi del calcolo del costo del lavoro per posizione contributiva.

1. Il primo passo consiste nell'analizzare i codici presenti nei **quadri B-C** del modello, che individuano gruppi omogenei di dipendenti ai fini contributivi.

In particolare distinguiamo due macro tipologie di codici:

“**Codici occupazione**”

“**Altri codici**”

2. Il secondo passo consiste nella scomposizione delle corrispettive somme a debito nelle due componenti: quella a carico del lavoratore (già compresa nella retribuzione imponibile) e quella a carico del datore di lavoro. A tale fine è necessario conoscere la corretta aliquota da applicare alla retribuzione imponibile. Tale informazione è desumibile dalla conoscenza del **CSC**, dagli eventuali **codici autorizzazione** e dalla **data di competenza** della contribuzione. Nel nostro esempio abbiamo:

- codice CSC: 10932 che individua un'azienda del settore dell'industria che si occupa della fabbricazione di altri prodotti in gomma, con più di 50 addetti (Numero dipendenti occupati quadro A).
- data di competenza: settembre 1995
- codici autorizzazione: non presenti

Nel nostro esempio abbiamo i seguenti codici:

Codice	Descrizione
Codici occupazione (individuano univocamente il numero dei dipendenti)	
10	Operai full-time con contratto a tempo indeterminato (senza particolari agevolazioni contributive). Sono specificati il numero di dipendenti, le giornate retribuite, l'ammontare della retribuzione imponibile e il corrispondente importo a debito.
11	Impiegati full-time con contratto a tempo indeterminato (senza particolari agevolazioni contributive) Sono specificati il numero di dipendenti, le giornate retribuite, l'ammontare della retribuzione imponibile e il corrispondente importo a debito.
3000	Dirigenti full-time con contratto a tempo indeterminato (senza particolari agevolazioni contributive) Sono specificati il numero di dipendenti, le giornate retribuite, l'ammontare della retribuzione imponibile e il corrispondente importo a debito.
2750	Lavoratori in mobilità assunti con qualifica di impiegato con contratto a tempo indeterminato (art. 25, comma 9, L.223/91). Per i datori di lavoro che assumono con contratto a tempo indeterminato lavoratori iscritti alle liste di mobilità, i contributi da versare sulle retribuzioni di detti dipendenti - per i primi 18 mesi - sono pari a quelli previsti per gli apprendisti. Tale agevolazione riguarda esclusivamente l'aliquota a carico del datore di lavoro, rimane ferma la quota a carico del dipendente nella misura prevista per la generalità dei lavoratori. Sono specificati il numero di dipendenti, le giornate retribuite, l'ammontare della retribuzione imponibile e il corrispondente importo a debito al lordo delle agevolazioni previste.
O000	Operai part-time con contratto a tempo indeterminato (senza particolari agevolazioni contributive) Sono specificati il numero di dipendenti, le ore retribuite, l'ammontare della retribuzione imponibile e il corrispondente importo a debito.
Y000	Impiegati part-time con contratto a tempo indeterminato (senza particolari agevolazioni contributive) Sono specificati il numero di dipendenti, le ore retribuite, l'ammontare della retribuzione imponibile e il corrispondente importo a debito.
Codici di cui	
S165	Contributo apprendisti NO INAIL, (riferito al codice 2750) per lavoratori assunti dalle liste di mobilità L.223/91. Sono specificati il numero di dipendenti, il numero delle settimane e l'importo dei contributi settimanali (in cifra fissa) dovuti.
22	Lavoro straordinario (contribuzione al 15%). Sono specificati il numero complessivo di dipendenti che hanno effettuato straordinario e il numero di ore.

Dalle informazioni, così individuate, risaliamo alle seguenti aliquote contributive di riferimento (in vigore dal 01/01/93 al 30/09/95):

	TOTALE INPS	DI CUI A CARICO DEL DIP.	DI CUI A CARICO DEL DAT. LAV.
OPERAI	45,35%	8,99%	36,36%
IMPIEGATI	43,13%	8,99%	34,14%
DIRIGENTI	10,28%		10,28%

	Quota fissa settimanale	Aliquota a carico del lavoratore
APPRENDISTI INAIL	4.772	5,34%
APPRENDISTI NO INAIL	4.592	5,34%

1. il terzo passo si riferisce all'analisi degli importi presenti nel **quadro D** (somme a credito del datore di lavoro). Nell'esempio, abbiamo:

CODICE	DESCRIZIONE
<i>Codici relativi ad anticipazioni</i>	
D800	Indennità riposi per allattamento art. 8 L.903/77
S110	Donatori di sangue

<i>Codici relativi a sgravi, fiscalizzazioni ed agevolazioni</i>	
L180	Lavoratori assunti dalle liste di mobilità. Riduzione art. 8 c.2 e art. 25 c.9 L.223/91

Possiamo allora procedere al calcolo del costo del lavoro, relativamente ai dati desunti dal DM10/2:

CODICE	RETRIBUZIONE	SOMME A DEBITO DEL DATORE DI LAVORO		SOMME A CREDITO	
		DATORE LAV.	LAVORATORE	CODICE	IMPORTO
10	62.338.000	22.666.000	5.604.000	D800	514.000
11	29.105.000	9.936.000	2.617.000	S110	106.000
3000	12.142.000	1.248.000		L180	2.528.000
O000	1.740.000	633.000	156.000		
Y000	2.587.000	883.000	233.000	TOT.	2.528.000
2750	7.404.000	2.528.000	666.000		
S165		92.000			
22	192.000	29.000			
TOT.	115.508.000	38.015.000			

COSTO DEL LAVORO =115.508.000+38.015.000-2.528.000=***150.995.000***

Analogamente è possibile calcolare il costo del lavoro per tutti i diversi codici occupazione che identificano i gruppi omogenei ai fini contributivi. Ad esempio in questo caso è possibile identificare inequivocabilmente il costo del lavoro dei seguenti gruppi di lavoratori: gli operai a tempo indeterminato e pieno (10), gli impiegati tempo indeterminato e pieno (11), gli operai a tempo parziale (O000), gli impiegati a tempo parziale (Y000), gli impiegati assunti in mobilità con contratto a tempo indeterminato (2750). Il costo del lavoro dei dirigenti (3000) non include una parte che viene versata direttamente alla INPDAI (si tratta principalmente dell'importo relativo all'aliquota di circa il 33% per IVS).

A questi importi andrebbe poi aggiunta la quota di accantonamento mensile del trattamento di fine rapporto e il premio INAIL.

DM 10/2

Il CSC non individua univocamente l'ATECO 91

N. dip. occupati \geq Σ N. dip. relativi ai cod. occupazione quadro B

Aiuta ad individuare l'attività e la corrispondente particolarità contributiva

A	Matricola	CSC	C. ISTAT	Cod. Fiscale	Cod. Autorizzazioni	Sede INPS					
	8700297874	10932	25130	276400108		3400					
	Data Compilazione		N. dip. occupati		Lav. Tempo Det.						
	95-09		51								
B	Dati retributivi				C	Somme a debito del datore di lavoro		D	Somme a credito del datore di lavoro		
	Codice	N. dip.	N. gior.	Retr.							
	10	30	770	62.338.000		28.270.000	D800		514.000		
	11	12	285	29.105.000		12.553.000	S110		106.000		
	3000	2	52	12.142.000		1.248.000	L180		2.528.000		
	2750	4	104	7.404.000		3.193.000					
	O000	2	168	1.740.000		789.000					
	Y000	1	84	2.587.000		1.116.000					
	S165	4	20			92.000					
	22	5	15	192.000		29.000					
Totale A					47.290.000	Totale B	3.148.000				
						Totale A - B	44.142.000				
						Totale B - A					

E' necessario distinguere tra codici "occupazione" e codici "di cui".

Giornate: in genere
Ore: per i part-time e straordinario
Settimane: per contribuzione tipo apprendista

In genere:
Contr. a carico del datore di lavoro
+
contr. a carico del lav.

Importi a credito per prestazioni anticipate per conto INPS o per agevolazioni contributive (non necessariamente riferiti ai contributi versati nel mese stesso)

3 LA QUALITÀ DEI DATI INPS: UN CONFRONTO CON L'INDAGINE ISTAT LPS PER L'ANNO 1995

3.1 Introduzione

Lo scopo del capitolo è quello di valutare la qualità dei dati INPS con riferimento al problema specifico riguardante il contenuto delle variabili amministrative ed il loro rapporto con le variabili statistiche.

La differenza tra le variabili amministrative e le variabili statistiche può dipendere da tre principali fattori:

1. le differenze definitorie;
2. la procedura utilizzata per il calcolo delle variabili di fonte INPS;
3. gli incentivi per le imprese a dare informazioni scorrette (ad esempio l'evasione contributiva).

Il lavoro è organizzato nel modo seguente. In primo luogo, si procede al confronto tra le definizioni utilizzate dall'Istat e dall'INPS per rilevare la grado di corrispondenza tra le variabili di fonte amministrativa e statistica. Quindi, sulla base dei risultati di una analisi univariata effettuata sulle differenze relative tra variabili amministrative e variabili statistiche, viene fornita, per ciascuna variabile oggetto di esame, una misura (in media e in mediana) della distorsione complessiva. Dove possibile, si tenta di isolare la componente imputabile esclusivamente alle differenze definitorie. Inoltre, allo scopo di valutare l'impatto sulla distorsione delle principali caratteristiche strutturali delle imprese (attività economica, dimensione, e localizzazione territoriale) si effettuano alcune analisi di regressione. Infine, viene eseguita un'analisi per strati (individuati in base alla qualifica ed al tempo di lavoro dei dipendenti) al fine di verificare la presenza o meno di problemi specifici, eventualmente, imputabili anche ad errori nella procedura di calcolo utilizzata.

L'analisi è effettuata su tre variabili: occupati dipendenti, retribuzioni lorde e costo del lavoro. Sono analizzati i livelli mensili (ultimi mesi di ogni trimestre: marzo, giugno, settembre, dicembre), i livelli medi trimestrali e il livello medio annuo. Inoltre sono analizzate le differenze relative delle variazioni mensili e trimestrali.

In particolare, il paragrafo 2 riporta le specifiche e i limiti dell'analisi effettuata. I paragrafi 3, 4 e 5, riguardano, rispettivamente, le analisi sull'occupazione dipendente, le retribuzioni lorde ed il costo del lavoro. Infine, nel paragrafo 6 si traggono alcune conclusioni.

3.2 Specifiche e limiti dell'analisi

La distorsione delle variabili amministrative può essere concepita come un problema di misurare una variabile vera con una variabile disponibile.

In particolare, definiamo:

Y_i misurazione della variabile y per l'osservazione i

X_i valore vero (o preferito) per l'osservazione i

$B_i = Y_i - X_i$ errore di misurazione assoluto per l'osservazione i (bias assoluto)

$b_i = Y_i - X_i / X_i$ errore di misurazione assoluto per l'osservazione i (bias relativo)

Gli scopi dell'analisi sono:

1. misurare la grandezza del bias relativo;
2. valutare la casualità o la sistematicità dell'errore;
3. valutare l'ipotesi che tale bias sia costante nel tempo;

Nel contesto specifico analizzato, i dati dei DM10 rappresentano le misure dei valori veri. Per approssimare questi ultimi usiamo i dati dell'indagine ISTAT LPS001. La scelta di un tale benchmark dipende dal fatto che, al momento, l'indagine LPS costituisce la sola fonte statistica a carattere congiunturale disponibile a parte l'indagine sulle grandi imprese. Quest'ultima non è utilizzabile, in questo ambito, in quanto, come noto, si riferisce soltanto alle imprese con più di 500

dipendenti, per le quali, oltretutto, si hanno i maggiori problemi di ricostruzione.

L'insieme di partenza è un set di imprese unilocalizzate rilevate dall'indagine LPS per l'anno 1995. Per esse sono disponibili i dati relativi ai dodici mesi dell'anno. Dall'universo dei DM10 riferiti al 1995 sono stati estratti quelli associati alle suddette imprese. La chiave di estrazione usata è il codice fiscale.

La scelta di usare le imprese unilocalizzate deriva dalla necessità di isolare, per quanto possibile l'ulteriore distorsione derivante dalla ricostruzione delle imprese di grande dimensione e quindi presumibilmente plurilocalizzate. L'analisi è stata condotta su un subset costituito da 5.648 imprese ottenuto dall'abbinamento delle imprese rilevate dalle due fonti. In particolare le imprese INPS sono state ricostruite riunendo tra loro le posizioni contributive con stesso codice fiscale³.

La Tabella 1 riporta la distribuzione delle imprese e dei dipendenti, per tipologia d'impresa rispetto al numero di posizione contributive INPS da cui essa è costituita.

Tabella 1 - distribuzione per tipologia dell'impresa rispetto al numero di posizione contributive da cui è costituita

TIPOLOGIA PER N° DI POSIZIONI ⁴	Imprese		Dipendenti INPS		Dipendenti LPS (inizio mese+fine mese)/2	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Uniposizionate	5.050	89,2	275.661	76	274.773	77
Pluriposizionate	611	10,8	87.111	24	80.396	23
Totale	5.661	100	362.772	100	355.169	100

Le Tabelle 2, 3 e 4 sintetizzano la composizione del subset di imprese relativamente all'attività economica, alla dimensione di impresa, alla localizzazione territoriale. I dipendenti sono calcolati come media annua. Le variabili di classificazione sono quelle derivanti dall'indagine LPS.

³ In questo lavoro, le imprese INPS sono ricostruite in maniera del tutto provvisoria. Infatti, in primo luogo, non viene effettuato nessun tipo di controllo sul codice fiscale (pertanto può accadere che ad alcune imprese vengano associate posizioni contributive di altre imprese) inoltre, la ricostruzione riguarda soltanto le variabili oggetto d'analisi (occupati dipendenti, retribuzioni lorde e costo del lavoro), mentre le variabili anagrafiche dell'impresa (codice di attività economica, localizzazione territoriale, dimensione, ecc.) sono prelevate direttamente dalle imprese LPS abbinate alle imprese INPS, ipotizzando in tal modo l'assenza di classificazioni INPS scorrette.

⁴ A partire dai dati forniti dall'archivio anagrafico INPS aggiornato al 1998 è effettuata un'operazione di conteggio del numero di posizioni contributive relative a ciascuna impresa. Sulla base di tale informazione si individuano le imprese del subset costituite effettivamente da una sola posizione contributiva (imprese uniposizionate) e le imprese costituite da più posizioni contributive (imprese pluriposizionate).

Tabella 2 - distribuzione per sezione-divisione

SEZIONE-DIVISIONE ⁵		Imprese		Dipendenti INPS		Dipendenti LPS (inizio mese+fine mese)/2	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
cejo	Estraz. miner. + Energia + Intermed. + Altri servizi pubblici	157	2,8	6.667	1,8	7.806	2,2
da	Alimentari	566	10,0	31.325	8,6	32.608	9,2
db	Tessili	598	10,6	43.809	12,1	43.235	12,2
dc	Conciarie	275	4,9	13.959	3,8	13.010	3,7
dd	Legno	241	4,3	8.861	2,4	8.737	2,5
de	Carta	267	4,7	17.045	4,7	17.397	4,9
df	Coke, petrolio	24	0,4	971	0,3	860	0,2
dg	Prodotti chimici	100	1,8	12.557	3,5	11.078	3,1
dh	Gomma, plastica	213	3,8	12.059	3,3	11.828	3,3
di	Minerali non metalliferi	270	4,8	17.098	4,7	16.592	4,7
dj	Metallo	487	8,6	38.445	10,6	37.582	10,6
dk	Apparecchi meccanici	299	5,3	34.160	9,4	32.417	9,1
dl	Macchine elettriche	312	5,5	27.519	7,6	27.489	7,7
dm	Mezzi di trasporto	178	3,1	19.521	5,4	20.003	5,6
dn	Altre industrie manifatturiere	298	5,3	13.226	3,6	12.775	3,6
g	Commercio	601	10,6	17.524	4,8	17.015	4,8
h	Alberghi	199	3,5	6.605	1,8	5.583	1,6
i	Trasporti	302	5,3	23.165	6,4	20.297	5,7
k	Attività immobiliari	274	4,8	18.257	5,0	18.857	5,3
	Totale	5.661	100	362.772	100	355.169	100

Tabella 3 - distribuzione per dimensione

CLASSE DIMENSIONALE ⁶	Imprese		Dipendenti INPS		Dipendenti LPS (inizio mese+fine mese)/2	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
< 15	1.504	26,6	18.127	5,0	17.017	4,8
15-50	2.445	43,2	70.549	19,4	66.022	18,6
50-100	874	15,4	64.252	17,7	61.878	17,4
>100	838	14,8	209.844	57,8	210.253	59,2
Totale	5.661	100	362.772	100	355.169	100

⁵ Tale variabile è un ibrido tra la sezione e la sotto sezione di attività economica. In particolare, le sezioni di attività economica poco significative sono aggregate tra loro, viceversa, per le sezioni di attività economica più significative si scende nel dettaglio della sotto sezione.

⁶ Alle imprese viene attribuita una classe dimensionale calcolata facendo riferimento al numero medio annuo di dipendenti LPS. I dipendenti mensili LPS sono stati calcolati come semi somma dei dipendenti alla fine del mese corrente e alla fine del mese precedente.

Tabella 4 - distribuzione per macro regione

MACRO REGIONE ⁷	Imprese		Dipendenti INPS		Dipendenti LPS (inizio mese+fine mese)/2	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Nord-est	1.558	27,5	125.482	34,6	123.138	34,7
Nord-ovest	1.455	25,7	97.792	27,0	96.217	27,1
Centro	1.360	24,0	75.907	20,9	74.403	20,9
Sud	971	17,2	48.118	13,3	47.681	13,4
Isole	317	5,6	15.474	4,3	13.729	3,9
Totale	5.661	100	362.772	100	355.169	100

3.3 Occupati dipendenti

3.3.1 Definizioni a confronto

Il numero di dipendenti desumibile dai modelli DM10 dell'INPS è pari al totale di quanti hanno percepito, nel mese di riferimento, compensi soggetti a contribuzione obbligatoria, cioè, presentano almeno una giornata retribuita. I dipendenti vengono, inoltre, suddivisi per gruppi omogenei ai fini contributivi. Nel computo non sono compresi, anche se in forza, i lavoratori che sono risultati assenti per tutto il mese senza diritto a retribuzione e i lavoratori presenti in libro paga ma che hanno percepito importi non imponibili (Alemanni, 1996).

L'indagine LPS rileva gli occupati alle dipendenze costituiti dal personale in forza all'impresa, legato ad essa da un contratto di lavoro diretto in base al quale percepisce una retribuzione.

L'indagine rileva in maniera puntuale sia il numero di dipendenti alla fine del mese di riferimento che quello alla fine del mese precedente. Inoltre registra il movimento occupazionale avvenuto in ciascun mese rilevando il numero dei dipendenti entrati e di quelli usciti nel corso di esso (Note alla compilazione del questionario LPS).

Risulta evidente, quindi, che esistono delle differenze concettuali tra le definizioni adottate dalle due fonti, differenze che riflettono, come è naturale, i diversi scopi conoscitivi che ciascuna è chiamata soddisfare.

In particolare, rispetto alle definizioni statistiche, l'occupazione dipendente rilevabile dai moduli DM10 può risultare differente per due ordini di motivi: 1) i lavoratori in libro paga, ma con zero giornate retribuite, non sono registrati nei moduli DM10; 2) rientra nell'occupazione INPS anche un dipendente non in forza alla fine del mese ma che abbia percepito retribuzione imponibile nel corso del mese. Le due fonti di distorsione operano in maniera opposta: la prima può determinare una sottostima della variabile desiderata, mentre la seconda una sovrastima. In particolare quest'ultima è tanto maggiore quanto maggiore è la variabilità a breve dell'occupazione.

3.3.2 Una valutazione della distorsione dell'occupazione dipendente

Come accennato precedentemente, lo scopo del lavoro è quello di misurare l'errore derivante dall'utilizzo dei dipendenti INPS come misura della variabile statistica occupazione dipendente. Come approssimazione della incognita variabile vera è usata l'occupazione dipendente LPS calcolata come semi somma dei dipendenti alla fine del mese di riferimento e alla fine del mese

⁷ Vengono prese in considerazione 5 ripartizioni (nord-est, nord-ovest, centro, sud, isole) ricavate a partire dalle regioni dell'indagine LPS.

precedente. Viene, inoltre, considerata una seconda variabile relativa all'occupazione dipendente LPS, il cui contenuto definitorio è sostanzialmente omogeneo a quello della variabile INPS. Tale variabile è calcolata come somma dei dipendenti alla fine del mese e dei dipendenti usciti nel corso del mese. Sono stati pertanto calcolati due misure di bias relativo. La prima, che chiameremo bias complessivo, si riferisce alla distorsione della variabile INPS rispetto alla variabile statistica desiderata (variabile LPS 1^a definizione). La seconda, bias netto, si riferisce alla distorsione rispetto alla variabile omogenea dal punto di vista definitorio (variabile LPS 2^a definizione).

In formule, per la generica impresa i nel mese t ⁸, si indicano con:

$LPS D_{i t}$	i dipendenti rilevati dall'indagine LPS alla fine del mese t ,
$LPS D_{i (t-1)}$	i dipendenti rilevati dall'indagine LPS alla fine del mese $t-1$,
$LPS U_{i t}$	i dipendenti rilevati dall'indagine LPS usciti nel corso del mese t ,
$LPS^1 D_{i t} = (LPS D_{i t} + LPS D_{i (t-1)})/2$	la variabile da misurare
$LPS^2 D_{i t} = LPS D_{i t} + LPS U_{i t}$	la somma dei dipendenti LPS alla fine del mese t e dei dipendenti usciti nel mese,
$INPS D_{i t}$	i dipendenti ricavabili dai DM10 INPS ⁹ ,
$(INPS D_{i t} - LPS^1 D_{i t}) / LPS^1 D_{i t}$	il bias relativo complessivo,
$(INPS D_{i t} - LPS^2 D_{i t}) / LPS^2 D_{i t}$	il bias relativo netto ¹⁰ .

Si comincia ad analizzare il bias complessivo. I principali risultati dell'analisi univariata (media, deviazione standard, principali quantili della distribuzione) sono riportati nella tabella 5.

La tabella mostra come, in media, i dipendenti INPS risultino superiori a quelli rilevati dall'indagine LPS, in tutti i casi considerati. Sia le differenze relative dei livelli che delle variazioni non oltrepassano l'8%. Al variare del mese il bias complessivo medio risulta pressoché costante. Ne deriva che le differenze delle variazioni risultano sistematicamente inferiori a quelle dei livelli. Comunque, sia per i livelli che per le variazioni gli scostamenti maggiori si riscontrano nel mese di dicembre e nel trimestre corrispondente. Tali evidenze si rafforzano notevolmente guardando ai quantili della distribuzione. In particolare, la mediana risulta, nella maggior parte dei casi, nulla e, l'80% delle osservazioni presentano scostamenti relativi, in valore assoluto, non superiori al 10%. Con riferimento al bias medio annuo questo si attesta intorno al 6% in media ed intorno al -1,5% in mediana.

Al fine di studiare gli scostamenti tra le due fonti non imputabili a differenze definitorie la tabella 6 riporta l'analisi per il bias netto.

I risultati su tale variabile sono, come atteso, in massima parte migliori rispetto a quelli ottenuti sulla distorsione complessiva.

La componente definitoria è stata isolata semplicemente portando in detrazione dalle misure centrali del bias complessivo le misure centrali del bias netto. La tabella 7 mostra una stima dell'errore distorsione per il totale dei dipendenti, per settore di attività economica e per classe dimensionale

⁸ Naturalmente, con l'indice t si può indicare qualunque periodo temporale diverso dal mese (trimestre, anno, ecc.).

⁹ Il calcolo dei dipendenti INPS è ottenuto sommando i dipendenti associati ai codici occupazione del modulo DM10. Per una dettagliata analisi dei codici si rimanda al capitolo precedente.

¹⁰ Oltre ai bias dei livelli è possibile calcolare bias delle variazioni. In formule si indica con:

$LPS^1 D_{i (t/t-1)} = LPS^1 D_{i t} / LPS^1 D_{i (t-1)}$ la variazione tra il mese t ed il mese $(t-1)$ della variabile da misurare,

$LPS^2 D_{i (t/t-1)} = LPS^2 D_{i t} / LPS^2 D_{i (t-1)}$ la variazione tra il mese t ed il mese $(t-1)$ della variabile LPS omogenea,

$INPS D_{i (t/t-1)} = INPS D_{i t} / INPS D_{i (t-1)}$ la variazione tra il mese t e $(t-1)$ dei dipendenti ricavabili dai DM10

$(INPS D_{i (t/t-1)} - LPS^1 D_{i (t/t-1)}) / LPS^1 D_{i (t/t-1)}$ il bias relativo complessivo,

$(INPS D_{i (t/t-1)} - LPS^2 D_{i (t/t-1)}) / LPS^2 D_{i (t/t-1)}$ il bias relativo netto.

dell'impresa.

Nel caso dei livelli, l'errore di definizione medio annuo risulta in mediana pari a circa l'1%. Tale errore risulta, in media, più elevato e si attesta intorno al 2%

Tabella 5 – analisi univariata sul bias complessivo dei dipendenti

Variabile	Riferimento temporale	Numero imprese	Media	Dev. Stand.	P1	P10	Q1	Mediana	Q3	P90	P99	
BIAS RELATIVO COMPLESSIVO	L I V E L L I	Marzo	5639	0,070	1,263	-0,697	-0,034	0,000	0,00000	0,027	0,091	1,154
		Giugno	5638	0,075	1,275	-0,699	-0,018	0,000	0,00000	0,033	0,097	1,174
		Settembre	5636	0,076	1,293	-0,657	-0,031	0,000	0,00000	0,032	0,098	1,273
		Dicembre	5622	0,081	1,247	-0,689	-0,011	0,000	0,00000	0,029	0,100	1,316
		1° trim	5653	0,061	1,264	-0,689	-0,036	0,000	0,00546	0,025	0,073	1,164
		2° trim	5657	0,067	1,257	-0,677	-0,020	0,000	0,00715	0,027	0,079	1,176
		3° trim	5659	0,064	1,285	-0,655	-0,030	0,000	0,00422	0,023	0,070	1,148
		4° trim	5654	0,068	1,264	-0,718	-0,025	0,000	0,00654	0,026	0,083	1,209
		Anno	5661	0,062	1,254	-0,656	-0,038	-0,002	0,00736	0,023	0,065	1,115
	V A R I A Z I	Giu/Mar	5619	0,018	0,277	-0,261	-0,046	-0,012	0,00000	0,023	0,070	0,400
		Set/Giu	5616	0,023	0,605	-0,293	-0,062	-0,023	0,00000	0,013	0,053	0,467
		Dic/Set	5599	0,059	2,303	-0,320	-0,055	-0,015	0,00000	0,022	0,070	0,667
		2° trim/1° trim	5650	0,019	0,246	-0,324	-0,035	-0,011	0,00000	0,017	0,054	0,536
		3° trim/2° trim	5655	0,010	0,330	-0,334	-0,048	-0,017	0,00000	0,010	0,034	0,500
	·	4° trim/3° trim	5652	0,011	0,186	-0,345	-0,035	-0,011	0,00000	0,016	0,051	0,500

Tabella 6 - analisi univariata sul bias netto dei dipendenti

Variabile	Riferimento temporale	Numero imprese	Media	Dev. Stand.	P1	P10	Q1	Mediana	Q3	P90	P99	
BIAS RELATIVO NETTO	L I V E L L I	Marzo	5638	0,048	1,256	-0,700	-0,056	-0,008	0,00000	0,000	0,024	1,143
		Giugno	5638	0,053	1,273	-0,729	-0,040	0,000	0,00000	0,000	0,030	1,143
		Settembre	5635	0,050	1,279	-0,680	-0,053	-0,009	0,00000	0,000	0,029	1,158
		Dicembre	5622	0,059	1,230	-0,726	-0,031	0,000	0,00000	0,000	0,042	1,200
		1° trim	5653	0,039	1,257	-0,724	-0,056	-0,020	0,00000	0,000	0,026	1,098
		2° trim	5657	0,046	1,232	-0,692	-0,039	-0,009	0,00000	0,000	0,029	1,113
		3° trim	5659	0,044	1,258	-0,677	-0,049	-0,014	0,00000	0,000	0,025	1,143
		4° trim	5654	0,049	1,248	-0,723	-0,043	-0,011	0,00000	0,000	0,032	1,154
		Anno	5661	0,040	1,235	-0,690	-0,062	-0,016	-0,00328	0,000	0,021	1,061
	V A R I A Z I	Giu/Mar	5618	0,015	0,262	-0,215	-0,029	0,000	0,00000	0,002	0,050	0,364
		Set/Giu	5615	0,020	0,647	-0,290	-0,049	-0,003	0,00000	0,000	0,028	0,304
		Dic/Set	5598	0,062	2,312	-0,267	-0,022	0,000	0,00000	0,010	0,063	0,500
		2° trim/1° trim	5650	0,022	0,283	-0,318	-0,027	-0,001	0,00000	0,014	0,045	0,533
		3° trim/2° trim	5655	0,016	0,484	-0,333	-0,038	-0,010	0,00000	0,001	0,027	0,500
	·	4° trim/3° trim	5652	0,046	2,281	-0,344	-0,029	-0,002	0,00000	0,011	0,038	0,500

Tabella 7 - stima dell'errore di definizione per i dipendenti INPS

Riferimento temporale	Stima dell'errore di definizione in media	Stima dell'errore di definizione in mediana
Marzo	2,21%	0,00%
Giugno	2,20%	0,00%
Settembre	2,62%	0,00%
Dicembre	2,15%	0,00%
1° trimestre	2,22%	0,54%
2° trimestre	2,08%	0,71%
3° trimestre	2,04%	0,42%
4° trimestre	1,92%	0,65%
Anno	2,12%	1,06%
Giugno/Marzo	0,27%	0,00%
Settembre/Giugno	0,30%	0,00%
Dicembre/Settembre	-0,30%	0,00%
2° trim./1° trim.	-0,30%	0,00%
3° trim./2° trim.	-0,57%	0,00%
4° trim./3° trim.	-3,54%	0,00%

Riferimento temporale	Sezione di attività	Stima dell'errore di definizione in media	Stima dell'errore di definizione in mediana
ANNO	(c) Estrazione minerali	1,52%	1,03%
	(d) Ind. Manifatturiera	1,95%	1,13%
	(e) Energia elettrica, gas, acqua	1,12%	0,51%
	(g) Commercio	1,79%	0,55%
	(h) Alberghi e ristoranti	<u>7,27%</u>	<u>2,61%</u>
	(i) Trasporti	2,72%	0,93%
	(j) Intermediazione monetaria e fin.	0,70%	0,23%
	(k) Att. Immob., noleggio, infor., ricerca	1,70%	0,85%
	(o) Altri servizi pubblici sociali e personali	<u>3,08%</u>	<u>2,66%</u>
	Totale		2,1%

Riferimento temporale	Dimensione	Stima dell'errore di definizione in media	Stima dell'errore di definizione in mediana
ANNO	<15	2,25%	0,78%
	15 - 50	2,22%	1,14%
	50 - 100	2,05%	1,05%
	> 100	1,69%	1,05%
	Totale		2,1%

Con riferimento alle variazioni, l'errore di definizione risulta nullo in mediana mentre, in media, varia tra il -3.5% e lo 0.27% a seconda del periodo considerato.

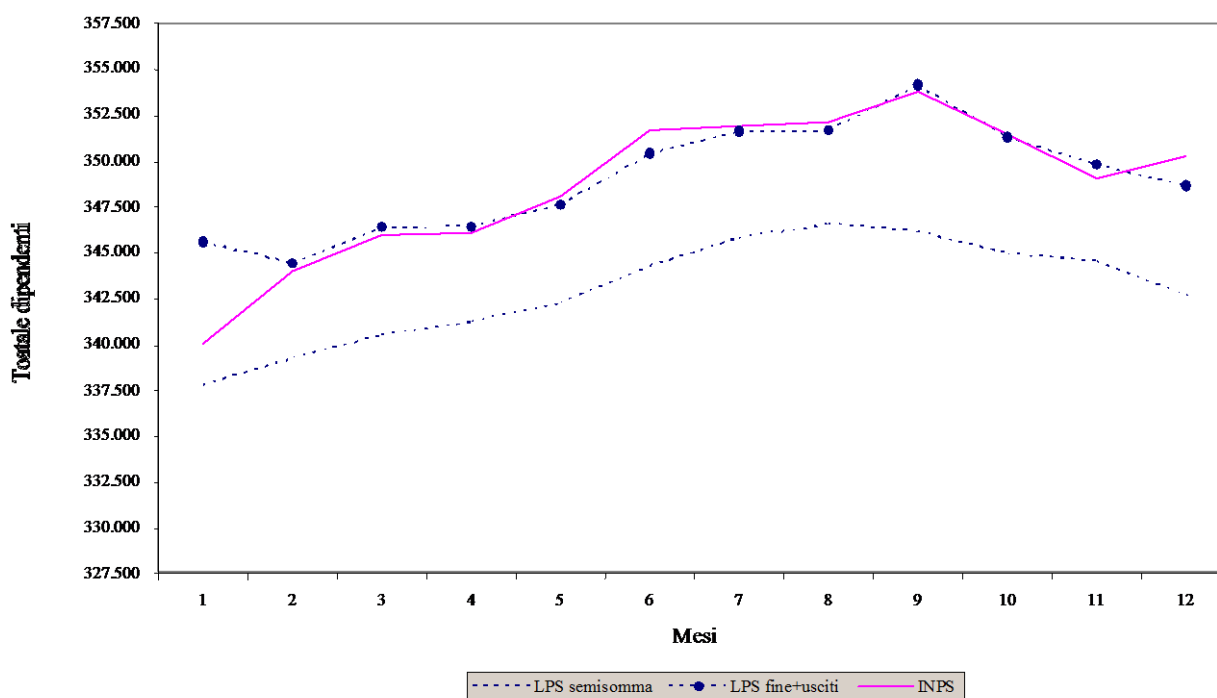
L'analisi per sezione di attività economica e classe dimensionale mostra chiaramente l'effetto del turn over sull'errore di definizione. In particolare, la tabella evidenzia un errore di definizione notevolmente superiore alla media (mediana) per i settori Alberghi e Ristoranti e Altri Servizi Pubblici Sociali e Personali che, come è noto, sono tra i settori a più elevato turn over. L'errore stimato in media risulta intorno al 7% e al 3% rispettivamente per gli alberghi e i servizi; l'errore stimato in mediana risulta per entrambi i settori intorno al 2.6%. Anche l'analisi per classe

dimensionale mostra una tendenza all'aumento dell'errore di definizione con il decrescere della dimensione d'impresa (tipicamente le imprese di piccola dimensione sono quelle a più elevata variabilità occupazionale a breve).

Per cogliere l'andamento nel tempo delle differenze tra i dipendenti di fonte amministrativa e statistica viene effettuata un'analisi "macro" sul campione troncato dell'1% a destra e a sinistra rispetto al bias complessivo medio annuo. In particolare, il grafico 1 riporta la distribuzione dei dipendenti INPS, LPS definizione 1 e LPS definizione 2 per i 12 mesi dell'anno. A conferma di quanto già riscontrato attraverso le analisi precedenti, dal grafico emergono chiaramente i seguenti aspetti:

- per gran parte dell'anno, i dipendenti INPS presentano livelli molto vicini a quelli dei dipendenti LPS ottenuti come somma dei dipendenti a fine mese e degli usciti (definizione 2);
- i dipendenti INPS risultano sistematicamente sovrastimati, nei livelli, rispetto ai dipendenti LPS ottenuti come semisomma dei dipendenti all'inizio e alla fine del mese considerato (definizione 1). Nelle variazioni, tuttavia tali differenze si attenuano notevolmente;
- gennaio e dicembre rimangono i mesi più problematici. In particolare, i dipendenti INPS e LPS definizione 2 presentano, in tali mesi, andamenti completamente opposti.

Grafico1: Analisi macro -Totale dipendenti



Per cogliere l'impatto di alcune caratteristiche strutturali dell'impresa (settore di attività, dimensione, localizzazione territoriale) sul bias, è, infine, condotta un'analisi multivariata. In particolare, sono effettuate analisi di regressione sull'intero campione, sul sotto insieme di imprese uniposizionate e sul campione troncato dell'1% a destra e dell'1% a sinistra.

Si considera un modello in cui la variabile dipendente è data dal bias complessivo. Le variabili esplicative inserite nel modello sono 17 dummy per l'attività economica:

ad esempio, d_da è la dummy per il sottosezione DA (Alimentari). Per questioni di numerosità sono state aggregate le sottosezioni DF, DG, DH; e le sezioni C, E, J, O che quindi rappresentano un settore contenitore residuale. Il settore scelto come benchmark è il Commercio (Sez. G).

4 dummy per la dimensione:

d_dim2, d_dim3 e d_dim4 sono, rispettivamente le dummy per le classi dimensionali 15-50, 50-100 e >100 dipendenti. La classe scelta come benchmark è <15 dipendenti.

5 dummy per la localizzazione territoriale:

ad esempio, d_no è la dummy per la macro regione nord-ovest. La macro regione scelta come benchmark è il centro;

2 dummy per la tipologia dell'impresa con riferimento al numero di posizioni da cui è costituita:

d_npos2 è la dummy per le imprese pluriposizionate. La tipologia d'impresa scelta come benchmark è l'uniposizionata.

Le regressioni presentano un R^2 molto basso. In ogni caso si possono trarre alcune indicazioni provvisorie. Con riferimento alla regressione sull'intero campione solo due settori presentano bias significativamente diversi da zero il Conciario e gli Apparecchi meccanici. I coefficienti corrispondenti indicano che per tali settori il bias risulta maggiore di quello riscontrato per il settore del commercio (scelto come settore di confronto) rispettivamente del 22% e del 49%.

Con riferimento alla dimensione si riscontra un bias significativamente diverso da zero soltanto in corrispondenza dell'ultima classe (>100 dipendenti). In questo caso il bias risulta inferiore di circa il 15% a quello riscontrato per le imprese con meno di 15 dipendenti. Un tale risultato conferma, quindi, quanto emerso dall'analisi univariata per classi dimensionali, cioè che il bias diminuisce al crescere della dimensione presumibilmente per effetto del turn over. Può essere data un'ulteriore spiegazione di tale risultato se si tiene conto che tra le imprese di grande dimensioni si concentra, verosimilmente, la maggior parte delle pluriposizionate per le quali possono esservi problemi di ricostruzione parziale. Tale ipotesi, da un lato, sembra essere confermata dai risultati ottenuti sul campione delle sole imprese uniposizionate (infatti, il coefficiente dell'ultima classe dimensionale si riduce notevolmente, in valore assoluto), dall'altro, contrasta con il bias ottenuto per le imprese pluriposizionate (infatti, tale bias risulta superiore di circa il 39% rispetto al bias delle uniposizionate). Tuttavia, i risultati della regressione sul sotto campione ottenuto eliminando l'1% delle osservazioni ai due estremi della distribuzione, indicano come tale risultato dipenda essenzialmente dalla presenza di outliers. La stessa conclusione è valida per tutte le altre dummy alle quali, nella precedente regressione, sono associati bias molto elevati.

3.4 Retribuzione lorda

3.4.1 Definizione a confronto

La retribuzione cui si riferiscono i dati INPS è la retribuzione imponibile ai fini del calcolo dei contributi di previdenza e assistenza di cui all'art.12 L.153/69. In particolare, la retribuzione imponibile è costituita da "tutto ciò che il lavoratore riceve dal datore di lavoro in denaro o in natura, al lordo di qualsiasi ritenuta, in dipendenza del rapporto di impiego".

Poiché, si fa riferimento a quanto viene erogato dal datore di lavoro a proprio carico, restano escluse (in quanto esenti da contribuzione) le somme corrisposte dal datore per conto di terzi, come, ad esempio, le prestazioni anticipate dal datore di lavoro per conto di gestioni previdenziali e mutualistiche (assegni familiari, indennità in caso di malattia, di infortunio, di gravidanza, assegni per congedo matrimoniale, compenso ai donatori di sangue, ecc.) e le somme percepite dal lavoratore a titolo di mancia.

L'indagine LPS rileva il totale guadagni lordi cioè l'ammontare dei compensi a carattere continuativo e delle gratifiche, indennità e premi a carattere non continuativo erogati dall'azienda nel mese di riferimento al personale dipendente, al lordo delle trattenute erariali e delle ritenute per contributi sociali e assicurativi a carico del lavoratore. Inoltre l'indagine evidenzia come di cui gli importi anticipati per conto degli Istituti di Previdenza e Assistenza (assegni familiari, di maternità, di infortunio, CIG, ecc.).

Dall'analisi delle numerosissime voci retributive considerate nelle indagini Istat emergono non poche differenze concettuali con le voci presenti nella retribuzione imponibile INPS¹¹. Tuttavia, nella pratica, tali differenze dovrebbero risultare trascurabili in primo luogo perché il peso di tali importi sul totale delle retribuzioni risulta molto modesto, in secondo luogo perché le imprese, una volta effettuato il calcolo della retribuzione imponibile ai fini contributivi, tenderanno a riportare il medesimo importo anche sui moduli Istat (senza effettuare nessuna correzione).

L'analisi che segue è stata effettuata con l'obiettivo, quindi, di cogliere la dimensione e soprattutto la natura (sistematica o meno) dello scostamento tra retribuzioni INPS ed LPS.

Si ricorda che, nel caso delle retribuzioni, alle possibili fonti di errore già individuate per i dipendenti si aggiunge la possibilità che si verificano tra le due indagini degli sfasamenti temporali nell'attribuzione di alcuni importi.

3.4.2 Una valutazione della distorsione della retribuzione lorda

Da quanto precedentemente illustrato emerge come per le retribuzioni lorde la componente dell'errore dovuta a differenze definitorie possa considerarsi, nella pratica, trascurabile. Pertanto, per tale variabile si è proceduto esclusivamente alla misura del bias complessivo.

La variabile vera che si intende misurare sulla base dei dati INPS è data dalla retribuzione pro-capite LPS. Tale variabile si ottiene rapportando al monte retributivo LPS i dipendenti LPS ottenuti come semisomma dei dipendenti alla fine del mese precedente e alla fine del mese corrente. La valutazione dell'errore rispetto a tale variabile risulta complicata dalla presenza del denominatore (dipendenti). Il primo passo, quindi, consiste nel misurare la distorsione rispetto al monte retributivo LPS. Tale variabile viene calcolata sottraendo, dal totale guadagni lordi, gli importi anticipati dal datore per conto degli enti di previdenza ed assistenza.

Successivamente, si procede alla valutazione dell'errore rispetto alla retribuzione pro-capite LPS.

Adottando una simbologia del tutto analoga a quella utilizzata nel caso dipendenti, si può scrivere che per la generica impresa i al tempo t , il bias relativo complessivo è dato dall'espressione seguente:

$$(\text{INPS}W_{it} - \text{LPS}W_{it}) / \text{LPS}W_{it}$$

dove con $\text{INPS}W_{it}$ si indica il monte retributivo INPS e con $\text{LPS}W_{it}$ il monte retributivo LPS ottenuto come spiegato. Entrambi i monti retributivi si riferiscono al totale dei dipendenti al netto dei dirigenti. L'esclusione delle retribuzioni corrisposte ai dipendenti con tale qualifica è indispensabile in quanto l'indagine LPS non rileva questa variabile.

I principali risultati dell'analisi effettuata sui bias complessivi di livelli e delle variazioni sono riportati nella tabella 8.

¹¹ Per un confronto dettagliato delle voci incluse ed escluse nelle definizioni INPS e Istat si rimanda a Revelli, 1995 ed Alemanni, 1996

Tabella 8 – analisi univariata sul bias dei monti retributivi

	Riferimento temporale	Numero imprese	Media	Deviazione standard	P1	P10	Q1	Mediana	Q3	P90	P99
LIVELLI	Marzo	5584	0,050	1,1060	-0,6787	-0,0758	-0,00605	0,00007	0,02423	0,07920	1,0961
	Giugno	5569	0,072	1,3078	-0,6594	-0,0709	-0,00513	0,00003	0,02106	0,08017	1,1899
	Settembre	5554	0,060	1,2457	-0,6702	-0,0797	-0,00673	0,00004	0,02132	0,07329	1,1467
	Dicembre	5579	0,084	0,7368	-0,6652	-0,0603	-0,00241	0,00007	0,01599	0,15721	1,8975
	1° trimestre	5613	0,037	1,0969	-0,6717	-0,1271	-0,01333	0,00157	0,02364	0,07259	0,9362
	2° trimestre	5612	0,045	1,1270	-0,6836	-0,0927	-0,01191	0,00096	0,02124	0,07550	0,9535
	3° trimestre	5600	0,046	1,1465	-0,6722	-0,1021	-0,01151	0,00072	0,02048	0,06729	1,1120
	4° trimestre	5627	0,071	1,4205	-0,6861	-0,1091	-0,00964	0,00169	0,02133	0,12559	1,2618
	Anno	5648	0,036	1,0920	-0,7211	-0,1146	-0,01953	0,00209	0,02271	0,08156	0,9314
VARIAZIONI	Giugno/Marzo	5541	0,038	0,4618	-0,3927	-0,0567	-0,01512	0,00000	0,01329	0,07221	1,0164
	Settembre/Giugno	5517	0,020	0,6421	-0,4966	-0,0689	-0,01479	0,00000	0,01331	0,06146	0,7237
	Dicembre/Settembre	5508	0,078	1,4338	-0,4934	-0,0677	-0,01632	0,00000	0,01718	0,15026	1,4630
	2° trim./1° trim.	5597	0,023	0,2893	-0,4072	-0,0551	-0,01510	-0,00005	0,01200	0,07454	0,6978
	3° trim./2° trim.	5577	0,016	0,3288	-0,4748	-0,0641	-0,01484	0,00000	0,01301	0,06117	0,7161
	4° trim./3° trim.	5586	0,079	1,6054	-0,4523	-0,0649	-0,01373	0,00000	0,01681	0,12877	1,0481

La tabella mostra come i monti retributivi INPS risultino superiori, tanto in media che in mediana, a quelli rilevati dall'indagine LPS, nella maggior parte dei casi considerati. Nel caso del bias medio annuo, ad esempio, questo risulta pari al 3.6% in media e al 2% in mediana.

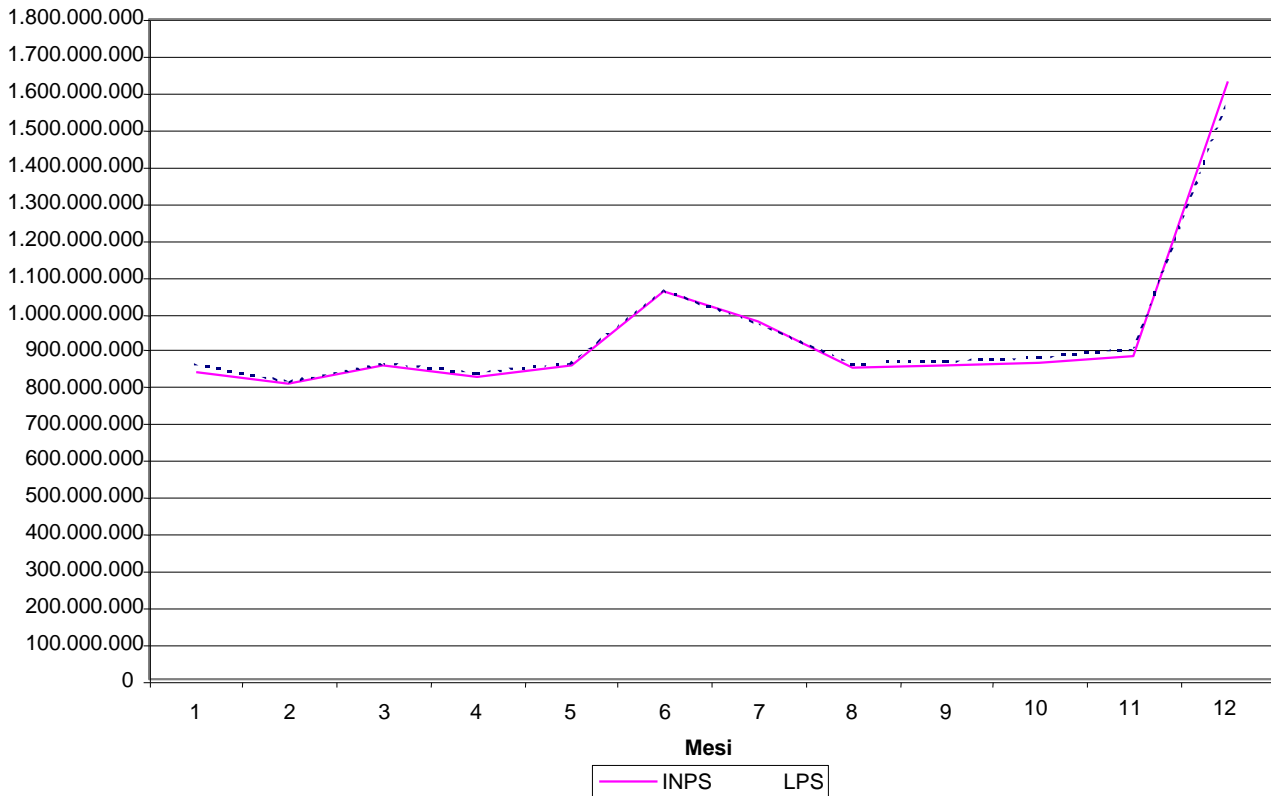
Per i livelli, si riscontrano, bias sistematicamente inferiori ai bias complessivi riscontrati per i dipendenti. D'altra parte un tale risultato era prevedibile se si considera che, per le retribuzioni, l'errore di definizioni può considerarsi, nella pratica, molto contenuto. Una ulteriore conferma è data dal fatto che il bias complessivo retributivo risulta molto simile al bias netto per i dipendenti.

Per le variazioni, viceversa, si ottengono bias di entità maggiore rispetto a quelli dei dipendenti.

Per evidenziare l'andamento temporale dei monti retributivi INPS ed LPS viene effettuata un'analisi "macro" sul campione troncato dell'1% a destra e a sinistra rispetto al bias complessivo medio annuo. In particolare, il grafico 2 riporta la distribuzione del monte retributivo INPS e di quello LPS per i 12 mesi dell'anno. Il grafico mostra chiaramente la presenza di una elevata corrispondenza nei livelli dei monti retributivi rilevati dalle due fonti.

Infine, sulla base di un modello del tutto analogo a quello utilizzato per i dipendenti, sono state effettuate alcune analisi di regressione i cui risultati non si discostano particolarmente da quelli relativi ai dipendenti.

Grafico 2: analisi macro - Monti retributivi



Passiamo, ora alla valutazione della distorsione tra le retribuzioni pro-capite. In simboli, si può scrivere che, per la generica impresa i al tempo t , il bias relativo complessivo è dato dall'espressione seguente:

$$\frac{(INPS_{W_{it}} - LPS_{W_{it}})}{LPS_{W_{it}}}$$

dove con $INPS_{W_{it}}$ si indica la retribuzione pro-capite INPS e con $LPS_{W_{it}}$ la retribuzione pro-capite LPS. Entrambi i pro-capite si riferiscono al totale dei dipendenti al netto dei dirigenti. I principali risultati dell'analisi effettuata sui bias complessivi dei livelli sono riportati nella tabella 9.

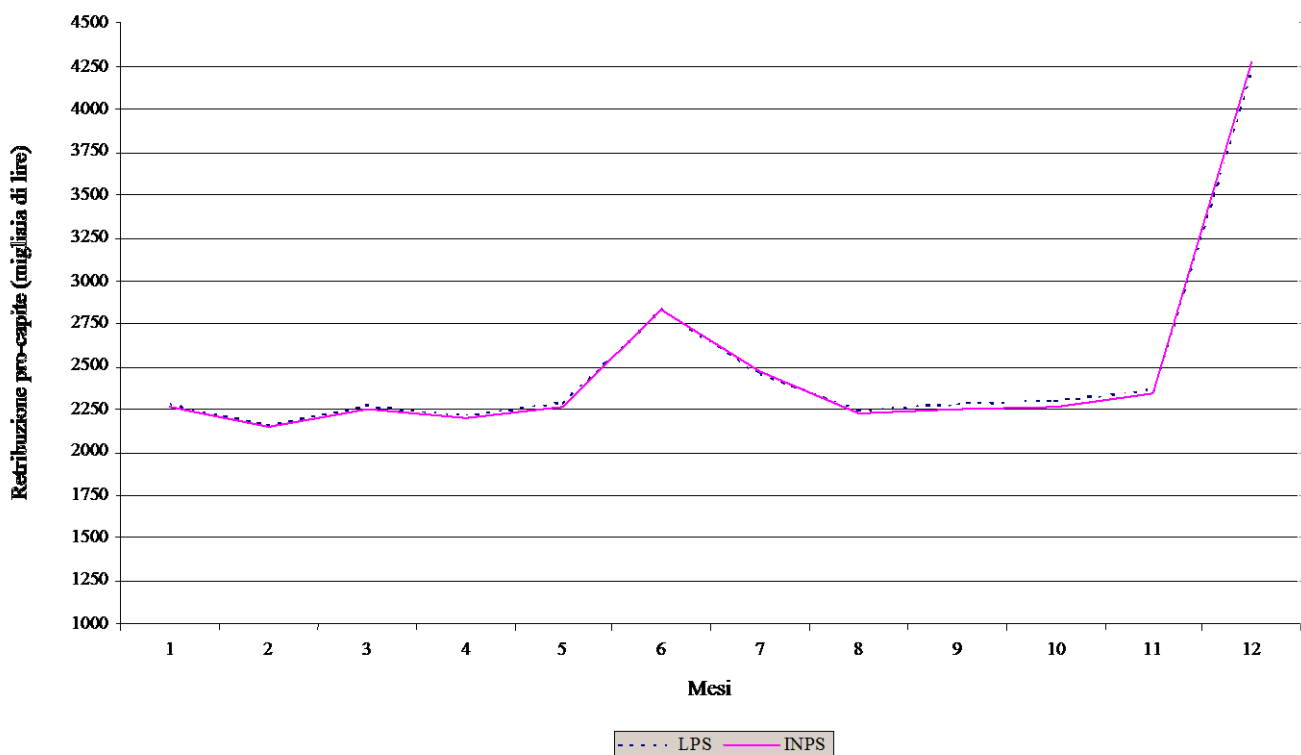
Tabella 9 – analisi univariata sul bias delle retribuzioni pro-capite

Riferimento temporale	Numero imprese	Media	Deviazione standard	P1	P10	Q1	Mediana	Q3	P90	P99	
LIVELLI	Marzo	5635	0,007	0,22097	-0,49515	-0,10701	-0,03122	0,00005	0,03576	0,09511	0,57851
	Giugno	5635	0,021	0,40840	-0,50501	-0,11266	-0,03719	0,00000	0,02883	0,08765	1,00211
	Settembre	5629	0,009	0,28372	-0,50010	-0,11542	-0,03670	0,00000	0,03319	0,08841	0,74117
	Dicembre	5616	0,039	1,26193	-0,56870	-0,11531	-0,03135	0,00000	0,02627	0,13830	1,37655
	1° trimestre	5650	-0,003	0,18323	-0,48088	-0,12374	-0,03177	0,00000	0,03201	0,08409	0,54949
	2° trimestre	5654	0,000	0,18925	-0,49612	-0,11066	-0,03437	-0,00039	0,02671	0,08328	0,64122
	3° trimestre	5656	0,002	0,20825	-0,50844	-0,10867	-0,02992	0,00001	0,02873	0,07966	0,66558
	4° trimestre	5650	0,011	0,26930	-0,53885	-0,12378	-0,03253	0,00000	0,02800	0,11031	0,76477
	Anno	5658	-0,003	0,16665	-0,47540	-0,10358	-0,03280	-0,00294	0,02714	0,07900	0,53021

Nel complesso, la tabella mostra bias nelle retribuzioni pro-capite sensibilmente inferiori a quelli riscontrati per i monti retributivi. Un tale risultato dipende, essenzialmente, dal fatto che la sovrastima INPS dei dipendenti risulta maggiore della sovrastima INPS dei monti retributivi. Tale circostanza, in particolare, determina un'inversione di segno del bias relativo al primo trimestre e all'intero anno, indicando, per tali periodi, una sottostima delle retribuzioni pro-capite INPS rispetto a quelle LPS.

Infine, l'analisi "macro" (grafico 3) evidenzia la presenza di una elevata corrispondenza anche nei livelli delle retribuzioni pro-capite rilevate dalle due fonti.

Grafico 3: Analisi macro - Retribuzioni pro-capite



3.5 Costo del lavoro

3.5.1 Definizioni a confronto

Secondo la definizione data dal SEC il costo del lavoro è costituito dal compenso complessivo, in denaro o in natura, riconosciuto da un datore di lavoro al dipendente quale corrispettivo per il lavoro svolto da quest'ultimo durante il periodo contabile. Esso si compone delle retribuzioni lorde (in denaro e in natura) e da quella parte dei contributi sociali (effettivi e figurativi) che è a carico del datore di lavoro. Secondo tale definizione rientrano a far parte del costo del lavoro anche le quote accantonate per trattamento di fine rapporto e per le provvidenze al personale.

Se, da un lato, i dati INPS consentono in maniera relativamente semplice l'individuazione delle retribuzioni lorde (ricavabili in base ai così detti codici occupazione), dall'altro, non altrettanto immediato risulta il calcolo dei contributi sociali a carico del datore di lavoro. Prima di analizzare le diverse questioni relative all'individuazione dei contributi, occorre premettere che i contributi dovuti all'INPS sono solo una parte del complesso dei contributi che devono essere versati per il

sistema di previdenza ed assistenza sociale. Il sistema previdenziale italiano prevede, infatti, che le assicurazioni legali (obbligatorie) siano gestite da istituti di diritto pubblico (l'INPS, le varie Casse nazionali di previdenza e assistenza, l'INAIL per la gestione assicurativa relativa ad infortuni e malattie professionali) e che l'erogazione di trattamenti contributivi speciali (contribuzione contrattuale volontaria) sia affidata ad altri istituti di diritto privato o di origine contrattuale (PREVINDAI, PREVINDAPI, CASAGIT, ecc.).

Nel modello DM10 viene riportato, per ciascuna categoria contributiva, l'importo complessivo degli oneri sociali, cioè sia la parte a carico del datore di lavoro che quella a carico del lavoratore dipendente¹². Piuttosto difficile risulta cercare di desumere da tale aggregato la sola quota a carico del datore di lavoro, in quanto la maggior parte dei contributi deriva dall'applicazione di aliquote percentuali alla corrispondente retribuzione imponibile. Naturalmente, le caratteristiche impositive non risultano omogenee in quanto dipendono da numerosi fattori come ad esempio le origini storiche differenziate del sistema previdenziale del nostro paese, le scelte politiche sociali ed economiche periodicamente prese, il tipo di attività economica svolta dall'impresa, la dimensione aziendale, il contratto applicato ai lavoratori (apprendistato, CFL, ecc.) ed altro.

Inoltre, l'importo complessivo di tali oneri sociali risulta, spesso, al lordo di benefici contributivi, sgravi e fiscalizzazioni.

D'altra parte nel modello esiste una sezione delle somme a credito del datore di lavoro (quadro D) in cui vengono riportati, tra gli altri, anche le somme a titolo di beneficio e sgravio degli oneri poste a conguaglio nel mese e quindi non necessariamente riferite ai contributi versati nel mese stesso.

Infine, notevoli problemi si pongono anche per l'identificazione delle voci relative al TFR e alle provvidenze per il personale.

L'indagine LPS rileva il totale degli oneri contributivi a carico del datore di lavoro al lordo della eventuale compensazione degli importi anticipati per conto degli Enti previdenziali ma la netto degli sgravi degli oneri.

Inoltre, come già evidenziato, dall'indagine è possibile ricavare le retribuzioni lorde semplicemente detraendo dal totale guadagni gli importi anticipati dal datore per conto degli Enti. Infine, vengono rilevate le indennità di fine rapporto effettivamente corrisposte.

Pertanto, risulta evidente che, mentre l'indagine LPS rileva direttamente tutte le variabili necessarie per il calcolo del costo del lavoro, viceversa, la base informativa desumibile dai modelli DM10 è scarsamente utilizzabile senza l'ausilio di informazioni aggiuntive.

3.5.2 Un esercizio di calcolo del costo del lavoro INPS

In questo lavoro viene svolto un primo esercizio nel quale si procede al calcolo del costo del lavoro per singola denuncia contributiva, cioè per singolo DM10 senza scendere, quindi, nel dettaglio delle qualifiche o del tempo di lavoro dei dipendenti. Inoltre, poiché, come precedentemente spiegato, esiste una grandissima eterogeneità contributiva, si restringe il campo di analisi ad un subset di imprese con caratteristiche tali da rendere più semplice e quindi, teoricamente, più corretta l'identificazione di tale variabile. In primo luogo, per eliminare i possibili problemi connessi con la ricostruzione dell'impresa, si considerano unicamente le imprese costituite da una sola posizione contributiva INPS (questo, naturalmente, non elimina completamente il problema in quanto la procedura in base alla quale si è proceduto al conteggio delle posizioni di ciascuna impresa è ancora provvisoria). In secondo luogo, l'analisi viene ristretta alle imprese senza dirigenti. Tale scelta è dettata dal fatto che entrambe le fonti non forniscono informazioni (LPS) o le

¹² Tali importi sono quelli esposti nel quadro B del modello in corrispondenza dei così detti codici occupazione e contribuzione, esclusi gli importi associati a particolari codici contribuzione che non identificano dei reali contributi.

forniscono soltanto in maniera parziale (INPS) per il calcolo del costo dei dipendenti con tale qualifica. Infine, per ridurre il numero di aliquote contributive da applicare, si considerano, esclusivamente, dati relativi ai primi 9 mesi dell'anno¹³ e alle imprese del Legno, Alimentare, Tessile e Gomma. L'attività economica è individuata facendo riferimento al Codice Statistico Contributivo (CSC) dell'INPS poiché è sulla base di tale codice che si diversificano le aliquote contributive¹⁴.

In tal modo, il sub set sul quale si lavora risulta composto da 811 imprese con le caratteristiche descritte sopra (la distribuzione di tale sub set con riferimento al settore di attività, alla dimensione e alla localizzazione territoriale è riportata nelle tavole A1, A2 e A3 in appendice al capitolo).

Il costo del lavoro INPS si ottiene, quindi, come somma, in valore assoluto, di singoli addendi. In particolare, la determinazione della retribuzione lorda non pone particolari problemi in quanto tale variabile è ottenuta come somma degli importi associati ai codici occupazione. Viceversa, la determinazione degli oneri sociali a carico del datore di lavoro risulta decisamente più complessa.

In particolare, si procede, dapprima, all'individuazione della quota degli oneri a carico del lavoratore e quindi alla loro detrazione dal complesso contributivo (totale importi quadro B)¹⁵. La quota degli oneri a carico del lavoratore si ricava, come accennato, applicando le opportune aliquote contributive alle retribuzioni imponibili corrispondenti. Nel nostro esercizio, poiché si fa riferimento ad un sub set di imprese con le caratteristiche sopra descritte, l'eterogeneità contributiva risulta essere funzione, essenzialmente, della dimensione dell'impresa e della categoria cui il lavoratore appartiene (apprendista, operaio, impiegato, viaggiatore). Per cui l'esercizio consiste nell'individuazione delle diverse retribuzioni imponibili (sulla base dei codici occupazione) e delle diverse aliquote contributive da applicarvi.

Il secondo passo della procedura di calcolo del costo del lavoro riguarda l'individuazione, tra gli importi a credito del datore di lavoro (importi associati ai codici esposti sul quadro D del modello), di quelli corrispondenti a benefici, sgravi e fiscalizzazioni degli oneri sociali. Soltanto tali importi¹⁶, quindi, vengono posti in detrazione dalle altre componenti del costo, calcolate in base alla procedura descritta.

Come precedentemente spiegato, il costo del lavoro ricostruito unicamente sulla base delle informazioni che transitano sul DM10 risulta senz'altro sotto stimato. Pertanto, un'ulteriore passo consiste nell'integrare tale variabile con informazioni aggiuntive. In particolare, in questa fase del lavoro, il costo del lavoro INPS è stato integrato unicamente con i contributi per il Servizio Sanitario Nazionale al lordo delle fiscalizzazioni previste (tale istituto era ancora in vigore nel 1995). Viceversa, non è stata affrontata la stima delle quote INAIL e delle quote accantonate per il TFR e per le provvidenze al personale.

¹³ A partire da ottobre 1995, infatti, le aliquote previste dalla normativa subiscono dei cambiamenti.

¹⁴ Secondo la classificazione in vigore fino al 1996, i CSC che individuano i settori del Legno, Alimentare, Tessile e Gomma sono, rispettivamente, pari a 103, 104, 107 e 109.

¹⁵ E' solo il caso di accennare che la modalità di esposizione degli importi del quadro B del modello DM10 non è univoca. Nella maggior parte dei casi, tali importi si riferiscono al complesso degli oneri sociali a carico del datore e del lavoratore al lordo di benefici e sgravi. In altri casi, tuttavia, le somme riportate nel quadro B possono identificare soltanto gli oneri a carico del lavoratore o soltanto del datore di lavoro; in questo caso, inoltre, tali oneri risultano al netto di benefici e sgravi in quanto costituiscono un'ulteriore esposizione cui già corrisponde, nel modello, un'esposizione di tipo generale. Si, sottolinea, infine, che la metodologia seguita per l'individuazione dei contributi a carico del datore sembra essere corretta indipendentemente dal modo con cui avviene l'esposizione degli importi del quadro B del modello.

¹⁶ Altri importi del quadro D del modello, come ad esempio la restituzione degli importi anticipati dal datore per conto dell'INPS, non costituiscono elementi della contribuzione e quindi non devono essere posti in detrazione.

3.5.3 Una valutazione della distorsione del costo del lavoro

La variabile scelta come benchmark per valutare la distorsione del costo del lavoro di fonte INPS è costituita dal costo del lavoro LPS definito come la somma della retribuzione lorda (totale guadagni lordi al netto degli importi anticipati dal datore per conto degli enti di previdenza ed assistenza) e del totale degli oneri contributivi a carico del datore di lavoro al lordo della eventuale compensazione degli importi anticipati per conto degli Enti previdenziali ma al netto degli sgravi degli oneri. Nel computo di tale variabile si è deciso di tralasciare l'addendo relativo alle quote accantonate per il TFR e per le provvidenze al personale per ridurre al minimo le differenze di definizione tra le variabili delle due fonti.

Quindi, in modo del tutto analogo a quanto effettuato nel caso delle altre variabili viene effettuata un'analisi sul bias complessivo.

Adottando una simbologia del tutto simile a quella utilizzata nei casi precedenti, si può scrivere che, per la generica impresa i al tempo t , il bias è dato dalla seguente espressione:

$$(\text{INPS}C_{it} - \text{LPS}C_{it}) / \text{LPS}C_{it}$$

dove con $\text{INPS}C_{it}$ si indica il costo del lavoro INPS e con $\text{LPS}C_{it}$ il costo del lavoro LPS.

Un'ulteriore analisi viene effettuata sui bias relativi delle due componenti del costo: retribuzioni e contributi a carico del datore di lavoro. In tutti i casi, si considerano dati ottenuti come media dei primi 9 mesi del 1995. I principali risultati di tali analisi sono riportati nella tabella 10.

Tabella 10 - analisi univariata sul bias del costo del lavoro, del monte retributivo e dei contributi a carico del datore di lavoro

Variabile	Numero imprese	Media	Deviazione standard	P1	P10	Q1	Mediana	Q3	P90	P99
BIAS RETRIBUTIVO	811	0,0108	0,1127	-0,2193	-0,0383	-0,0091	0,0034	0,0213	0,0504	0,4658
BIAS CONTRIBUTIVO	807	-0,0857	0,3658	-1,3028	-0,3730	-0,1793	-0,0686	0,0303	0,1856	0,8864
BIAS DEL COSTO DEL LAVORO	811	-0,0214	0,1256	-0,4031	-0,1183	-0,0520	-0,0155	0,0166	0,0515	0,3832

La tabella mostra come il bias del costo risulti tanto in media che in mediana piuttosto contenuto. In particolare, per i motivi precedentemente esposti, il costo INPS risulta leggermente sotto stimato rispetto a quello LPS.

Dall'analisi del bias retributivo e di quello contributivo, appare chiaramente che la sottostima del costo è dovuta ad una sottostima degli oneri contributi a carico del datore di lavoro di entità tutt'altro che trascurabile. In particolare, l'errore si attesta, in valore assoluto, intorno all'8.5% in media e intorno al 7% in mediana; inoltre, almeno il 25% delle osservazioni presentano un bias superiore, in valore assoluto, al 18%. Al momento non è del tutto chiara l'origine di una distorsione di tale entità. L'introduzione delle quote INAIL¹⁷ potrebbe consentire la riduzione di parte dell'errore. E' comunque auspicabile un esame più approfondito delle altre possibili fonti di distorsione. In particolare, occorre appurare se:

1. nella fase di scomposizione del complesso contributivo (totale importi del quadro B del DM10) si sovrastima la quota a carico del lavoratore;
2. l'informazione con cui integrare i dati INPS riguarda, oltre i contributi INAIL, altre forme di

¹⁷ Le aliquote medie INAIL previste per gli operai sono, approssimativamente, pari al: 6,5% per gli Alimentari, 3,6 % per il Tessile, 6,9% per il Legno, 3,6% per la Gomma e Plastica. (nostre ipotesi su aliquote Inail)

contribuzione contrattuale o volontaria;

3. esiste uno sfasamento temporale tra il mese in cui avviene l'anticipazione ed il mese in cui avviene la compensazione, e se, in particolare, tale compensazione ha luogo alla fine dell'anno (in tal caso, infatti, si sovrastimerebbe la variabile contributi sociali al lordo della composizione degli importi anticipati di fonte LPS).

Infine, i risultati della regressione condotta sul bias contributivo, evidenziano la presenza di errori specifici per il settore del legno, non a caso, il settore che presenta l'aliquota INAIL più elevata (6,9%).

3.6 Conclusioni

L'obiettivo del capitolo è stato quello di studiare le distorsioni delle variabili ricavabili dal modulo DM10 dell'INPS. Esse sono intese come una possibile misurazione di ignote variabili vere. Per approssimare queste ultime si è usata l'indagine LPS001 riferita al 1995. Le analisi svolte quindi si fondano sull'ipotesi che la fonte usata come benchmark sia affidabile. Su tale questione è bene avanzare le dovute cautele.

I risultati ottenuti evidenziano come la maggior parte dei problemi convenzionali legati al contenuto delle variabili di fonte INPS, risulta nella pratica di entità accettabile. In, particolare:

1. per tutte le variabili considerate esistono differenze formali tra definizioni amministrative e statistiche. Nonostante questo nei dati le differenze riconducibili a problemi di definizione sono di modesta entità. In particolare, in media per i dipendenti tale errore si attesta intorno al 2%. Esso appare almeno in parte dipendere dal turn-over occupazionale.
2. Tutte le variabili presentano dei bias complessivi dei livelli praticamente nulli in mediana. In media tali bias indicano la presenza di una sovrastima INPS tanto dei dipendenti (pari a circa il 6%) tanto dei monti retributivi (pari a circa il 3,6%). Viceversa, nel caso del costo del lavoro, il bias medio segnala una sottostima INPS di circa il 2%. Esso è imputabile ad una sottostima di circa l'8% dei contributi a carico del datore di lavoro. L'integrazione di tale variabile con le quote contributive INAIL dovrebbe riportare il bias sui valori riscontrati per le altre variabili;
3. Le variazioni congiunturali risultano previste dai dati INPS con un errore accettabile. Qualche dubbio rimane sulla stima delle variazioni dei periodi finali dell'anno. in tutti i casi eccetto che alla fine dell'anno.
4. La procedura di riaggregazione dei codici presenti sul DM10 su cui si basa tanto il calcolo dei dipendenti che quello dei monti retributivi non sembra essere affetta da errori;
5. L'analisi multivariata evidenzia due punti principali. In primo luogo, il bias complessivo, sia per i dipendenti che per i monti retributivi, è maggiore per le imprese con più posizioni contributive. In secondo luogo, specialmente per i monti retributivi, il bias presenta una caratterizzazione settoriale,. Appaiono maggiormente problematici i settori del tessile, del conciario, dei Trasporti.

APPENDICE AL CAPITOLO 3

SOTTO INSIEME DI IMPRESE INPS e LPS USATO PER L'ANALISI SUL COSTO DEL LAVORO - ANNO 1995

TABELLA A1 - DISTRIBUZIONE PER SEZIONE-DIVISIONE

SEZIONE-DIVISIONE		Frequency	Percent
da	Alimentari	248	30,6
db	Tessili	114	14,1
dc	Conciarie	11	1,4
dd	Legno	149	18,4
de	Carta	1	0,1
df	Coke, petrolio	3	0,4
dg	Prodotti chimici	34	4,2
dh	Gomma, plastica	98	12,1
di	Minerali non metalliferi	3	0,4
dj	Metallo	8	1,0
dk	Apparecchi meccanici	1	0,1
dl	Macchine elettriche	4	0,5
dm	Mezzi di trasporto	4	0,5
dn	Altre industrie manifatturiere	115	14,2
g	Commercio	12	1,5
h	Alberghi	1	0,1
i	Trasporti	2	0,2
k	Attività immobiliari	3	0,4
	Totale	811	100

TABELLA A2 - DISTRIBUZIONE PER DIMENSIONE

CLASSE DIMENSIONALE	Frequency	Percent
0- 15	194	23,9
15-50	460	56,7
50-100	121	14,9
>100	36	4,4
Totale	811	100

TABELLA A3 - DISTRIBUZIONE PER MACRO REGIONE

MACRO REGIONE	Frequency	Percent
Nord-est	186	22,9
Nord-ovest	210	25,9
Centro	198	24,4
Sud	172	21,2
Isole	45	5,5
Totale	811	100

4 ALCUNE VALUTAZIONI SULLA QUALITÀ DELLA CODIFICA DELL'ATTIVITÀ ECONOMICA NEGLI ARCHIVI AMMINISTRATIVI

4.1 La codifica del settore di attività economica negli archivi INPS

All'atto dell'iscrizione di un'azienda presso la sede provinciale INPS di competenza, alla posizione contributiva neo nata viene attribuito un codice di classificazione denominato "CSC" (codice statistico contributivo). Tale codice ha la funzione di fornire informazioni sul tipo di attività economica svolta dall'azienda, ai fini dell'applicazione delle norme vigenti in tema di obblighi contributivi. Come tale, il CSC non ha una immediata corrispondenza con i codici convenzionali delle classificazioni per branche di attività economica utilizzate dall'istat (attualmente ATECO 91 che corrisponde alla classificazione internazionale NACE Rev. 1). Nelle analisi correnti finalizzate alla produzione di indicatori su occupazione, retribuzioni e costo del lavoro sulla base degli archivi amministrativi dell'INPS, il settore di attività economica rappresenta, tuttavia, una delle principali variabili d'interesse. Al fine di dedurre dai dati INPS informazione sui settori di attività economica assimilabili ai criteri classificatori nazionali e internazionali, si rendono dunque necessarie una serie di operazioni di trascodifica del CSC. A tal proposito, un supporto operativo di fondamentale importanza è fornito dalle tabelle ufficiali messe a disposizione dall'INPS agli operatori incaricati all'iscrizione delle posizioni contributive. In tali tavole, accanto alla descrizione dettagliata delle branche di attività economica, è riportato il corrispondente codice CSC. Il codice ATECO può, in breve, essere dedotto da tale descrizione. Tuttavia, all'aumentare del numero di cifre convertite per ogni codice statistico contributivo crescono sia problemi di carattere formale (il CSC non permette di identificare alcuni settori definiti dalla classificazione ATECO) che sostanziale (imprecisa definizione di alcune attività, incentivo alla cattiva dichiarazione di attività da parte delle imprese per usufruire di benefici legislativi, uso indiscriminato di codici residui, lacune nella tassonomia etc.). Tali problematiche, in alcuni casi, non consentono di identificare univocamente o correttamente i codici istat, con il rischio di commettere errori di attribuzione o mancate imputazioni.

In realtà, con decorrenza dal gennaio 1996, vige l'obbligo da parte delle sedi INPS di affiancare alla classificazione ufficiale dei datori di lavoro tradizionalmente basata su criteri amministrativi, la corrispondente classificazione di attività economica attenendosi ai parametri ISTAT (ATECO91). L'applicazione della direttiva avviene in risposta alle indicazioni fornite nell'ambito del Sistema Statistico Nazionale, secondo cui l'armonizzazione delle diverse fonti di dati deve divenire un obiettivo fondamentale ai fini della fornitura dell'informazione statistica. Al fine di perseguire la prevista armonizzazione, l'INPS ha provveduto sia mediante l'adozione di una uniforme nomenclatura delle attività, sia con la memorizzazione dei codici ISTAT negli archivi aziende. Le indicazioni fornite per completare la classificazione delle posizioni contributive con la nuova informazione, prevedono che l'individuazione ed attribuzione dei codici ATECO91 debba avvenire consultando le descrizioni delle varie attività economiche contenute in un manuale operativo appositamente divulgato. Nei casi in cui l'immediata individuazione del codice ISTAT sulla base dell'attuale codifica non sia possibile, il provvedimento adottato indica la possibilità di ricorrere all'esame del fascicolo aziendale. In breve, l'attribuzione dei nuovi codici deve seguire delle procedure rigorose che consentano di limitare l'ambiguità nella definizione della conformazione dell'azienda.

La disponibilità della variabile ATECO 91 dagli archivi INPS rappresenterebbe un importantissimo valore aggiunto in termini informativi ed un primo vantaggio consiste nella possibilità di evitare laboriose e complesse operazioni di trascodifica dall'esito non sempre certo. L'inserimento della variabile, attualmente, è tuttavia in fase di transizione. Mentre, infatti, per le posizioni nate successivamente al gennaio 1996 tale informazione è immediatamente immessa negli archivi di

anagrafica, non è del tutto chiaro come e con quale gradualità l'ATECO verrà attribuito anche alle posizioni nate in precedenza all'entrata in regime della disposizione.

Nel presente capitolo vengono riportati alcuni risultati sintetici di un'analisi volta a valutare la qualità della classificazione ATECO nei dati di fonte INPS, prima procedendo ad esaminare i risultati della trascodifica del CSC e, in seguito, accertando la validità dell'informazione traendola direttamente dalla nuova variabile presente negli archivi. In tal senso, l'obiettivo è quello di fornire un ulteriore elemento per trarre delle conclusioni sull'eventualità di ricorrere ad un suo diretto uso nelle analisi correnti. Come benchmark di confronto viene considerato l'Archivio ASIA riferito alle imprese attive nel 1996 (con informazioni anagrafiche aggiornate al giugno 1999), in cui l'attribuzione dei codici di attività economica risulta dall'integrazione di dati desunti da più fonti amministrative (Finanze, CCIAA, INPS, Inail, Enel) e statistiche (Sirio-Nai, indagini strutturali sulle imprese).

I confronti vengono effettuati abbinando per codice fiscale le imprese di ASIA con l'universo dei DM10 dell'INPS relativo al mese di dicembre 1996 (data di aggiornamento delle variabili occupazionali dell'archivio ASIA di giugno 1999) e, in seguito, con le posizioni contributive risultanti dal campione dei DM10 per il mese di gennaio 1999.

4.2 Alcune considerazioni generali

Una serie di controlli generali consentono di trarre delle prime indicazioni sulla qualità della variabile ATECO91 negli archivi INPS. Una prima verifica, a tal proposito, viene realizzata sull'anagrafica delle posizioni contributive risultanti iscritte presso l'INPS nel mese di giugno 1999 (AZISTAT). L'analisi evidenzia che anche per le aziende già in vita alla data di diffusione della direttiva la nuova variabile, in qualche modo, è stata attribuita. Di fatto non sono presenti valori di missing nel campo. Tuttavia, dall'analisi emerge la presenza di codici "atipici", da cui non è possibile risalire direttamente ad alcun codice di attività economica senza avvalersi di informazioni ausiliarie:

- "99999", attribuito sempre in connessione al CSC 70601, che nel manuale operativo INPS corrisponde alla categoria economica "proprietari di fabbricati" e che formalmente non è previsto in alcun codice ATECO91 (anche se vi è una divisione pari a 99 equivalente alla sezione Q). L'analisi micro sulle ragioni sociali dei casi caratterizzati da tale codifica evidenzia prevalentemente attività svolte in condomini o attribuite a società immobiliari. Di fatto, in seguito a informazioni dirette acquisite presso degli operatori INPS, si è avuta la conferma che tale codice viene attribuito nei casi in cui il datore di lavoro diretto è il condominio. In una prima fase di analisi questo caso potrebbe essere risolto attribuendo il codice ATECO "70320" ossia "Amministrazione e gestione di beni immobili per conto terzi", che sembra racchiudere la descrizione del tipo di attività in questione;
- "00000", a cui corrispondono i più disparati codici CSC. Tuttavia, in seguito ad un'analisi condotta anche in relazione alla data di costituzione dell'azienda, si è osservato che tali casi sono concentrati, in termini relativi, prevalentemente in riferimento ad aziende costituite prima del gennaio 1997 evidenziando, dunque, casi in cui l'attribuzione dell'ATECO non è stata ancora effettuata, come informazioni dirette acquisite presso operatori INPS hanno in seguito confermato. Il recupero di tali casi potrebbe basarsi su due strategie gerarchiche. In particolare, l'esperienza sui vari archivi di dati INPS suggerisce, in primo luogo, di acquisire il codice ATECO delle aziende problematiche dagli archivi più ravvicinati da cui la variabile è deducibile. In secondo luogo procedere, nei casi in cui non è stato possibile recuperare la variabile, con la trascodifica del CSC. Tale procedura dovrebbe tuttavia essere applicata solo per i casi in cui viene generata una soluzione non univoca.

4.3 Analisi nell'universo DM10 di dicembre 1996

Escludendo dall'analisi i codici della variabile AZISTAT risultati "atipici", un primo confronto è basato sull'universo delle posizioni contributive in vita a dicembre 1996. Tale universo corrisponde all'insieme delle aziende abbinato con le imprese dell'archivio ASIA con variabili occupazionali aggiornate a dicembre 1996. Rispetto alle imprese presenti nell'archivio ASIA ed in vita nello stesso periodo, si osserva una coincidenza per sezione nel 91% dei casi abbinati, per una quota pari al 90% degli occupati INPS. La corrispondenza per divisione scende all'82%, equivalente al 79% degli occupati rilevati dall'INPS (tav.1). Rispetto alla variabile ottenuta mediante la trascodifica del CSC, la corrispondenza per sezione supera il 94%, assorbendo il 93% degli occupati rilevati nell'archivio. La quota di abbinato per divisione scende al 92%, con il 91% degli occupati. Il gap nelle corrispondenze rilevato dal confronto con i due archivi in parte può essere imputato al fatto che diverse sono le unità di analisi alla base delle comparazioni. L'archivio ASIA, infatti, contiene imprese, mentre le unità di analisi di fonte INPS utilizzate in questo esercizio sono posizioni contributive. Più posizioni contributive possono dare luogo ad un'impresa, pur svolgendo attività economiche classificabili in diversi settori. Questo fenomeno, come rilevato dall'analisi dei dati, è ovviamente più rilevante a livello di divisione, in cui il dettaglio nella classificazione dell'attività economica è più ampio.

Tav. 1 – Corrispondenza dei codici di attività economica ATECO91 presenti nell'universo dei DM10 con quelli ottenuti da trascodifica del CSC (presenti nello stesso archivio) e con i codici tratti dall'archivio ASIA 96 per posizioni e relativi occupati – valori assoluti e percentuale

Coerenza codici di attività economica	Posizioni		Occupati	
	n.	%	n.	%
<i>ATECO91 DM10 - ATECO91 ASIA ' 96</i>				
Sezione	965.957	90,6	7.336.040	90,1
Divisione	876.343	82,2	639.4924	78,6
<i>ATECO91 DM10 - Trascodifica CSC DM10</i>				
Sezione	1.003.067	94,1	7.566.772	92,9
Divisione	903.178	92,3	6.683.588	91,2

Estendendo l'analisi delle coincidenze a livello di singole sezioni si osservano alcune particolarità. La corrispondenza supera il 95% delle posizioni abbinato rispetto ai CSC trascodificati per le sezioni C-H, con un picco in corrispondenza della sezione F che assorbe il 99,9% delle posizioni ed il 99,8% degli occupati. In prossimità dell'80% è comunque la coincidenza per le sezioni residue. Mediamente inferiori appaiono le coincidenze rispetto alle sezioni dedotte dall'archivio ASIA, prossime al 90% per le sezioni D, F-K. Per la sezione C secondo ASIA, ad una quota equivalente al 72% di posizioni abbinato, corrisponde una percentuale pari a poco meno del 57% di occupati. Nella sezione C tradizionalmente è presente un numero limitato di imprese di grandi dimensioni, per cui la missclassificazione di anche una sola azienda può influenzare notevolmente il risultato complessivo.

Tav. 2 - Corrispondenza delle sezioni di attività economica ATECO91 presenti nell'universo dei DM10 con le sezioni ottenute da trascodifica del CSC (presenti nello stesso archivio) e con i codici tratti dall'archivio ASIA 96 – valori in percentuale

Sezione	ASIA 12/96		Trascodifica CSC	
	% posizioni	% occupati	% posizioni	% occupati
C	72,5	56,9	96,8	82,3
D	88,3	93,1	96,6	97,5
E	85,0	98,5	98,3	99,5
F	95,4	91,8	99,9	99,8
G	92,5	90,7	98,2	97,9
H	93,1	91,6	97,2	83,4
I	88,9	76,8	77,1	86,9
J	88,3	96,5	76,7	90,9
K	88,5	86,1	82,5	80,8

4.4 Analisi nel campione DM10 di gennaio 1999

Una speculare analisi viene condotta sul campione dei DM10 relativo al mese di gennaio 1999. Il confronto con le imprese ASIA evidenzia corrispondenza tra i due archivi nell'85% dei casi rispetto alla sezione e al 79% dei casi per divisione (tav.3). Inoltre, l'80% degli occupati relativi alle aziende presenti in entrambe gli archivi sono classificati con la stessa sezione, quota che sale all'83% nel dettaglio della divisione. Delle posizioni abbinate la coincidenza per sezioni supera il 90% per i settori G, H, I ed F (tav.4). In quest'ultima, in particolare, la quota di posizioni con sezione coincidente raggiunge il 95%. Tra 88 ed 89 punti percentuali si attesta la quota di coincidenze nelle sezioni D, J e K. Nella sezione D la quota equivale a 88% e, infine, è pari a 75% nella sezione C. Le posizioni residue rispetto alla sezione C, in particolare, nell'archivio ASIA sono distribuite per il 14% nella sezione D e per un ulteriore 7% in F.

Il secondo confronto viene effettuato rispetto all'universo delle dichiarazioni contributive aggiornato al mese di febbraio 1999. Si osserva, a tal proposito, come i due archivi siano corrispondenti per sezione nel 95% delle posizioni abbinate, equivalenti a poco meno del 99% dei relativi occupati (tav.3). Rispetto alla divisione, la quota di posizioni coincidenti equivale al 93% dei casi, assorbendo il 98% circa degli occupati. Una breve analisi per singola sezione evidenzia una coincidenza completa per i settori E, F, I, e K mentre superiore al 99% è la quota di corrispondenti nelle sezioni D G e H e comunque equivalente a 97% nella sezione J (tav.4). Nella sezione C, invece, la percentuale di posizioni con sezione equivalente nei due archivi è pari al 65% e, un'analisi più approfondita, ha evidenziato che il restante 35% è classificato nel settore K. L'elevato livello di coincidenza tra questi ultimi due archivi sembra ancora una volta confermare che l'assegnazione della variabile AZISTAT viene effettuata attenendosi effettivamente alle tabelle ufficiali di trascodifica divulgate dall'INPS.

Tav. 3 – Corrispondenza dei codici di attività economica ATECO91 presenti nell'campione gennaio'99 dei DM10 con quelli ottenuti da trascodifica del CSC (presenti nello stesso archivio) e con i codici tratti dall'archivio ASIA 96 per posizioni e relativi occupati – valori assoluti e percentuale

Coerenza codici di attività economica	Posizioni		Occupati	
	n.	%	n.	%
<i>ATECO91 DM10 - ATECO91 ASIA' 96</i>				
Sezione	148.224	85,3	1.634.108	80,9
Divisione	140.280	79,1	1.509.671	83,2
<i>ATECO91 DM10 - Trascodifica CSC DM10</i>				
Sezione	151.215	95,2	1.438.152	98,8
Divisione	150.628	93,4	1.434.059	97,7

Tav. 4 - Corrispondenza delle sezioni di attività economica ATECO91 presenti campione gennaio'99 dei DM10 con le sezioni ottenute da trascodifica del CSC (presenti nello stesso archivio) e con i codici tratti dall'archivio ASIA 96 – valori in percentuale

Sezione	ASIA 12/96		Trascodifica CSC	
	% posizioni	% occupati	% posizioni	% occupati
C	73,9	33,24	97,8	65,8
D	88,1	96,1	98,8	99,7
E	81,6	98,1	100	100
F	95,0	90,9	100	100
G	91,9	90,0	99,6	99,7
H	91,6	74,0	99,7	99,6
I	90,3	85,8	100	100
J	87,9	90,8	60	97,4
K	88,8	86,0	100	100

4.5 Conclusioni

Le analisi svolte consentono di trarre alcune indicazioni di massima sulla qualità della variabile AZISTAT e sulla possibilità di farne diretto uso nella costruzione degli indici d'interesse. In particolare, si è osservato che la potenzialità informativa della variabile allo stato originario sia preponderante e molto limitata risulta la probabilità di commettere errore attribuendo alle aziende l'attività economica senza ricorrere a complesse operazioni di trascodifica del CSC. In particolare, ciò che sembra emergere con chiarezza dall'analisi è che l'attribuzione dell'ATECO negli archivi INPS corrisponde, con un certo grado di sistematicità, al CSC. Ciò lascia presumere che gli operatori preposti all'inserimento delle informazioni sull'attività economica consultino effettivamente le guide che consentono di identificare l'ATECO91 corrispondente al codice statistico contributivo. Tuttavia, poiché non sempre vi è corrispondenza univoca tra CSC e codice ATECO91 ci si deve attendere che, inevitabilmente, l'assegnazione del codice di attività economica segua, in mancanza di ulteriori informazioni, una valutazione personale da parte dell'addetto. In

quei casi, d'altra parte, la trascodifica del CSC non consentiva di giungere all'identificazione di un corrispondente codice di attività economica (convenzionalmente questi casi venivano codificati come "Z").

5 LA RICOSTRUZIONE DELLE IMPRESE A PARTIRE DALLE POSIZIONI CONTRIBUTIVE DEGLI ARCHIVI INPS.

5.1 Introduzione

Obiettivo di questo capitolo è illustrare la metodologia studiata e le procedure messe a punto per ricostruire l'impresa a partire dalle unità elementari in base alle quali sono organizzati gli archivi INPS, le *posizioni contributive*.

La posizione contributiva non sempre coincide con l'impresa; infatti un'impresa può aprire presso l'INPS più posizioni contributive.

La ricostruzione dell'impresa è un passo ritenuto necessario al fine di ottenere una stima degli addetti dipendenti, delle retribuzioni e del costo del lavoro ad essi associate, utilizzando le informazioni contenute negli archivi delle posizioni contributive dell'INPS. Alla base di tale scelta vi sono due considerazioni:

1. la procedura di stima statistica degli addetti, delle retribuzioni e del costo del lavoro richiede una precisa definizione dell'unità di analisi e l'unità elementare contenuta negli archivi dell'INPS non soddisfa tale requisito;
2. vi potrebbe essere la necessità di attribuire caratteri (mancanti o errati) alle unità sulla base di archivi (ASIA) in cui l'unità elementare è costituita dall'impresa.

I principali problemi che sono stati affrontati nello studiare una metodologia per la ricostruzione delle imprese sono:

- la natura delle unità elementari presenti negli archivi di fonte INPS, - una posizione contributiva può fare riferimento sia a singole unità locali o a unità funzionali dell'impresa, sia a qualifiche degli occupati dipendenti (operai, impiegati, etc.) - e l'aggiornamento delle variabili in essi contenute;
- la scelta delle chiavi di aggancio e la definizione delle regole che consentano di individuare le unità 'impresa con dipendenti' mediante opportune aggregazioni delle posizioni contributive (regole di *link*);
- l'assegnazione dei principali caratteri di tipo anagrafico e strutturale (attività economica, addetti, etc.) alle nuove unità individuate.

Il capitolo è articolato come segue: nel paragrafo 2 si descrive la situazione informativa su cui si basano le procedure per l'individuazione delle imprese; nel paragrafo 3 si illustrano le caratteristiche delle principali chiavi di aggancio che possono essere utilizzate per aggregare le posizioni contributive; nel paragrafo 4 viene delineata la metodologia seguita per la ricostruzione delle imprese e vengono esposti i principali risultati di un esercizio di ricostruzione con riferimento al Dicembre 1996; nel paragrafo 5 vengono presentati i risultati di una prima valutazione dell'attendibilità della procedura di ricostruzione, basata su confronti effettuati con informazioni tratte dall'archivio ASIA.

5.2 La situazione informativa

Le fonti informative su cui si basano le procedure per la ricostruzione dell'impresa sono costituite da due principali archivi amministrativi dell'INPS:

- *l'archivio anagrafico delle posizioni contributive*. Tale archivio contiene le informazioni che le imprese forniscono all'INPS al momento dell'iscrizione (moduli DM68). In particolare sono presenti variabili di tipo anagrafico (codice fiscale, ragione sociale, data di costituzione etc.), territoriale (provincia, indirizzo etc.) e strutturale (attività economica, numero di dipendenti al momento dell'iscrizione nell'archivio). Si noti che alcune variabili, quali ad esempio il codice

fiscale e la ragione sociale, si riferiscono all'impresa mentre altre, quali la data di costituzione o l'attività economica, sono relative alla posizione contributiva. Inoltre le posizioni contributive non hanno sempre la stessa natura, potendo far riferimento ora ad una unità locale dell'impresa, ora ad una tipologia di occupati dipendenti, a seconda di come l'impresa decida di organizzare i propri versamenti contributivi all'INPS;

- *l'archivio delle dichiarazioni contributive*, fornite ogni mese dalle imprese all'INPS (moduli DM10). Da questo archivio sono deducibili le variabili strutturali relative al numero di occupati dipendenti, alle retribuzioni totali e agli oneri contributivi, distinte per tipologia di occupati (operai, impiegati, dirigenti e apprendisti). Tali variabili sono riferite alla posizione contributiva.

I due archivi differiscono sia per il grado di copertura dell'universo di riferimento (l'universo delle imprese con dipendenti), sia per l'aggiornamento delle informazioni in essi contenute. In particolare, l'archivio anagrafico, contenente le posizioni contributive *formalmente* aperte presso l'INPS, è aggiornato in tempo reale, con una buona precisione per quanto riguarda l'apertura di nuove posizioni contributive e qualche ritardo per quanto riguarda la registrazione delle cessazioni. Non contiene, tuttavia, informazioni aggiornate per quanto attiene le variabili di analisi, quali il numero di addetti. Inoltre, dato il meccanismo di aggiornamento delle informazioni, quando si hanno variazioni non accompagnate da movimenti demografici delle posizioni contributive, come ad esempio quando muta l'attività economica, oppure in presenza di alcuni particolari tipi di trasformazione giuridica, si perde l'informazione relativa al periodo di riferimento, con la conseguenza di attribuire alla posizione contributiva caratteri riferiti al periodo corrente in cui l'informazione è aggiornata¹⁸.

L'archivio delle dichiarazioni contributive, viceversa, contiene informazioni più affidabili sulle variabili di analisi, ma presenta una copertura dell'universo delle imprese con dipendenti abbastanza completa solo relativamente ad un periodo di riferimento¹⁹ che precede di circa dodici mesi il periodo corrente.

Nella tabella 1 sono elencati gli effetti, sull'archivio anagrafico, delle principali trasformazioni giuridiche, delle fusioni e degli scorpori.

¹⁸ Si utilizza l'archivio delle posizioni contributive fornito all'ISTAT per il progetto ASIA. In questa base dati è amplificato il problema della discrepanza temporale delle informazioni contenute nell'archivio anagrafico con quelle contenute nell'archivio delle dichiarazioni contributive, essendovi un intervallo temporale che può arrivare fino a 36 mesi

Tabella 1 - Gli effetti di variazioni demografiche sulle variabili dell'archivio anagrafico

<i>Tipo di Variazione</i>	<i>Effetto sull'Archivio Anagrafico</i>
TRASFORMAZIONI GIURIDICHE	
ditte individuali ↔ società di persone o di capitali società di fatto ↔ società regolarmente costituite	nuovo numero di matricola
Società di persona ↔ società di capitale Società di capitale ↔ altro tipo società di capitale	stesso numero di matricola
FUSIONI	
Fusione mediante costituzione di una nuova società e cessazione delle società che si estinguono	nuovo numero di matricola per le posizioni della nuova società
Fusione mediante incorporazione di una o più società, che si estinguono, verso un'altra che rimane in vita	stesso numero di matricola per le posizioni della società incorporante
SCORPORI	
Scorporo di un'impresa da una precedente	nuovo numero di matricola per la società scorporata

Fonte: INPS – Manuale di istruzioni in materia di iscrizione e classificazione delle aziende.

5.3 Dalla posizione contributiva all'impresa: le chiavi di aggancio

Se una posizione contributiva non può fare riferimento a imprese diverse, più posizioni contributive possono invece appartenere ad una stessa impresa. E' necessario dunque individuare le chiavi di aggancio che consentano di collegare tra loro le posizioni contributive che si riferiscono ad una stessa impresa.

Negli archivi dell'INPS sono presenti le seguenti variabili che possono essere utilizzate come chiavi di aggancio:

- il codice fiscale;
- la partita IVA;
- la matricola di riferimento;
- la ragione sociale.

Una analisi delle caratteristiche di ciascuna di queste variabili permette di poter definire la regola di *link*, ossia quella strategia di abbinamento delle posizioni contributive che sceglie la migliore chiave di aggancio in funzione del suo contenuto informativo.

- Il codice fiscale e la partita IVA

Il codice fiscale è un codice assegnato in maniera univoca dal Ministero delle Finanze ad una impresa al momento della sua nascita; esso è di tipo alfanumerico o numerico a seconda che l'impresa sia costituita da una *persona fisica* (imprese individuali e liberi professionisti) o da *persona giuridica* (società di persone, società di capitali). In particolare, alle persone giuridiche (società) viene assegnato un codice fiscale di tipo numerico che, all'atto della nascita, coincide con la partita IVA. Nel caso in cui un'impresa varia la propria sede spostandosi in una diversa provincia, avviene l'attribuzione di un nuovo codice di partita IVA, mentre il codice fiscale numerico rimane invariato.

Poiché il codice fiscale è una variabile che identifica l'unità economico giuridica in maniera univoca e, teoricamente, è posseduta da tutte le posizioni contributive, essa viene considerata come chiave principale su cui basare l'aggregazione.

- *La matricola di riferimento*

Un'azienda deve presentare la dichiarazione dei versamenti contributivi nella sede INPS della provincia in cui svolge attività con dipendenti. Nel caso in cui l'attività viene svolta in più provincie, la dichiarazione deve essere presentata presso ogni sede per i dipendenti di competenza, a meno che l'impresa non sia stata autorizzata dall'Ispettorato Provinciale del Lavoro a presentare in un'unica sede le dichiarazioni che fanno capo a province diverse (*accentramento contributivo*).

Nel caso in cui l'impresa ha aperto in una stessa sede provinciale più posizioni contributive, queste sono distinte in posizione *principale* e posizioni *secondarie*, a seconda che si tratti della prima posizione aperta oppure delle posizioni aperte successivamente alla prima. Le posizioni secondarie presentano sui modelli di rilevazione dell'INPS nello spazio riservato alla *matricola di riferimento* la matricola della posizione principale dell'azienda a cui la posizione secondaria si riferisce.

La variabile '*matricola di riferimento*' è collegata alla variabile '*tipo azienda*'. Quest'ultima indica se una posizione contributiva appartiene a imprese autorizzate o meno all'accentramento contributivo e se la posizione contributiva è unica nell'ambito della provincia. Nel caso in cui l'impresa è autorizzata all'accentramento contributivo, ossia è autorizzata a presentare nell'ambito di una stessa provincia le dichiarazioni contributive di competenza di sedi provinciali diverse è possibile, mediante la matricola di riferimento, ricostruire l'impresa. Per le imprese non autorizzate all'accentramento contributivo le variabili tipo azienda e matricola di riferimento non permettono di ricostruire totalmente l'impresa, in quanto non consentono di agganciare tra loro le eventuali posizioni contributive di una stessa impresa aperte in sedi provinciali diverse.

Inoltre nel caso in cui si utilizzi come chiave di aggancio la matricola di riferimento, si possono avere agganci di posizioni contributive che non sono coerenti con quelli ottenuti utilizzando come chiave di aggancio il codice fiscale. Questi casi si ritiene debbano essere imputati a variazioni dell'archivio anagrafico che portano a modifiche del codice fiscale, nell'ambito di una stessa posizione contributiva, non seguite dall'aggiornamento della variabile matricola di riferimento. Occorre dunque procedere con cautela nell'uso della matricola di riferimento come chiave di aggancio.

- *La Ragione Sociale*

La ragione sociale, ossia la denominazione dell'impresa, è una variabile alfanumerica presente negli archivi INPS che costituisce una ulteriore potenziale chiave identificativa dell'impresa. Essa è associata al codice fiscale in modo univoco e, come tale, presenta una notevole forza esplicativa. Dal punto di vista applicativo, tuttavia, la ragione sociale non è di immediato utilizzo, data la estrema varietà di combinazioni testuali in cui l'informazione sulla ragione sociale può essere inserita nell'archivio, in assenza di indicazioni sistematiche dell'INPS circa la registrazione dell'informazione.

La regola di *link* scelta nella procedura di ricostruzione delle imprese si basa principalmente su tre chiavi di aggancio, considerate in modo gerarchico, in funzione del diverso potere esplicativo che ognuna di esse ha nell'individuare univocamente un'impresa: codice fiscale, partita IVA e matricola di riferimento.

La chiave principale è il codice fiscale che, come abbiamo visto, individua in modo univoco un'impresa. Le variabili partita IVA e matricola di riferimento vengono invece utilizzate come chiavi secondarie, in due momenti successivi: inizialmente nel tentativo di recuperare il codice fiscale nei casi in cui questo sia mancante oppure errato; successivamente per associare fra loro le posizioni contributive per cui non è stato possibile individuare l'impresa mediante il codice fiscale.

Gli errori commessi nella ricostruzione delle imprese possono dipendere da due circostanze:

- a) vengono erroneamente aggregate posizioni contributive che in realtà appartengono ad imprese differenti; ciò si può verificare, ad esempio, nel caso di un mancato aggiornamento del codice fiscale in seguito a trasformazioni giuridiche dell'impresa (ad esempio scorpori, fusioni etc.);

- b) non vengono agganciate posizioni contributive appartenenti alla stessa impresa, ad esempio, quando si è in presenza di un'errata registrazione del codice fiscale, che pur rimanendo formalmente corretto, risulta essere diverso da quello di un'impresa a cui una posizione contributiva effettivamente si riferisce.

5.4 La procedura di ricostruzione delle imprese

5.4.1 Una descrizione generale della metodologia

Il processo di ricostruzione delle imprese, a partire dalle posizioni contributive presenti nell'archivio anagrafico e nell'archivio delle dichiarazioni contributive, si articola nelle seguenti due fasi:

I. una prima fase il cui obiettivo è quello di produrre un archivio delle posizioni contributive in vita in un generico istante temporale, raggruppate in famiglie secondo una regola di *link* che collega tra loro le posizioni appartenenti ad una stessa impresa; questa fase si compone dei seguenti passi (figura 1):

- verifica della correttezza formale del codice fiscale e della partita IVA e correzione dei codici fiscali errati, utilizzando come chiave di recupero i codici di partita IVA corretti oppure la matricola aziendale, ove presente;
- formazione delle *famiglie* di posizioni contributive, utilizzando come regola di *link* quella che utilizza come chiave di aggancio, in successione, il codice fiscale, la partita IVA e matricola di riferimento, a seconda della loro esistenza e correttezza.
- imputazione del codice di attività economica alla eventuale posizione contributiva che ne risulti priva, mediante l'assegnazione del codice di attività economica della posizione con il maggior numero di addetti nell'ambito della famiglia;

Fig. 1 - I passi della prima fase della ricostruzione delle imprese

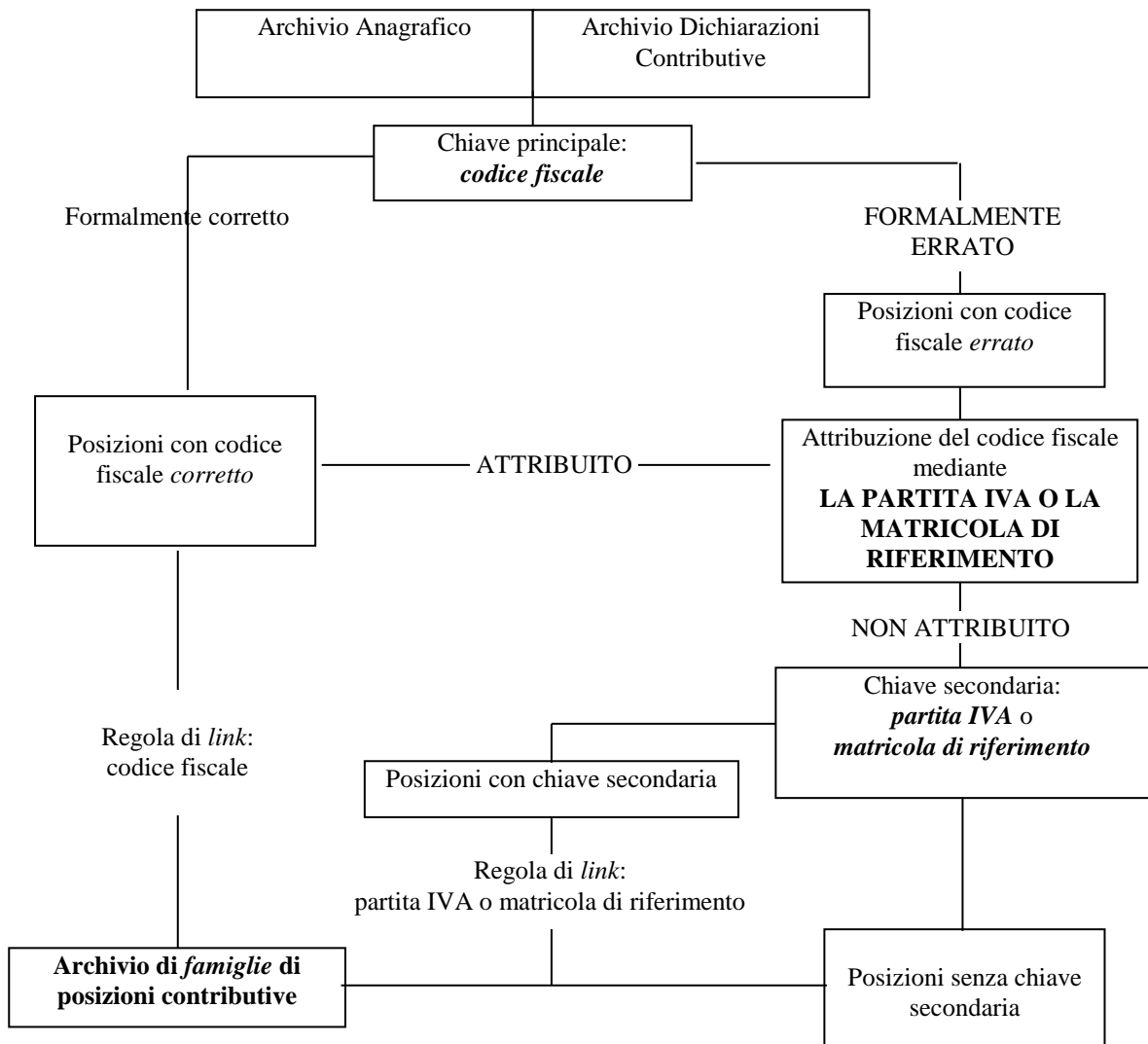


Tabella 3 – Regole di assegnazione dei caratteri anagrafici e strutturali alle imprese

Variabile dell'impresa	Regola di attribuzione
<i>Codice fiscale</i>	Se non unico, quello corrispondente alla sede amministrativa
<i>Ragione sociale</i>	Se non unica, quella corrispondente alla posizione nata per prima
<i>Data di costituzione</i>	Quella della posizione nata per prima
<i>Stato di attività</i>	Per definizione tutte le imprese risultano attive nel periodo di riferimento considerato
<i>Divisione di attività economica</i>	Quella prevalente in termini di numero di occupati
<i>Sezione di attività economica</i>	Quella corrispondente alla divisione attribuita
<i>Numero di occupati dipendenti</i> , per ciascuna posizione professionale	La somma degli occupati dipendenti delle posizioni che costituiscono l'impresa
<i>Ammontare delle retribuzioni</i> , per ciascuna posizione professionale	La somma delle retribuzioni delle posizioni che costituiscono l'impresa

5.5 Un esercizio di ricostruzione dell'impresa al mese di dicembre 1996

Della procedura implementata viene realizzata una prima applicazione sulla base dei dati degli ultimi aggiornamenti disponibili dell'archivio anagrafico e dell'universo dei modelli DM10.

La base dati considerata si riferisce alle posizioni contributive delle aziende presenti nell'anagrafica nel periodo aprile 1996 – febbraio 1999. Per tali posizioni le informazioni relative all'occupazione e alle retribuzioni, derivanti dai moduli DM10, sono disponibili con dettaglio mensile, ma limitatamente al periodo aprile 1996 – aprile 1998.

L'analisi viene condotta sulla base di due fondamentali output, la cui predisposizione costituisce la prima fase della procedura:

1. il totale delle posizioni contributive in vita nel periodo aprile 1996 – aprile 1998;
2. l'insieme delle imprese ricostruite. Quest'ultimo va riferito ad un dato istante temporale, per convenzione rappresentato, in riferimento all'universo disponibile, dal mese di dicembre 1996²⁰. La necessità di riferirsi ad un singolo mese come istante di riferimento della ricostruzione deriva dalla constatazione della natura stessa delle unità di analisi, la cui conformazione è suscettibile di modificazioni nel corso del tempo. A tal proposito, è inoltre necessario precisare che la mancanza di allineamento nell'aggiornamento delle informazioni di natura anagrafica e strutturale, ha delle importanti implicazioni sulla qualità delle informazioni relative all'impresa ricostruita: le informazioni anagrafiche si riferiscono al mese di febbraio 1999 che non coincide con l'istante di riferimento delle variabili relative all'occupazione.

La procedura di individuazione dei due output viene realizzata seguendo alcuni passi consecutivi che prevedono, in fase iniziale, l'esame ed il controllo della validità dei codici fiscali delle posizioni contributive disponibili. In seguito, sulla base dei codici corretti, si procede alla ricostruzione degli identificativi delle posizioni risultate non complete, adottando delle procedure di interrelazione di volta in volta implementate in funzione delle informazioni ad ogni livello disponibili.

²⁰ Nell'attuale contesto di analisi la scelta del mese di dicembre 1996 consente di predisporre una base dati coerente con le informazioni deducibili dall'archivio delle imprese ASIA, al momento disponibile, appunto, per il dicembre '96.

Delle posizioni contributive con codice fiscale corretto o ricostruito vengono selezionate le unità in vita (attive e temporaneamente sospese) nel mese di dicembre 1996, riportate a livello di impresa effettuando un'operazione di aggregazione per codice identificativo.

5.6 Alcuni risultati

La ricostruzione delle imprese ha inizio su una base dati costituita da un totale di 1.624.617 posizioni contributive (fig. 2). In seguito ad un primo check sulla qualità dei codici fiscali si rileva che tale identificativo risulta corretto per il 99,8% delle posizioni disponibili.

Utilizzando la partita IVA e la matricola di riferimento come chiavi di *link* per l'individuazione dei codici fiscali errati, vengono recuperati gli identificativi di 111 posizioni contributive. L'operazione di ricostruzione dei codici fiscali comporta un residuo di 3.026 posizioni a cui non è stato possibile associare un identificativo. Tra queste, in particolare, prevalgono i casi in cui il codice è mancante (52%) e per cui, pur risultando formalmente corretto, non lo è per il digit finale di riscontro (42%).

Una breve analisi sui casi a cui non è stato possibile attribuire un codice fiscale, evidenzia una certa caratterizzazione delle posizioni problematiche per settore di attività economica (tav.1): il 49% di esse sono classificate nella sezione dell'intermediazione monetaria e finanziaria ed oltre il 18% nella sezione delle attività immobiliari, noleggio, informatica, ricerca, altre attività professionali ed imprenditoriali.

Tav. 1-Posizioni contributive INPS senza codice fiscale per sezione di attività economica

Settore di attività economica	Posizioni		
	N.	% sul totale senza codice fiscale	% sul totale con e senza codice fiscale
Estrazione di minerali	2	0,1	0,00
Attività manifatturiere	233	7,7	0,01
Prod. e distr. di en. Elet. , gas e acqua	100	3,3	0,01
Costruzioni	299	9,9	0,02
Com. all'ingr. e al det.; rip. Autov., motocicli e di beni personali e per la casa	74	2,4	0,00
Alberghi e ristoranti	29	1,0	0,00
Trasporti, magazzinaggio e comunicazioni	21	0,7	0,00
Intermediazione monetaria e finanziaria	1.494	49,4	0,09
Att. immob., noleggio, informatica, ricerca, altre att. profes. ed imprendit.	560	18,5	0,03
Altri ^(a)	774	26,0	0,05
Totale	3.026	100	0,19

(a) Agricoltura, pesca, pubblica amministrazione e difesa, istruzione, sanità, altri servizi pubblici. Sono inoltre incluse alcune imprese per cui non è stato possibile individuare univocamente la sezione di attività economica in seguito alla trascodifica del codice di attività ISTAT assegnato dall'INPS.

Sono queste sezioni che raggruppano tipologie di attività economica recentemente interessate da considerevoli modificazioni giuridiche e sociali, per cui la più ampia mobilità di imprese o parti di imprese da e verso di esse può in parte spiegare la maggiore possibilità di commettere errori in fase di registrazione dei dati.

Le posizioni non recuperate danno luogo ad una perdita media, calcolata rispetto ai mesi di attività, di circa 11.000 occupati dalla base di analisi²¹. In particolare, tra esse vi è una

²¹ Un analogo calcolo effettuato su tutte le posizioni contributive presenti nella base dati analizzata, stima l'occupazione complessiva a 11,4 milioni di occupati.

posizione contributiva a cui sono associati 1.349 dipendenti. Si osserva, inoltre, come il 25% delle posizioni perse ha almeno 1 dipendente e il numero medio di occupati per posizione contributiva è pari a 3,55.

Delle posizioni con codice identificativo corretto o assegnato, l'88% sono singole rispetto al codice fiscale (fig.3). Poiché queste ultime costituiscono, per definizione, un'impresa, è sul rimanente 12% (196.815 posizioni contributive) che va applicata la procedura di ricostruzione dell'impresa.

In seguito alla selezione sulla data di costituzione e di cessazione si ottengono 1.340.064 posizioni in vita nel mese di dicembre 1996 (attive o temporaneamente sospese). Di queste, 1.188.019 derivano dalle posizioni precedentemente classificate "singole" rispetto al codice fiscale e 152.045 dalle "multiple".

Prima di procedere all'aggregazione dei codici fiscali per l'individuazione dell'impresa nei casi di posizioni multiple, la procedura di ricostruzione dell'impresa prevede l'attribuzione di un codice di divisione economica alle posizioni contributive che ne sono prive (la sezione è invece sempre presente). Del totale delle 152.045 posizioni di origine, in circa 7 mila unità non è presente il codice di divisione economica. Di queste, il 18% può essere potenzialmente recuperato attraverso posizioni contributive non in vita nel periodo di riferimento. Per valutare la bontà dell'attribuzione del codice di divisione in questi casi si effettua un raffronto tra il codice sezione così recuperato e la sezione originariamente presente. I due codici sono coincidenti nel 29% dei casi. Solo per essi viene dunque attribuita la divisione. Tale operazione consente di attribuire la variabile per il 5% delle posizioni che ne erano prive.

Aggregando le posizioni contributive inizialmente classificate come "multiple" per codice fiscale risultano individuate 73.770 imprese. Di queste, tuttavia, solo 53.356 continuano ad essere multiple nel mese di dicembre 1996 essendo, parte di quelle originarie, cessate oppure non ancora sorte in riferimento al periodo di osservazione selezionato. Il raggruppamento delle posizioni singole viene dunque incremento dell'insieme dei codici fiscali che hanno perso degli abbinamenti in seguito alla selezione sul mese di dicembre 1996 dando luogo, nel complesso a 1.208.433 imprese (fig.4).

Una breve analisi sulle imprese così costituite evidenzia la presenza di una serie di codici fiscali formalmente corretti ma "atipici". Nella maggior parte di essi i digit dei codici fiscali si presentano con una certa sistematicità, lasciando pensare a delle forzature effettuate in fase di registrazione delle informazioni. A tali codici corrisponde, frequentemente, un elevato numero di posizioni contributive caratterizzate da diverse ragioni sociali, presenti nelle più disparate province, e con molteplici settori di attività economica. Tali codici fiscali vengono eliminati, poiché potenziale fonte di distorsione nella stima di alcune variabili di interesse.

In seguito a tali depurazione si giunge ad individuare un numero di imprese pari a 1.261.780 unità. Di esse il settore manifatturiero ricopre il 25% dei casi, seguito dal commercio con il 24%, dalle costruzioni e dalle attività professionali ed imprenditoriali, rispettivamente con il 14% ed il 12% (tav.2). Un'analisi delle imprese individuate in riferimento al numero di dipendenti dichiarati, evidenzia che il 72% dei casi risulta avere da 1 a 9 occupati. Dall'analisi emerge, inoltre, come il 16% delle stesse risulta non avere dipendenti nel mese di dicembre 1996 (tav.3).

Tav. 2 – Imprese INPS in vita a dicembre 1996 per sezione di attività economica
Sezioni di attività economica

	Imprese	
	N.	%
Estrazione di minerali	2.967	0,24
Attività manifatturiere	311.990	24,73
Prod. e distr. di energia elettrica, gas e acqua	771	0,06
Costruzioni	178.903	14,18
Commercio all'ingrosso e al dettaglio; rip. autoveicoli, motocicli e di beni personali e per la casa	299.614	23,75
Alberghi e ristoranti	94.551	7,49
Trasporti, magazzinaggio e comunicazioni	32.143	2,55
Intermediazione monetaria e finanziaria	16.533	1,31
Attività immobiliari, noleggio, informatica, ricerca, altre att. professionali ed imprenditoriali	150.579	11,93
Altre	173.729	13,77
Totale	1.261.780	100

Tav. 3 – Imprese INPS in vita a dicembre 1996 per classe dimensionale

Classe dimensionale	Imprese	
	N.	%
0	198.849	15,76
1-9	902.935	71,56
10-19	91678	7,27
20-49	46.560	3,69
50-99	12.225	0,97
100-199	5.284	0,42
200-499	2.708	0,21
500-999	787	0,06
1000+	754	0,06
Totale	1.261.780	100

5.7 Validazione dei risultati

5.7.1 Confronto delle imprese ricostruite da fonte INPS con le imprese dell'archivio ASIA

Un primo riscontro sulla validità della procedura di ricostruzione viene effettuato mediante uno studio comparativo tra le caratteristiche delle imprese individuate dalle fonti INPS e quelle delle imprese attive dell'archivio ASIA relativo al mese di dicembre 1996. La comparazione viene effettuata abbinando le imprese presenti nei due archivi per mezzo del codice fiscale.

Del complesso delle 915 mila imprese individuate in INPS, si abbinano con ASIA 1.066.092 (84%) unità, di cui le imprese pluriposizionate rappresentano il 3,6% e le uniposizionate il 96,4% .

I risultati di una breve analisi sulle caratteristiche delle imprese presenti in INPS ma non rilevate in ASIA sono riportati nelle tavv. 4 e 5.

La tav. 4 riporta la distribuzione di tali imprese per sezione di attività. E' interessante notare che il 25% delle non abbinata compaiono nella sezione K.

La tav. 5 riporta la distribuzione delle imprese non abbinate per classe dimensionale. In particolare, il 63% di esse appartiene alla classe dimensionale da 1 a 9 addetti.

Tav. 4 – Imprese INPS non abbinate con le imprese ASIA per sezione di attività economica

Sezione di attività economica	Imprese non abbinate		
	N.	% sul totale delle non	% sul totale per classe dimensionale
Estrazione di minerali	238	0,12	8,02
Attività manifatturiere	19.401	9,91	6,22
Prod. e distr. di energia elettrica, gas e acqua	75	0,04	9,73
Costruzioni	21.708	11,09	12,13
Commercio all'ingrosso e al dettaglio; rip. autoveicoli, motocicli e di beni personali e per la casa	20.317	10,38	6,78
Alberghi e ristoranti	8.733	4,46	9,24
Trasporti, magazzinaggio e comunicazioni	3.074	1,57	9,56
Intermediazione monetaria e finanziaria	1.920	0,98	11,61
Attività immobiliari, noleggio, informatica, ricerca, altre att. professionali ed imprenditoriali	49.021	<u>25,05</u>	32,56
Altri	71.201	36,38	40,98
Totale	195.688	100	15,51

Tav. 5 – Imprese INPS non abbinate con le imprese ASIA per classe dimensionale

Classe dimensionale	Imprese non abbinate		
	N.	% sul totale delle non abbinate	% sul totale per classe dimensionale delle imprese INPS
0	53.184		26,75
1-9	124.265	<u>63,5</u>	13,76
10-19	9.835	5,0	10,73
20-49	5.454	2,8	11,71
50-99	1.537	0,8	12,57
100-199	642	0,3	12,15
200-499	337	0,2	12,44
500-999	134	0,1	17,03
1000+	344	0,2	<u>45,62</u>
Totale	195.688	100	15,51

In totale, le imprese INPS non presenti nell'archivio ASIA assorbono 1.876.356 occupati, equivalenti al 18% dei dipendenti distribuiti sul complesso delle imprese ricostruite.

La Tav. 6 riporta, la distribuzione del differenziale assoluto tra gli occupati registrati dai due archivi. Per le imprese uniposizionate tale differenziale, in media pari a -0,19 occupati, è concentrato in poche unità. Per quanto riguarda le imprese pluriposizionate, per cui il differenziale medio è pari a -4,79, la distribuzione mostra una maggiore dispersione del fenomeno tra le unità osservate. D'altra parte, la più ampia dimensione potenzialmente caratterizzante le aziende pluriposizionate spiega in parte le maggiori differenze assolute riscontrate.

Tav. 6 - Differenza assoluta: occupati INPS - occupati ASIA

Indicatori	Tipologia impresa per numero di posizioni contributive	
	Imprese uni posizionate	Imprese pluri posizionate
	N. osservazioni	1.027.806
N. osservazioni per cui la variabile è diversa da 0	181.297	15.277
Media	-0,19	-4,79
Deviazione standard	85,14	1.195,46
<i>Quantili:</i>		
100% massimo	2759	89.043
75%	0	0
50%	0	0
25%	0	0
0% minimo	-86.013	-181.105
99%	5	47
90%	0	4
10%	-1	-2
1%	-6	-40
Moda	0	0

L'analisi comparativa viene successivamente estesa evidenziando il differenziale degli occupati per sezione di attività economica e classe di addetti. A tal proposito si ipotizza che le informazioni contenute in ASIA siano la base per il riscontro e, dunque, le imprese confrontate vengono classificate secondo le caratteristiche indicate da tale archivio.

L'analisi condotta, di tipo "micro", si basa sullo studio della distribuzione dell'indicatore:

$$\sigma = \frac{(OCC_{INPS} - OCC_{ASIA})}{((OCC_{INPS} + OCC_{ASIA})/2)} 100 \quad \text{con} \quad -200 \leq \sigma \leq +200.$$

In particolare, σ assume il suo valore minimo quando $OCC_{INPS}=0$ e $OCC_{ASIA} \neq 0$ il suo valore massimo quando $OCC_{ASIA}=0$ e $OCC_{INPS} \neq 0$. Infine σ assume valore 0 quando $OCC_{INPS}=OCC_{ASIA}$.

L'indicatore considerato, quindi, risulta molto sensibile a grosse differenze tra i dipendenti considerati in INPS e quelli in ASIA, in particolare ai casi in cui uno dei due archivi registra 0 dipendenti. Nell'insieme di imprese considerato, in particolare, 170.009 imprese uniposizionate e 4.198 imprese pluriposizionate risultano avere zero dipendenti in almeno uno degli archivi. Ai fini dell'analisi seguente le suddette imprese sono escluse per rendere il confronto più leggibile.

Ai fini dell'individuazione delle imprese possono essere commessi due tipi di errori. Il primo riguarda le imprese che hanno zero dipendenti in INPS e dipendenti positivi in ASIA Per le uniposizionate si tratta di 81.870 imprese per un totale di 222.076 dipendenti ASIA (tav.7). Per le pluriposizionate si tratta di 1.556 imprese per un totale di 12.854 dipendenti ASIA (tav. 8). Per esse occorre un supplemento di indagine. Il secondo tipo di errore riguarda le imprese con dipendenti positivi in INPS e zero dipendenti in ASIA. Per le uniposizionate si tratta di 26.739 imprese per un totale di 99.399 dipendenti INPS (tav. 7). Per le pluriposizionate si tratta di 1.759 imprese per un totale di 24.310 dipendenti INPS (tav. 8).

Da tale esercizio emergere il problema dell'aggiornamento degli archivi INPS. Una parte di queste discrepanze, sorprendenti particolarmente per quanto riguarda le imprese

uniposizionate, può essere ascritta ai continui aggiornamenti dell'archivio INPS. Ricordiamo, infatti, che il confronto in atto riguarda un file INPS di un anno più aggiornato rispetto a quello utilizzato per la costruzione di ASIA. Un esempio potrà chiarire il punto. Tra le imprese uniposizionate con 0 dipendenti INPS cade il 99% delle imprese sospese. Di queste ultime solo però il 45% risulta con zero dipendenti in ASIA.

Tav. 7 - La presenza di imprese con zero dipendenti – Imprese uniposizionate

INPS						
ASIA		Dipendenti 0		Dipendenti>0		Totale
	Dipendenti 0	N. imprese	61.400	N. imprese	26.739	88.139
		Dip. INPS	0	Dip. INPS	99.399	99.399
		Dip. ASIA	0	Dip. ASIA	0	0
	Dipendenti>0	N. imprese	81.870	N. imprese	857.797	939.667
		Dip. INPS	0	Dip. INPS	5.783.039	5.783.039
		Dip. ASIA	222.076	Dip. ASIA	5.857.172	6.079.248
	Totale	N. imprese	143.270	N. imprese	884.536	1.027.806
		Dip. INPS	0	Dip. INPS	5.882.438	5.882.438
		Dip. ASIA	222.076	Dip. ASIA	5.857.172	6.079.248

Tav. 8 - La presenza di imprese con zero dipendenti – Imprese pluriposizionate

INPS						
ASIA		Dipendenti 0		Dipendenti>0		Totale
	Dipendenti 0	N. imprese	883	N. imprese	1.759	2.642
		Dip. INPS	0	Dip. INPS	24.310	24.310
		Dip. ASIA	0	Dip. ASIA	0	0
	Dipendenti>0	N. imprese	1.556	N. imprese	34.088	35.644
		Dip. INPS	0	Dip. INPS	2.473.536	2.473.536
		Dip. ASIA	12.854	Dip. ASIA	2.668.581	2.681.435
	Totale	N. imprese	2.439	N. imprese	35.847	38.286
		Dip. INPS	0	Dip. INPS	2.497.846	2.497.846
		Dip. ASIA	12.854	Dip. ASIA	2.668.581	2.681.435

L'analisi sulla base dell'indicatore σ è effettuata per sezioni e classi di addetti, nella tabelle 18 e19 rispettivamente per le imprese uniposizionate e pluriposizionate. Da esse sono escluse tutte le imprese che hanno 0 dipendenti in ASIA o in INPS. L'analisi è circoscritta al dominio C-K delle sezioni di attività economica.

Tav. 9 – Confronto INPS-ASIA per sezione di attività economica e classi di dipendenti: posizioni singole - numero imprese abbinate, media, mediana, primo e nono decile dell'indice: $((OCC_{INPS} - OCC_{ASIA}) / ((OCC_{INPS} + OCC_{ASIA}) / 2)) \cdot 100$

SEZIONI DI ATTIVITA' ECONOMICA		CLASSI DI ADDETTI								Totale
		1-9	10-19	20-49	50-99	100-199	200-499	500-999	1000 e +	
Estrazione di minerali	N.	1.543	403	157	13	6	2	—	—	2.124
	Media	1,68	0,01	-1,02	5,30	-7,39	8,37	—	—	1,17
	Q90	0	7,40	5,13	0	0,97	1,67	—	—	4,44
	Mediana	0	0	0	-2,04	-0,33	0,84	—	—	0
	Q10	0	0	-53	-16,95	-42,69	0	—	—	0
Attività manifatturiere	N.	175.365	39.817	21.277	5.340	2.308	857	131	40	245.135
	Media	0,48	-0,55	-0,18	-2,90	-2,53	-3,82	-3,97	-5,23	0,13
	Q90	0	0	2,74	6,06	0,67	0	0	0	0
	Mediana	0	0	0	-1,17	0	0	0	0	0
	Q10	0	0	0	-11,32	-4,47	-2,44	-2,02	-4,38	0
Energia elettrica, gas e acqua	N.	403	94	87	29	16	8	3	2	642
	Media	-1,46	0,34	-3,70	-19,27	-14,24	-75,48	-54,26	-0,04	-3,78
	Q90	0	0	6,67	1,60	1,31	0	0	0	0
	Mediana	0	0	0	-1,09	0	-34,28	0	-0,04	0
	Q10	0	0	-4,88	-79,67	-23,67	-198,84	-162,79	-0,09	-0,90
Costruzioni	N.	107.302	8.815	3.041	431	111	41	9	—	119.750
	Media	0,77	-1,94	-1,79	-10,13	-12,27	-5,62	-15,94	—	0,45
	Q90	0	0	4,44	9,52	4,58	4,17	1,12	—	0
	Mediana	0	0	0	0	0	0	0	—	0
	Q10	0	-6,45	-4,44	-29,21	-35,02	-16,22	-137,89	—	0
Commercio e riparazioni	N.	200.161	10.839	4.204	703	256	123	41	24	216.351
	Media	0,21	-1,33	-1,42	-3,17	-0,35	-0,96	1,74	-5,38	0,09
	Q90	0	0	4,08	8,55	4,69	11,72	6,28	6,26	0
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	0	-3,51	-10,91	-2,69	-2,71	-2,63	-2,11	0
Alberghi e ristoranti	N.	50.934	2.769	810	136	46	13	4	1	54.713
	Media	0,27	-2,17	-3,18	-14,13	-6,58	-0,25	-4,40	19,34	0,06
	Q90	0	0	6,78	22,22	33,12	41,07	0	19,34	0
	Mediana	0	0	0	0	0	0	0	19,34	0
	Q10	0	0	-4,00	-94,74	-63,26	-19,17	-17,62	19,34	0
Trasporti, magazzinaggio e comunicazioni	N.	24.926	2.692	1.438	392	151	84	11	5	29.699
	Media	1,02	-1,63	-2,91	-2,88	-2,30	-4,92	0,76	-37,36	0,50
	Q90	0	6,06	3,64	10,17	5,71	1,12	3,08	19,72	0
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	-6,45	-4,44	-19,78	-10,07	-6,52	0	-199,92	0
Intermediazione monetaria e finanziaria	N.	15.108	642	480	211	116	60	22	15	16.654
	Media	0,36	-0,11	-0,15	-1,82	-1,09	-3,16	1,39	1,15	0,28
	Q90	0	6,90	4,44	4,21	4,16	4,67	3,69	3,98	0
	Mediana	0	0	0	0	0	0,28	1,04	0,72	0
	Q10	0	0	-6,07	-6,56	-2,87	-2,08	-0,13	-1,75	0
Att. immob. e altre att. profes. ed imprendit.	N.	93.101	4.668	2.287	621	301	135	25	3	101.141
	Media	0,26	-2,13	-1,53	-2,18	-2,99	-5,82	0,53	2,18	0,08
	Q90	0	0	4,44	13,08	1,99	8,81	2,37	6,53	0
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	0	-3,77	-15,15	-10,20	-6,55	0	0	0
Totale	N.	668.843	70.739	33.781	7.876	3.311	1.323	246	90	786.209
	Media	0,42	1,05	-0,77	-3,48	-2,79	-4,24	-2,93	-5,35	0,17
	Q90	0	0	3,51	7,25	1,50	1,83	3,10	3,91	0
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	0	0	-12,24	-5,29	-3,32	-2,19	-3,04	0

Tav. 10 – Confronto INPS-ASIA per sezione di attività economica e classi di dipendenti: posizioni multiple - numero imprese abbinate, media, mediana, primo e nono decile dell'indice: $((OCC_{INPS} - OCC_{ASIA}) / ((OCC_{INPS} + OCC_{ASIA}) / 2)) \cdot 100$

SEZIONI DI ATTIVITA' ECONOMICA		CLASSI DI ADDETTI								Totale
		1-9	10-19	20-49	50-99	100-199	200-499	500-999	1000 e +	
Estrazione di minerali	N.	101	78	63	12	6	1	1	1	263
	Media	16,16	1,92	-0,23	-7,88	-7,07	32,91	0	-194,13	5,58
	Q90	72,00	11,11	3,77	3,63	0	32,91	0	-194,13	28,57
	Mediana	0	0	0	-3,89	0	32,91	0	-194,13	0
	Q10	-13,33	-9,52	-6,90	-29,01	-40,96	32,91	0	-194,13	-11,76
Attività manifatturiere	N.	2.826	1.382	1.291	651	488	486	189	160	7.473
	Media	7,17	-1,65	-1,25	-3,30	-4,86	-3,16	-5,87	-7,64	1,06
	Q90	40,00	12,50	10,75	11,11	2,51	1,74	0,68	3,51	18,18
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	-10,52	-8,70	-12,86	-9,06	-3,91	-7,37	-17,79	-9,52
Energia elettrica, gas e acqua	N.	35	32	33	17	24	17	10	7	175
	Media	14,56	3,14	6,73	-3,75	0,10	4,84	-6,61	-35,82	3,06
	Q90	66,67	13,33	18,18	3,24	4,20	10,21	53,54	34,80	16,66
	Mediana	0	0	0	0	0	0	0	-2,53	0
	Q10	0	0	-3,39	-18,80	-1,09	-5,47	-87,83	-190,04	-5,40
Costruzioni	N.	2.527	938	670	200	94	55	17	7	4.508
	Media	9,64	-1,37	-1,94	-4,02	-1,80	-3,11	-5,59	-6,23	4,55
	Q90	66,67	18,18	18,35	18,11	16,14	13,92	0	18,36	36,36
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	-15,38	-14,29	-24,45	-9,97	-20,35	-22,22	-72,45	-9,52
Commercio e riparazioni	N.	5.423	2.008	1.224	300	106	58	16	13	9.148
	Media	6,08	-0,43	-0,63	-704	-11,60	-13,73	-1,19	2,22	2,97
	Q90	40,00	9,52	9,52	9,85	8,55	5,40	0	10,07	25,00
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	-10,52	-4,88	-24,75	-38,50	-42,59	-3,58	0	-8,00
Alberghi e ristoranti	N.	1.319	372	325	124	54	31	13	13	2.251
	Media	8,99	-1,83	-3,47	-27,06	-15,51	-10,55	-5,79	13,02	2,50
	Q90	66,67	22,22	13,33	19,82	34,87	0	0	42,24	40,00
	Mediana	0	0	0	-1,84	0	0	0	0,90	0
	Q10	0	-22,22	-10,53	-138,46	-110,07	-53,79	-16,92	-1,62	-23,16
Trasporti, magazzinaggio e comunicazioni	N.	880	412	437	211	131	117	40	34	2.262
	Media	11,04	1,54	1,16	-5,55	-0,19	-0,69	-0,14	-9,36	4,09
	Q90	55,27	23,53	15,05	15,13	9,43	5,61	1,60	10,19	33,50
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	-11,76	-9,52	-9,09	-20,83	-11,88	-2,81	0	-8,76	-11,11
Intermediazione monetaria e finanziaria	N.	263	59	51	27	36	51	35	50	572
	Media	12,79	4,73	2,52	12,72	1,63	0,25	3,12	-0,86	7,43
	Q90	66,67	26,08	16,00	45,07	4,44	5,82	6,61	9,96	25,00
	Mediana	0	0	0	3,51	0,73	0	2,47	2,79	0
	Q10	0	0	0	-2,78	0	-0,21	0	-0,07	0
Att. immob. ed altre att. Profess. ed imprend.	N.	1.929	621	639	340	212	137	39	18	3.935
	Media	7,64	0,62	-1,06	-3,12	-2,44	2,41	5,52	-7,33	3,20
	Q90	40,00	9,52	18,18	16,04	4,25	4,24	27,60	24,40	22,22
	Mediana	0	0	0	0	0	2,24	0	0	0
	Q10	0	-8,69	-9,09	-21,10	-18,67	-6,74	0	-89,88	-7,41
Totale	N.	15.303	5.902	4.733	1.882	1.151	953	360	303	30.587
	Media	7,80	-0,60	-0,98	-5,56	-4,46	-3,27	-2,91	-6,62	2,92
	Q90	46,15	14,28	12,24	12,82	6,11	4,13	4,28	7,72	27,02
	Mediana	0	0	0	0	0	0	0	0	0
	Q10	0	-10,52	-8,00	-20,56	-15,11	-5,35	-3,00	-8,76	-9,23

Le principali evidenze risultanti dalle tavv. 9 e 10 sono di seguito descritte:

1. Le imprese uniposizionate, come era prevedibile, risultano avere nel complesso errori molto contenuti.
2. Ad eccezione di due settori, le distorsioni medie non superano lo 0,5% e per la parte centrale della distribuzione non esistono differenze tra le due misure. Per il settore C il 10% delle imprese registra una sovrastima dell'INPS superiore al 4,44%. Il settore E è caratterizzato da poche imprese che determinano una distorsione media negativa.
3. L'analisi per dimensione mette in luce una prevalenza di sottostime INPS, senza una chiara direzione di aggravamento o allievemento del problema al variare della dimensione.
4. Per le imprese pluriposizionate il quadro è meno roseo. Sono compresenti errori di sovrastima e di sottostima, con una maggiore gravità dei primi. Nel complesso delle 30.587 imprese esaminate per l'ultimo 10% della distribuzione c'è un errore di almeno il 27%. Per il primo 10% si registra un errore di -9,23%.
5. Questo tipo di distribuzioni è comune a tutti i settori. Mostra errori particolarmente rilevanti, sia in eccesso che in difetto il settore degli Alberghi e Ristoranti. Il settore delle Costruzione presenta errori in eccesso sostanzialmente maggiore che nel totale generale.

5.8 Conclusioni

La ricostruzione dell'impresa è un'operazione necessaria al fine di individuare una precisa unità di analisi nella stima dei dipendenti e delle relative retribuzioni lavorative, ricorrendo alle informazioni contenute negli archivi amministrativi dei versamenti contributivi dell'INPS. La definizione di posizioni contributive presso ogni azienda, infatti, può avere diversa natura dipendendo, spesso, da ragioni di carattere giuridico, fiscale, organizzativo etc. A tal proposito, si ricorre ad una procedura di ricostruzione basata sull'abbinamento delle posizioni contributive. Al fine di interrelare ogni singola unità ad una famiglia potenzialmente individuante un'impresa, vengono utilizzate chiavi diverse di *link*. Ogni impresa ricostruita viene dunque qualificata con una serie di variabili anagrafiche e strutturali.

Un'applicazione della procedura sulla base dei dati degli ultimi aggiornamenti disponibili dell'archivio anagrafico e dell'universo dei modelli DM10, consente di ricostruire e di valutare l'adeguatezza della metodologia implementata riferendosi alle imprese in vita nel mese di dicembre 1996. A tal proposito, vengono effettuate delle comparazioni con l'archivio ASIA delle imprese attive. Una prima validazione dei risultanti consente di osservare alcune sostanziali differenziazioni tra gli archivi confrontati, particolarmente evidenti per alcuni settori di attività economica e dimensione dell'impresa. Tali differenze sono da imputare prevalentemente al disallineamento temporale tra le fonti di dati utilizzate. Questa constatazione suggerisce il ricorso ad aggiuntive fonti informative che consentano di integrare alcuni punti di debolezza della fonte INPS (ad esempio l'indagine Istat sulle Grandi Imprese).

SCHEMA SINTETICO DELLA IMPLEMENTAZIONE E VALUTAZIONE DELLA PROCEDURA DI RICOSTRUZIONE DELL'IMPRESA

Fig. 2 - Controllo sulla correttezza ed attribuzione dei codici fiscali

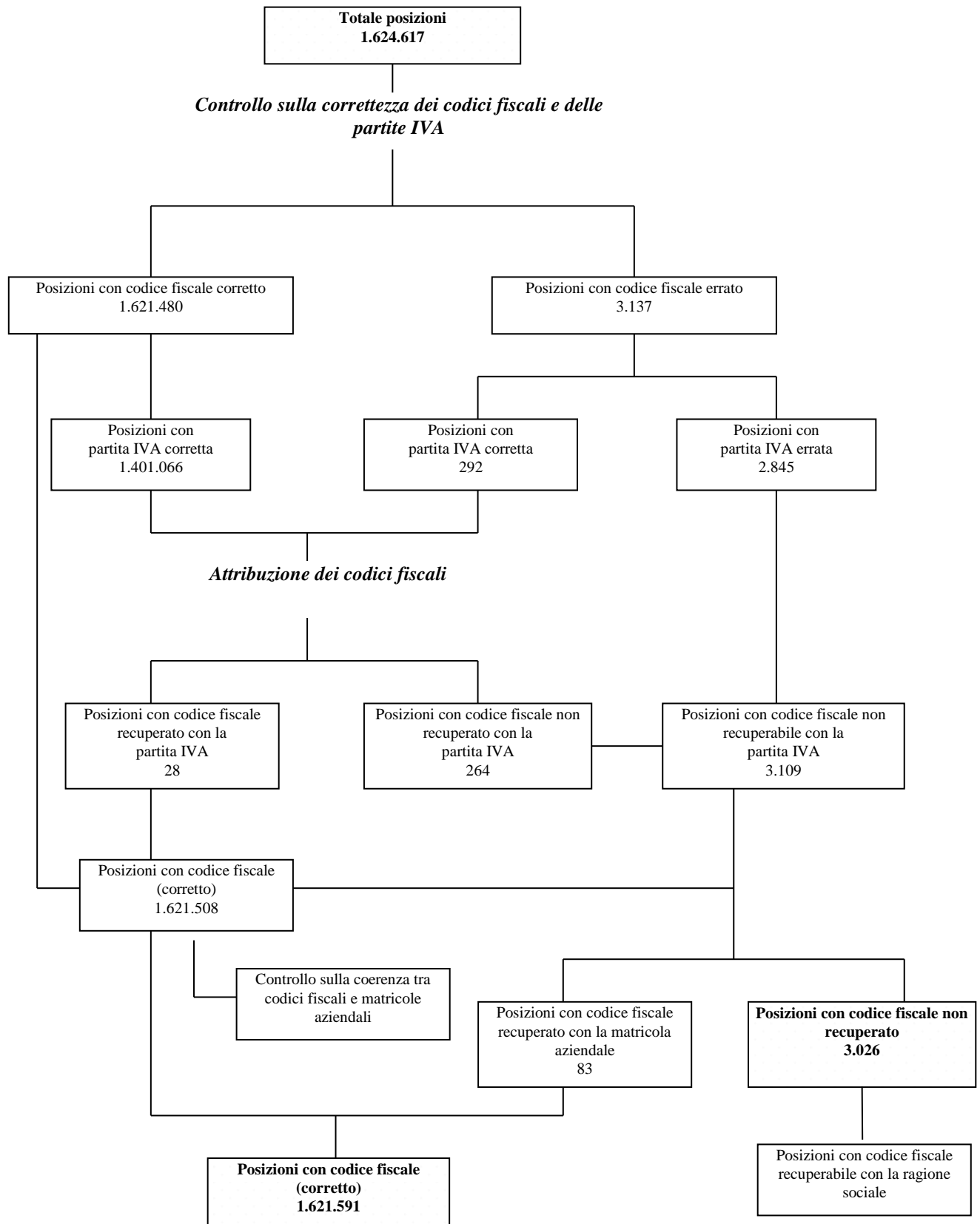


Fig. 3 - Analisi sui codici fiscali e attribuzione delle divisioni di attività economica

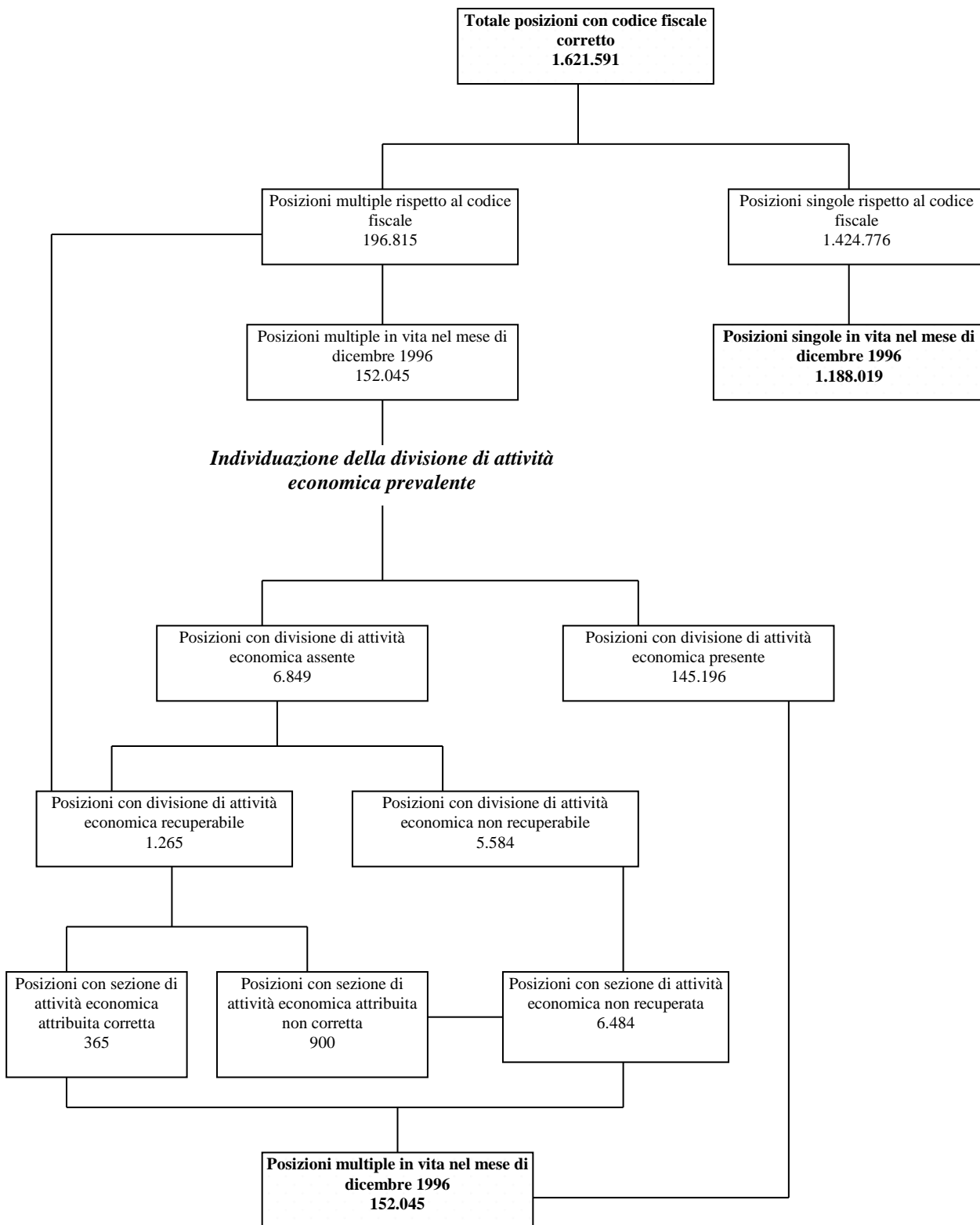
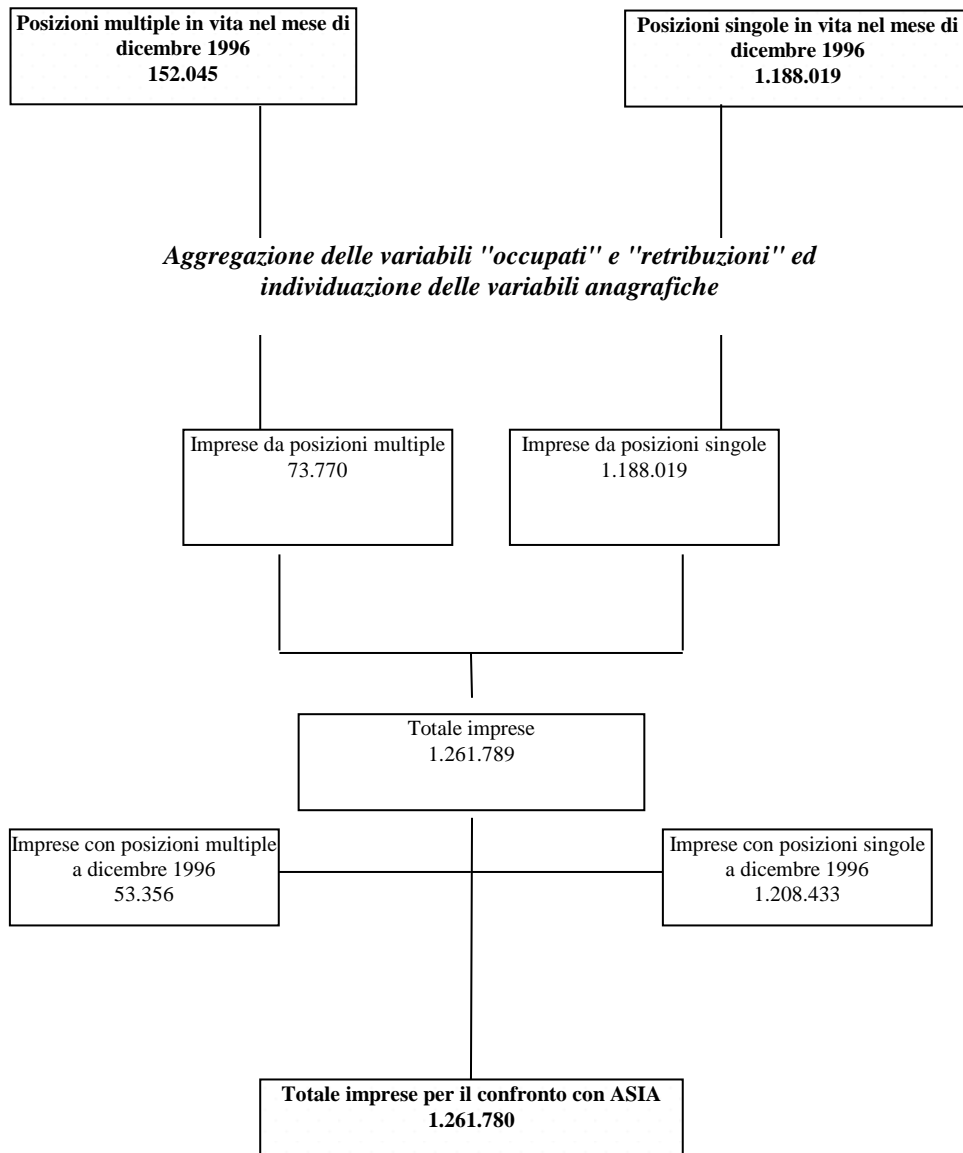


Fig. 4 - Ricostruzione dell'impresa nel mese di dicembre 1996



6 LA STIMA DEL TEMPO DI LAVORO DELLE DICHIARAZIONI CONTRIBUTIVE INPS. ALCUNE IPOTESI PRELIMINARI E PROBLEMI APERTI

6.1 Introduzione

Un ideale indice del costo del lavoro dovrebbe avere al denominatore il numero delle ore lavorate. In sostanza sia per il numeratore - il costo del lavoro - che per il denominatore - le ore lavorate - si dovrebbe disporre di informazioni almeno trimestrali:

$$LCt = TLCt / THt, \text{ dove}$$

TLCt è il costo totale del lavoro nel periodo t

THt è il numero totale delle ore lavorate nel periodo t.

Le ore lavorate al denominatore per riferirsi esattamente al costo del lavoro al numeratore dovrebbero provenire esattamente dalla stessa fonte. Ciò significa dovere esplorare la possibilità di rilevare congiuntamente tali variabili.

Arrivare ad una stima delle ore effettivamente lavorate e non di una qualsiasi altra unità di misura del tempo lavorato (ore retribuite, giornate lavorate, ecc.), tra l'altro, è necessario a soddisfare anche esigenze nazionali e del Regolamento Congiunturale sulle statistiche delle imprese. Risulta, quindi indispensabile tentare di sfruttare al meglio le potenzialità informative sul tempo di lavoro del modulo DM10 per arrivare a stimare le ore effettivamente lavorate.

Un'altra variabile strettamente connessa con le Ore effettivamente lavorate sono le Ore Retribuite. Il calcolo di una variabile di tempo retribuito può essere sia una tappa intermedia per il calcolo delle Ore Effettivamente Lavorate che una variabile importante *per se*. La raccolta di dati sulle ore pagate è già effettuata in molti paesi, tra gli altri, USA e Canada. Tra le altre cose, al di là, degli aspetti congiunturali, le ore retribuite rappresentano una fonte di informazione necessaria per il calcolo di salari orari e quindi per la stima di differenziali di diverso tipo. Proprio a questo riguardo la mancanza di una stima ufficiale ed affidabile di questo tipo di variabile è stata origine di un recente dibattito sull'andamento dei differenziali regionali. Recentemente, in letteratura è stata avanzata la tesi che i differenziali salariali tra Nord e Sud d'Italia si sono progressivamente ridotti negli ultimi anni. Tale conclusione sembra però inficiata dalla qualità dei dati degli archivi INPS sulla base dei quali è stata avanzata. Questo tipo di problemi deve essere considerato se la fonte di informazione primaria per la stima delle ore è la fonte INPS.

In questo capitolo si discute della possibilità di stimare le Ore Retribuite e le Ore Effettivamente Lavorate sulla base delle denunce contributive DM10. Le informazioni desumibili da questi moduli soffrono, ai fini della statistica ufficiale, di due difetti principali: un problema di insufficienza ed uno di qualità. L'insufficienza riguarda il fatto che da essi è solo desumibile la quantità di lavoro retribuita, non quella effettivamente prestata. Per quanto riguarda la qualità dei dati particolarmente rilevante per le variabili di tempo di cui si sta trattando è la tendenza all'under-reporting legata all'evasione contributiva. La stima delle Ore (anche solo quelle retribuite) non può fondarsi quindi unicamente sulla fonte INPS DM10. Per sopperire ad entrambe i difetti occorre affiancare alla fonte INPS altre fonti informative. Alcune fonti di informazione già presenti in ISTAT potranno essere utili allo scopo, mentre altre dovranno essere predisposte. Le fonti ausiliarie avranno due scopi. In parte serviranno per controllare ed in parte per integrare le informazioni desumibili dai DM10.

Tale sistema di fonti dovrebbe permettere di ottenere due output principali: le ore retribuite e le ore effettivamente lavorate. Nel processo produttivo immaginato le prime costituiranno, per così dire, l'input intermedio per le seconde. Le ore retribuite costituiscono comunque un output rilevante e

potrà permettere confronti internazionali con quei paesi che già dispongono di un'informazione simile.

Il procedimento che viene sotto delineato comporta un set di interventi differenti, per tentare di risolvere il problema della qualità e dell'insufficienza. Tale procedimento comporta una serie di errori. Essi vengono illustrati per potere essere controllati.

Il lavoro è strutturato come segue. Nel paragrafo 2 sono illustrate le principali richieste della Comunità Europea. Nel paragrafo 3 si affrontano i problemi della costruzione delle variabili in questione a partire dai DM10, mantenendo l'ipotesi che i dati di partenza siano sufficientemente affidabili, ovvero che il tempo retribuito non sia affetto da under-reporting. Il processo è in due fasi. Nella paragrafo 3.1 si delinea il procedimento per ottenere una stima delle ore retribuite, e si sottolineano gli errori cui si va incontro. Nel paragrafo 3.2 si espongono le difficoltà cui si va incontro per arrivare ad una stima delle Ore Lavorate. Nel Paragrafo 4 si espone il problema dell'under-reporting, un possibile modo di affrontarlo ed i suoi limiti.

6.2 Alcune linee per una stima delle ore sulla base dei dati INPS

L'INPS rappresenta il principale ente nazionale di previdenza e assistenza sociale. Ad esso le imprese italiane, mensilmente, denunciano le somme soggette a contribuzione e calcolano l'ammontare dei contributi dovuti attraverso il modulo DM10. In esso sono riportati distintamente per gruppi di lavoratori (che rappresentano particolarità contributive) oltre alle informazioni riguardanti il numero dei dipendenti, le retribuzioni lorde pagate ed i contributi a debito, anche informazioni sul tempo retribuito dall'impresa. Tale indicazione costituisce la base per il controllo sul rispetto dei minimi imponibili. La retribuzione minima imponibile è infatti calcolata come retribuzione per unità di tempo retribuito.

In generale i dati desumibili dai DM10 riguardano il numero di **giornate retribuite**. Per giornata retribuita si intende ogni giornata per cui il datore di lavoro paga una retribuzione soggetta a contribuzione obbligatoria, indipendentemente dal numero di ore per cui si versa tale retribuzione. A questa regola generale vanno affiancate alcune eccezioni.

Il lavoro straordinario è riportato in ore .

Per i lavoratori assunti con contratto part-time, il numero di ore retribuite.

Per gli apprendisti vengono riportate le settimane retribuite.

A parte queste eccezioni, dunque, il DM10 fornisce il monte giornate retribuite mensile per qualifica (e, potenzialmente, per alcuni tipi di contratto). Si tratta di trasformare il numero di giornate (o settimane) retribuite in ore effettivamente lavorate.

I passi profilati sono i seguenti.

Calcolo delle ore retribuite. In questa fase occorre trasformare le giornate retribuite (e le settimane) in ore retribuite. A questo scopo saranno utili le informazioni contrattuali.

Calcolo delle ore effettivamente lavorate. La trasformazione delle ore retribuite in ore effettivamente lavorate richiede di tenere conto di tutte le ore retribuite ma non lavorate (ad es. malattie, ferie, festività etc) e delle ore effettivamente lavorate ma non retribuite.

6.3 Il calcolo delle ore retribuite

6.3.1 L'informazione del DM10

L'informazione contenuta nel DM10 riguarda il numero di giornate che l'impresa retribuisce nel mese di competenza. Il numero di giornate è riportato ad una base di sei giorni settimanali: 40 ore settimanali su 5 giorni lavorativi costituiscono 6 giornate retribuite. Così un mese interamente retribuito è pari a 26 giornate; un anno interamente retribuito è pari a 312 giornate. Per un lavoratore con contratto a tempo pieno si usa quindi un coefficiente implicito di 1.2 per trasformare prestazioni lavorative inferiori a 5 giorni. Viceversa prestazioni su 6 giorni per un numero di ore inferiore a

quella della settimana lavorativa standard di sei giorni vengono computate come 6 giornate . (Revelli 95).

Un primo problema nasce dal modo in cui viene interpretata questa norma. Le 26 giornate mensili per dipendente a tempo pieno rappresentano un valore convenzionalmente seguito nella prassi.

Così accade non di rado che anche in mesi “corti”, come febbraio il numero di giornate retribuite medio per dipendente sia di 26. Inoltre, talvolta, viene anche superato il limite di 26 giornate retribuite per dipendente. Tale osservazione può essere la base di una regola di controllo per il numero di giornate retribuite, dato che secondo regola nessuno potrebbe dichiarare più di 26 giornate per dipendente. (R&P 1990)

Non è neanche chiaro come sono trattate le giornate di assenza non retribuite (permessi, scioperi, CIG). Ogni giornata di assenza, in un posto di lavoro che prevede il lavoro su cinque giorni, dovrebbe contare come 1.2 giornate in meno. Non è affatto chiaro quali siano le prassi delle imprese per approssimare i risultati che ne derivano. (R&P 1990)

Considerando la natura convenzionale dell’informazione sulle giornate retribuite nasce il sospetto se essa apporti qualche guadagno rispetto all’informazione sui dipendenti. Nell’eventuale confronto bisogna tenere presente che l’informazione sui dipendenti riguarda tutti i dipendenti che hanno anche solo una giornata retribuita. Almeno in linea di principio, come indicazione del tempo retribuito, la variabile giornate dovrebbe tenere conto dei rapporti di lavoro che sono cessati o sono iniziati durante il mese di riferimento.

6.3.2 L’uso delle informazioni contrattuali

Attualmente l’ISTAT con la rilevazione sulle Retribuzioni Contrattuali (RC) rileva e segue 80 Contratti Collettivi Nazionali di Lavoro (CCNL). Essi sono rappresentativi dei circa 300 attualmente in vigore. Oltre alle informazioni riguardanti la parte economica del contratto sono osservate alcune informazioni riguardanti il tempo di lavoro. Tra esse le più rilevanti sono il numero di ore contrattualmente stabilito e il numero di giornate previste per ferie, festività ed altri permessi. Poter risalire al contratto applicato dall’impresa sarebbe il modo più corretto per calcolare il numero di ore retribuite. In realtà tale informazione è difficilmente recuperabile. Anche ammettendo che da un codice attività economica (ATECO o CSC) si possa risalire al contratto collettivo nazionale che l’impresa con maggiore probabilità applica non si dispone in ISTAT (e non si pensa di disporre in futuro) delle informazioni su tutti i contratti.

L’alternativa più praticabile sembra essere quella di usare la stima per codice di attività economica delle ore contrattuali. A ciascuna impresa appartenente ad un certo settore di attività economica contraddistinto da un codice ATECO a due cifre può essere attribuito il numero di ore contrattuali medie per giornata (riportate su una base di sei giorni) stimate per quel settore di attività economica. Nell’ipotesi in cui ciascuna giornata retribuita indicata nei DM10 sia composta dal numero medio di ore previsto dal contratto passare dalle giornate retribuite alle ore retribuite è un’operazione relativamente innocua. I contratti di lavoro stabiliscono l’orario che deve essere prestato su base giornaliera, settimanale, o mensile. È così possibile — almeno in linea di principio- ottenere il numero di ore giornaliero retribuito dall’impresa per qualifica. Tale procedimento è soggetto a due fonti di errore. La prima riguarda l’imputazione delle ore medie sulla base del settore di attività economica. La seconda fonte di errore, più grave, nasce nel caso in cui viene misurata una giornata retribuita a fronte di poche ore retribuite. Alcuni esempi possono essere: permessi, scioperi, CIG parziale o integrazioni dell’impresa alla CIG, etc. In casi del genere la procedura illustrata ovviamente sovrastima sistematicamente le ore retribuite²².

Al numero di ore retribuite calcolate per il lavoro ordinario a tempo pieno vanno sommate le ore di

²² Un tale calcolo ovviamente pone anche un problema di coerenza tra le variabili Retribuzione e Costo del lavoro e la variabile Ore Retribuite. L’attribuzione della durata della giornata secondo contratto anche qualora sono retribuite solo poche ore per giornata comporta che la retribuzione per ora sia sottostimata.

lavoro dei contratti part-time e le ore di lavoro straordinario.

Si individua a questo punto un'altra fonte di errore che deriva dalla definizione di orario straordinario secondo le norme contributive (vedi, per maggiori dettagli, in appendice 2). Secondo tali norme sono straordinarie le prestazioni superiori alle 40 ore settimanali. Molti contratti di lavoro stabiliscono però un orario normale di lavoro inferiore alle 40 ore. Il metodo di calcolo delineato sopra non considera quindi le ore in eccesso alle ore normali di contratto e inferiori alle 40 ore. Questo problema può essere parzialmente affrontato. Nel DM10 sono riportate oltre alle ore di straordinario effettuate anche il numero di dipendenti che l'hanno effettuato. Ovviamente se è stato effettuato orario straordinario in senso contributivo è stato effettuato anche orario supplementare (chiamiamo supplementare il numero di ore comprese tra le ore contrattuali e le 40). Purtroppo non è possibile una stima puntuale esatta del numero di ore supplementari di una azienda. Essa infatti dipende dalla distribuzione dello straordinario tra i lavoratori e le settimane. Supponendo un mese di 4 settimane e la seguente notazione:

n =numero di dipendenti che effettuano lo straordinario

os =ore supplementari

oc =ore contrattuali

osi =ore straordinarie riportate nel DM10

allora abbiamo se $osi/4n \geq 1$

$n(40-oc) \leq os \leq 4n(40-oc)$

Il limite inferiore si ha nel caso in cui ciascun lavoratore abbia fatto straordinario per solo una settimana. Il limite superiore nel caso in cui tutti i lavoratori abbiano fatto straordinario per 4 settimane. Un valore medio è quindi $2.5n(40-oc)$.

Il calcolo così fatto riguarda solo tre delle qualifiche. Sono infatti esclusi gli apprendisti il cui tempo retribuito è misurato in settimane. L'ultimo passo sarà, quindi quello di aggiungere le ore retribuite degli apprendisti. Tali ore dovrebbero essere ottenute con un metodo analogo a quello presentato.

Nella tabella seguente è riportato un quadro sinottico del calcolo delle ore retribuite.

Calcolo del tempo per una impresa appartenente ad una cella contrassegnata da un ATECO a due cifre

Variabile	Fonte	
N° di giornate retribuite (lavoro ordinario, tempo pieno) per dirigenti, impiegati operai	DM10	*
N° di ore contrattuali giornaliere (riportate su una base di sei giorni)	RC	=
N° di ore retribuite (lavoro ordinario contratti a tempo pieno)		+
N° di ore retribuite contratti part time	DM10	+
N° di ore di straordinario ^a	DM10	=
N° di ore retribuite esclusi apprendisti		+
N° ore retribuite calcolate per gli apprendisti ^b		=
N° di ore retribuite		

^aNella tabella ci si riferisce alle ore di straordinario effettive. Esse sono date dalle ore di straordinario da DM10 + le ore supplementari stimate

^bN° ore retribuite calcolate per gli apprendisti=N° di settimane retribuite degli apprendisti (DM10)* N° di ore contrattuali settimanali(RC)

6.4 Il calcolo delle Ore Effettivamente Lavorate

La seconda fase, ovvero il passaggio dalle ore retribuite alle ore effettivamente lavorate, è più complesso.

Le ore effettivamente lavorate possono essere desunte dal seguente calcolo:

Ore effettivamente lavorate=Ore retribuite+Ore lavorate non retribuite-Ore retribuite non lavorate

Le Ore non retribuite ma lavorate non sono usualmente identificabili da fonti di impresa. Esse dovrebbero però costituire un ammontare rilevante principalmente per le qualifiche dirigenziali.

La maggior parte delle Ore non lavorate ma retribuite²³ può essere raggruppata sostanzialmente in tre categorie:

- 1) Assenza per ferie, festività, congedi retribuiti.
- 2) Assenza per malattia, infortunio, maternità.
- 3) Assenza per la partecipazione dei lavoratori a corsi di formazione ed istruzione, per la parte pagata dall'impresa.

Per le assenze di cui al punto 1) è corrisposta dall'impresa una retribuzione soggetta a contribuzione. Esse quindi dovrebbero risultare come tempo retribuito.

Per le assenze riguardanti il punto 2) ci sono trattamenti differenziati. Alcune note a riguardo della malattia sono riportate in appendice. In linea del tutto generale si può dire che la legge prevede che tali assenze siano remunerate dal datore di lavoro e/o dall'Istituto di previdenza. Anche laddove però la sostanziale quota di remunerazione è a carico dell'Istituto cui compete, i CCNL prevedono normalmente integrazioni a carico del datore di lavoro. In casi come questo l'impresa dovrebbe dunque dichiarare la giornata retribuita anche se non è lavorata.

Per le assenze riguardanti il punto 3) bisogna tenere presente che tra le componenti escluse dalla retribuzioni imponibili previste dal 1.1.98 ci sono "le spese relative ad oneri e servizi utilizzabili dalla generalità dei dipendenti o categorie di dipendenti volontariamente sostenute per specifiche finalità di educazione, istruzione, ..." (ANIV 1999). Ne deriva che qualora le imprese paghino corsi di istruzione o educazione che occupino giornate intere il calcolo delle ore retribuite non dovrebbe risentire di difetti. Qualora i corsi riguardino porzioni di giornata ricadiamo nel caso illustrato precedentemente di sovrastima delle ore retribuite.

Il manuale metodologico non include o esclude esplicitamente le cause di cui al punto 3).

Per quanto riguarda le cause di cui al numero 1) e 2) e più in generale per provare a stimare per impresa le ore effettivamente lavorate occorre integrare le informazioni INPS con altre fonti. Al momento in ISTAT si può contare su due tipi di fonti. La prima fonte è sempre quella relativa alle Retribuzioni Contrattuali. Da tale fonte è desumibile il numero di ore di ferie, festività e congedi previste da ciascun contratto per ogni tipo di qualifica (e quindi stimate per ATECO a due cifre). L'utilizzo di questa fonte va quindi incontro ai seguenti limiti: a) le ore previste sono ore di diritto b) sono computate su base annuale (vanno quindi eventualmente trimestralizzate²⁴) c) si tiene sotto controllo solo le cause di cui al punto 1) d) Sebbene le informazioni contrattuali specifichino le assenze di diritto separatamente per anzianità l'assenza di informazioni sull'anzianità aziendale sul DM10 non permette di controllare questo fattore di distorsione²⁵

Il secondo tipo di fonte sono alcune indagini già presenti in ISTAT. Nella tabella in appendice al capitolo sono riportate insieme alla fonte INPS integrata dalle informazioni contrattuali

²³ Prima di procedere occorre però ricordare che le ore retribuite desunte dal calcolo del paragrafo precedente si riferiscono alle ore per cui si paga una retribuzione esposta a contribuzione.

²⁴ In modo da tenere conto dei picchi di assenza dal lavoro in particolari periodi dell'anno (ferie estive e natalizie, festività natalizie etc.)

²⁵ Mentre è probabile che l'errore dovuto alla trimestralizzazione scompaia in aggregato lo stesso non si può presumere per l'errore che si commette a causa di strutture dell'occupazione caratterizzate da diversa anzianità. E' probabile che in alcuni settori si concentri una forza lavoro con anzianità aziendale sopra la media o viceversa. L'anzianità media all'interno delle imprese dipende dalle politiche di ricambio che esse attuano. Tali politiche possono essere differenziate a seconda di settori di appartenenza.

(DM10/RC), le principali fonti ISTAT dalle quali sono desumibili informazioni relative alle variabili di interesse (Ore Retribuite e Ore Effettivamente Lavorate). Sono state escluse dalla tabella quelle fonti che forniscono un'informazione solo parziale (es. Prodcum che riporta solo le ore effettivamente lavorate dagli addetti alla produzione, o la Stima provvisoria sul valore aggiunto che riporta solo le ore effettivamente lavorate da operai e apprendisti). La tabella illustra il campo di osservazione della rilevazione la variabile richiesta, il dettaglio di qualifiche per cui l'informazione è disponibile, la cadenza della informazione e la serie di dati ad oggi disponibile.

Tra le rilevazioni che potenzialmente possono dare un contributo alla stima delle ore lavorate trimestrale, da una parte abbiamo la Rilevazione sulle Grandi Imprese e dall'altra alcune rilevazioni strutturali.

La Rilevazione sulle Grandi Imprese a cadenza mensile rileva con riferimento ad impiegati ed operai le ore ordinarie effettivamente lavorate, le ore straordinarie, e le ore non lavorate ma retribuite dal datore di lavoro. Naturalmente trasformare le ore retribuite DM10/RC in ore lavorate secondo un coefficiente (rapporto tra ore lavorate e ore retribuite) desunto dalle Grandi Imprese può essere distortivo. Nelle grandi imprese è presumibile che alcuni fenomeni che determinano l'assenza dal lavoro si comportino in maniera differente da quanto avviene nelle imprese di dimensioni più piccole.

Tra le indagini strutturali che rilevano le ore effettivamente lavorate quella sul Sistema dei Conti delle Imprese (SCI), e quella sulle Piccole Imprese (PMI) garantiscono una serie storica di dati. Purtroppo tali serie di dati sono molto brevi e quindi si prestano con difficoltà ad eventuali metodi di trimestralizzazione.

6.5 Il problema dell'underreporting delle giornate retribuite

Finora si è assunto che la variabile di partenza (le giornate retribuite) non siano affette da errore. Come è stato accennato, però, le imprese hanno un incentivo a sottodichiararle. Questo particolare incentivo nasce a causa del disegno legislativo di riferimento, che nella sostanza impone che la base imponibile ai fini contributivi non possa essere inferiore ai minimi salariali calcolati su base giornaliera. Se l'impresa denuncia monti retributivi bassi, per far quadrare i conti non resta che correggere adeguatamente le giornate retribuite.

Tale incentivo all'under-reporting varia presumibilmente al variare di alcune caratteristiche dell'impresa, come la localizzazione geografica, il settore di attività economica, la dimensione, e probabilmente di altre variabili.

L'evasione contributiva attuata mediante l'under-reporting sia delle retribuzioni che delle giornate retribuite ha già dato origine ad alcune valutazioni probabilmente errate sull'andamento dei differenziali salariali tra le diverse aree del paese (vedi ad es. Lucifora e Origo 1997). In Ginzburg et al. (1998) si sostiene che tale tipo di pratica assieme alla procedura di normalizzazione adottata dall'INPS è causa dell'illusione che i divari salariali Nord-Sud si siano ridotti. Come è sottolineato dagli autori tali tipi di pratiche evasive si verificano più probabilmente al Sud e nei settori a prevalenza di piccole e piccolissime imprese.

Una prima indicazione del fenomeno si può trarre dall'analisi delle giornate retribuite nei dati O1M del 1994 (R&P 1998) riferiti agli operai in imprese dei rami 2,3,4 che indicano 12 mesi retribuiti. Il numero di giornate retribuite supera le 287 nell'89% dei casi nel Nord-Ovest, nel 92.3% nel Nord Est, nell'82.6% dei casi nel centro nel 56.6% nel Sud e nel 63.5% dei casi nelle Isole. Per le sole imprese appartenenti alle classi 1-5 e 6-19 per le cinque aree geografiche i valori scendono rispettivamente al 86.9% e 88.4% (NO), 88.7% e 90.7% (NE), 73.7% e 79% (C), 44.7% e 46.4% (S), 55.1% e 54.7% (Isole).

Una possibilità evidente per correggere i dati INPS è l'uso di fonti ausiliarie. L'efficacia di tale metodo però è fondata su una ipotesi abbastanza forte: i dati rilasciati alle fonti ISTAT dovrebbero essere più attendibili di quelli forniti all'INPS.

Una analisi approfondita della questione dovrebbe prendere in considerazione il modello di risposta

delle imprese alle rilevazioni esterne dei dati. Schematizzando, il rispondente di un'impresa nel rispondere ad un questionario recupera le informazioni da un sistema codificato di informazioni interne (*hard data* nella terminologia del Federal Committee on Statistical Methodology). Nel caso italiano è forse più giusto parlare di due sistemi informativi: uno ufficiale ed uno "in nero". I dati amministrativi italiani soffrono purtroppo del fatto che i rispondenti usano il primo sistema informativo, distorcendo quindi le informazioni. Questa caratteristica è abbastanza specifica del nostro paese e quindi non esiste, a nostra conoscenza, a livello internazionale, una letteratura sulla correzione dei dati inficiati da questo tipo di problema. La possibilità di correggere questo tipo di distorsioni utilizzando fonti ISTAT si fonda sull'ipotesi che nel rispondere a questo tipo di fonti le imprese utilizzino il vero sistema informativo, e questa assunzione equivale a fare ipotesi sulla razionalità dei rispondenti.

Sotto l'ipotesi mantenuta di buona fede nei confronti dell'ISTAT è possibile stimare un modello di correzione delle ore retribuite. I metodi di correzione possono essere almeno due. Uno basato sulla rilevazione delle vere ore retribuite per un sufficientemente alto numero di aziende di ciascun settore di attività economico. Il coefficiente di rapporto tra le ore retribuite calcolate sulla base del DM10 e quelle dichiarate all'indagine ausiliaria potrebbe essere quindi usato per correggere la stima delle ore retribuite effettuata sulla base della fonte INPS²⁶. Un'estensione di questo metodo può consistere nell'usare al posto della stima non condizionale dell'under-reporting delle ore il valore predetto da un modello di regressione. Il modello specificato ad esempio può regredire la sottostima delle ore retribuite per settore di attività economica su una serie di variabili esplicative probabilmente correlate al fenomeno in esame. Tra esse si pensa ad una variabile che indica la distribuzione sul territorio nazionale delle attività del settore, un indice della distribuzione per dimensione delle imprese nel settore, una stima del sommerso per settore di attività etc. A parte la necessità di un'indagine ausiliaria per ottenere una stima più attendibile per il tempo retribuito, non ci si nasconde le difficoltà che un metodo di regressione può nascondere. Un metodo analogo, e a cui questa idea è ispirata, è stato suggerito per la correzione del Censimento della popolazione negli Stati Uniti del 1980 ed è stata bersaglio di severe critiche. (vedi Freedman and Navidi 1986).

6.6 Conclusioni

La stima delle variabile Ore Effettivamente Lavorate a partire dai dati dei DM10 non sembra essere un problema facilmente risolvibile. Dalla fonte INPS integrata con le informazioni contrattuali si può ottenere una stima delle Ore Retribuite. Tale variabile, oltre ad essere utile per confronti internazionali e per il calcolo di costi e retribuzione per ora retribuita, rappresenta un passo intermedio verso la produzione di un indicatore di Ore Effettivamente Lavorate. Il calcolo delle Ore Retribuite calcolate per ATECO 91 a due cifre si basa sostanzialmente sulla attribuzione dell'orario medio per giornata stimato per quel settore di attività economica, così come desumibile dalla Rilevazione sulle Retribuzioni Contrattuali. Tale stima soffre di due maggiori tipi di difetto. Il primo deriva dal fatto che le imprese possono dichiarare una giornata a fronte di poche ore retribuite. Il secondo tipo di problema deriva dall'underreporting delle giornate retribuite causata dall'evasione contributiva.

La stima dell'underreporting dovrebbe essere basato su una fonte ausiliaria al momento non disponibile in ISTAT. Tale stima è però soggetta a sua volta ad alcune componenti di errore.

La stima delle Ore Lavorate non può avvenire senza l'ausilio di altre fonti. Al momento le soluzioni presenti in ISTAT non sembrano adeguate. L'espansione di un coefficiente di correzione basato sui dati delle Grandi Imprese può risultare in una stima distorta. D'altro canto l'idea di trimestralizzare le Ore Effettivamente Lavorate calcolate da alcune indagini strutturali, trova un vincolo nelle

²⁶ In ISTAT attualmente le Ore Retribuite sono rilevate solo dall'indagine sulle Grandi Imprese. Questa fonte ha il vantaggio di fornire dati congiunturali. Tuttavia espandere una correzione basata sul confronto dei dati INPS e dei dati ISTAT sulle grandi imprese all'universo di imprese potrebbe però risultare notevolmente distortivo.

lunghezza attuale delle serie.

Un indicazione finale sulla possibilità di trarre indicatori congiunturali del tempo di lavoro a partire dalla fonte INPS/DM10 potrà essere data solo dopo aver implementato ulteriori sperimentazioni.

Tali analisi dovrebbero mirare:

- a valutare la bontà di un indicatore di ore retribuite.
- a costruire e verificare relazioni strutturali tra ore retribuite secondo la fonte INPS e ore lavorate secondo le fonti istat (Grandi Imprese, Censimento Intermedio, Costo del Lavoro)
- verificare la possibilità di usare congiuntamente tali relazioni insieme alle informazioni INPS (ore retribuite ordinarie e, soprattutto, straordinarie) ed alla relazione tra ore retribuite e ore lavorate derivante dall'indagine Grandi Imprese per ottenere un indicatore congiunturale di Ore effettivamente lavorate.

APPENDICE al capitolo 6

1. Nota sulle particolarità contributive dello straordinario

Fino al 31/12/1995 per le sole imprese industriali era previsto un contributo al fondo disoccupazione per il lavoro straordinario. Con lavoro straordinario si intendeva quello eccedente le 48 ore settimanali. (L. 30.10.1955).

Dal 1/1/1996 la normativa è cambiata. Riassumendo si definisce lavoro straordinario (art.2 L. 28.12.95):

- a) quello che eccede le 40 ore nel caso di orario settimanale;
- b) quello che eccede la media di 40 ore settimanali nel caso di orario plurisettimanale previsto dal CCNL ovvero in applicazione di questi ultimi dagli accordi di livello inferiore. In tal caso il periodo di riferimento non può essere superiore a dodici mesi;

Si riferisce a tutte le categorie di lavoratori salvo quelle che svolgono mansioni direttive

La contribuzione varia a seconda della dimensione di impresa e del settore di attività:

Tipo Azienda	Aliquota	Limite orario
Qualsiasi settore con >15 dip.	5%	40 ore
Del settore industria >15 dip.	10%	44 ore
Settore industria >0	15%	48 ore

2. Nota sulle particolarità contributive della malattia

Per quanto riguarda la cause di cui al punto 2) occorre tenere presenti le norme in materia al fine di articolare una strategia di correzione delle ore retribuite.

La normativa in materia di indennizzo di malattia prevede un trattamento differenziato per taluni tipi di lavoratori. In generale si può dire che i contratti collettivi di lavoro prevedono una corresponsione a favore del lavoratore tale da garantirgli il mantenimento del trattamento che avrebbe ricevuto se avesse lavorato o di una quota significativa di esso. Normalmente i CCNL prevedono una integrazione da parte del datore di lavoro. Entrando più nello specifico si può schematizzare nella seguente maniera. Per i lavoratori che non hanno diritto all'indennità giornaliera da parte dell'INPS (ad es. impiegati, quadri e dirigenti dell'industria) è a carico del datore di lavoro l'intero trattamento economico previsto dal CCNL; per i lavoratori che abbiano diritto all'indennità giornaliera (ad es. gli operai dell'industria, nonché operai, quadri ed impiegati nel commercio) il datore di lavoro eroga la sola quota integrativa a tale indennità fino a concorrenza con il trattamento economico previsto dal CCNL. E' importante sottolineare questa distinzione perché ai fini contributivi solo la quota a carico del datore di lavoro rientra nell'imponibile. La cosa rilevante ai nostri fini è che nella generalità dei casi, quindi, almeno una parte del trattamento che riceve il lavoratore ammalato è imponibile. Per tutti i casi in cui sono previste integrazioni o corresponsioni consegue che le giornate di malattia sono considerate giornate retribuite da parte dell'azienda che compila il DM10.

La prassi può però discostarsi dal principio secondo cui i giorni di malattia per cui l'impresa corrisponde a suo carico parte o tutta la retribuzione sono da considerarsi giornate retribuite. Infatti soprattutto per le integrazioni salariali (quindi di entità parziale) per le aziende con pochi dipendenti e/o con retribuzioni basse introdurre tali giornate tra quelle retribuite potrebbe abbassare la retribuzione per giornata al di sotto dei minimali. Le imprese potrebbero quindi decidere di non considerarle giornate retribuite. In tali casi ovviamente si avrebbe una sottostima del tempo retribuito per le imprese (R&P 1990).

A parte questi casi comunque appare giusto collocazione all'interno di questo schema è quindi proprio a questo punto, ovvero nel calcolo delle ore non prestate da sottrarre all'orario di lavoro.

Non sembra affatto di facile utilizzazione l'informazione nel quadro D del DM10 relativa alle poste

messe a conguaglio per le indennità di malattia anticipate dall'INPS.

La stima delle giornate di malattia deve avvenire, dunque in qualche altro modo.

Analogamente, anche per maternità e infortuni esiste normalmente un trattamento integrativo a carico del lavoratore imponibile e quindi tali giornate rientrano nelle giornate retribuite.

APPENDICE 3: Le principali fonti ISTAT dal lato delle imprese per le variabili di tempo di lavoro

FONTE	CAMPO DI OSSERVAZIONE		VARIABILE RILEVATA	QUALIFICHE	CADENZA DELL'INFORMAZIONE	DISPONIBILITA' DELLA SERIE
	Dimensione	Settori ATECO(91)				
Grandi Imprese	>500 addetti	10-41 50-74	Ore retribuite	Impiegati; Operai;	Mensile	1995:1-1999:3
Grandi Imprese	>500 addetti	10-41 50-74	Ore Eff. Lavorate (ordinarie e straordinarie)	Impiegati; Operai;	Mensile	1995:1-1999:3
SCI	>20 addetti	10-93 (escluse 65, 66,67,91, 74.84.1)	Ore Eff. Lavorate	Imprenditori cooperatori e coadiuvanti; Dirigenti, quadri impiegati; Operai e commessi; Apprendisti; Lavoratori a domicilio; ^a	Annuale	89:96 per operai commessi e apprendisti 96 per tutte le categorie
PMI	1-19 addetti	10-93 (esclusi 65,66,91)	Ore Eff. Lavorate	Imprenditori,cooperatori ; Coadiuvanti familiari; Dirigenti, iuimpiegati, intermedi; Operai, Commessi; Apprendisti; Lavoratori a domicilio; ^a	Annuale	92:96 per operai e apprendisti 95:96 per tutte le categorie
Censimento Intermedio (Long Form)	Nessun limite	10-93 (esclusi 75, 80, 85, 91, 92.5, 92.6, 92.7.)	Ore Eff. Lavorate (ordinarie e straordinarie)	Dirigenti; Quadri, direttivi; Impiegati; Operai e personale ausiliario; Apprendisti;	Occasionale	1997
Rilevazione sulla flessibilità del mercato del lavoro	>9	10-93	Ore Eff. Lavorate (divise in ordinarie e straordinarie)	Lavoratori dipendenti	Occasionale	1995-1996
Rilevazione sulla Struttura delle Retribuzioni	>9	10-74	Ore Retribuite a tariffa intera	Dirigenti; Quadri; Impiegati; Operai; Apprendisti;	Quadriennale	1995
Rilevazione sulla struttura del costo del lavoro	>9	10-74	Ore Effettivamente lavorate	Lavoratori indipendenti; Dirigenti; Quadri, Impiegati; Operai e personale ausiliario; Apprendisti;	Quadriennale	1997

7 LE CARATTERISTICHE STRUTTURALI DELLO “PSEUDO-CAMPIONE” E DELL’UNIVERSO DEI DM10 INPS

7.1 Introduzione

La necessità di fornire informazioni congiunturali dal lato imprese su Occupazione, Retribuzioni, Ore lavorate e Costo del lavoro e sue componenti provengono oltre che da esigenze interne da richieste comunitarie. Attualmente l’Italia fornisce queste informazioni solo con riferimento alle imprese con più di 500 addetti. Si tratta di una porzione molto piccola dell’universo in termini di imprese anche se rilevante in termini di occupazione dipendente. La percentuale di imprese con più di 500 dipendenti sul totale delle imprese con dipendenti operanti nelle sezioni ATECO91 da C a K è pari 0.11%; in termini di dipendenti la quota è pari al 25.11% (Dati ASIA96 riferiti al 31 dicembre).

In sintesi le richieste comunitarie riguardano in primo luogo il Labour Cost Index (LCI) in secondo luogo lo Short Term Business Regulation (STBS). Nel primo caso sono sufficienti le seguenti aggregazioni di sezioni della NACE Rev.1: C+D+E, D, F, G+H+I, J+K e totale da C a K. Nel secondo vengono richieste stime a livello di due cifre NACE rev.1. Il manuale metodologico legato al Regolamento congiunturale raccomanda, inoltre, la suddivisione per quattro classi di addetti: 1-19, 20-99, 100-499, 500 e più. Entrambe le richieste richiedono una copertura dell’intero universo di imprese.

Alle richieste comunitarie vanno affiancate le esigenze nazionali. Il dettaglio necessario per soddisfare queste ultime potrebbe anche consentire una disaggregazione settoriale meno fine di quella richiesta dal regolamento congiunturale, ma dovrebbe consentire un incrocio con classi di dimensione aziendale e porzioni territoriali (ad esempio il livello ripartizionale).

Ai fini di soddisfare tali esigenze, l’ISTAT ha intrapreso la strada dell’uso dei dati amministrativi. Lo scopo è quello di fornire almeno le variabili occupati alle dipendenze, retribuzione lorda e costo del lavoro per il dominio comprendenti le sezioni da C a K.

La metodologia attualmente in corso di studio per la produzione di indicatori congiunturali richiede l’uso ad ogni periodo, t , (per semplicità poniamo il mese²⁷) di quattro fonti di dati, le prime tre di fonte INPS e l’ultima di fonte ISTAT. Esse sono: un Subset di DM10 di competenza del periodo t (ad esso ci si riferirà nel seguito, in maniera non del tutto propria, con il termine campione), il set completo di DM10 di competenza del periodo $t-12$ (ad esso ci si riferirà nel seguito con il termine universo), l’Anagrafica delle posizioni contributive aggiornata al periodo t , l’ultimo l’Archivio Statistico delle Imprese Attive (ASIA) disponibile. Lo scopo di questo breve documento è quello di fornire alcune informazioni sulle numerosità e sulle strutture delle tre fonti principali: subset mensile, universo ritardato di dodici mesi, e ASIA più recente. Il fine principale è quello di costituire una base per valutare il dettaglio al quale le informazioni possono essere fornite e per effettuare alcune altre scelte (e.g. date di richieste degli scarichi da parte dell’INPS). Rimane esclusa da questa analisi la valutazione della distorsione del campione.

Il capitolo ha la seguente struttura. Il paragrafo 2 illustra le numerosità ed i tempi degli arrivi dei DM10 su supporto magnetico. Il paragrafo 3 descrive il trattamento preliminare dei dati. Il paragrafo 4 riporta le principali strutture dei set di dati. Il paragrafo 5 descrive brevemente la situazione della persistenza delle posizioni nel campione. Il paragrafo 6 riassume i principali risultati.

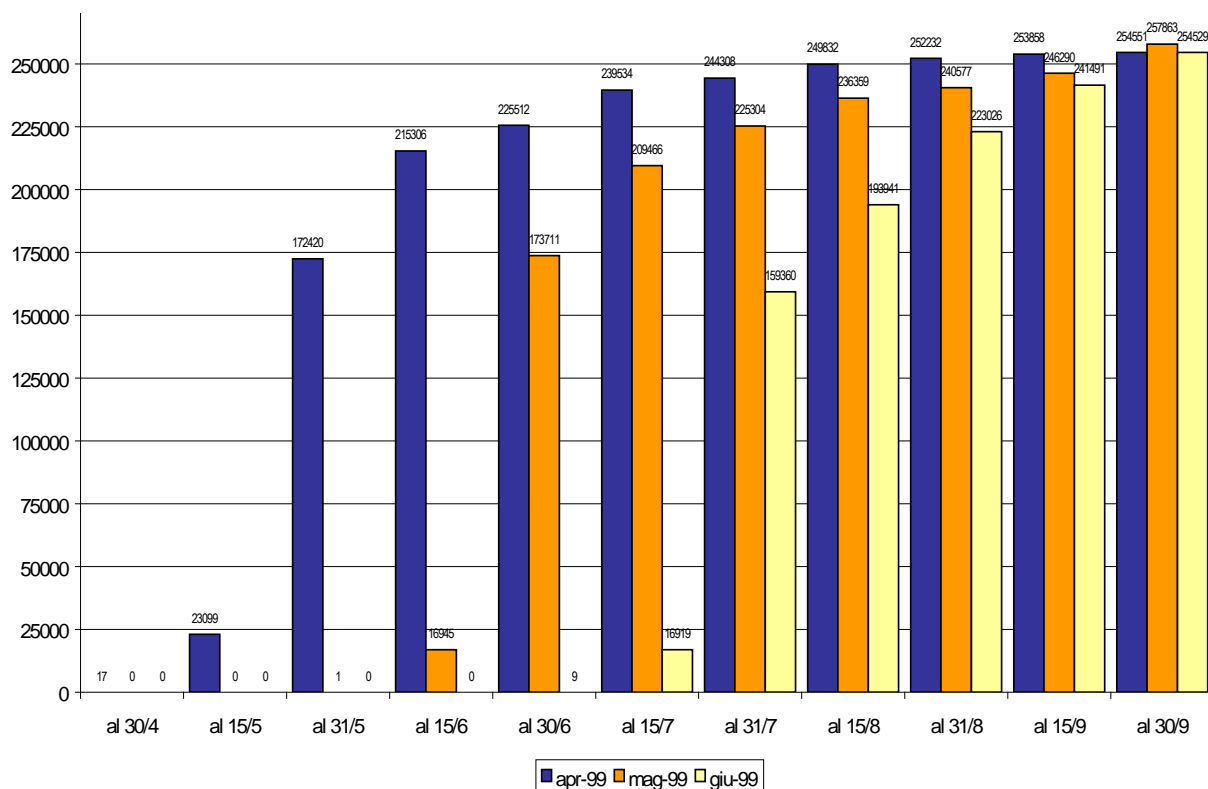
²⁷ Al momento in cui si scrive, però l’ipotesi su cui si lavora è quella di fornire informazioni trimestrali.

7.2 Gli arrivi dei DM10 su supporto magnetico

La rilevazione congiunturale basata sui dati INPS-DM10 può contare “a regime” su un campione di oltre 150000 posizioni contributive. Si tratta, va ricordato, dei DM10 compilati su supporto informatico dalle aziende o dai loro delegati e che quindi arrivano all’INPS con maggiore rapidità²⁸.

Per valutare la dimensione del campione occorre essere a conoscenza dell’intensità dei flussi di DM10 su supporto informatico dalle sedi periferiche dell’INPS alla sede centrale. La conoscenza delle quantità d’arrivo dei DM10 è una preziosa informazione per decidere la cadenza con cui richiedere gli scarichi e la possibilità di effettuare stime congiunturali mensili piuttosto che trimestrali. Una indicazione di questo tipo di fenomeno emerge dal grafico 1. Esso mostra la sequenza di arrivi di DM10 di competenza aprile, maggio e giugno 1999²⁹. Dal grafico si può notare che:

Grafico 1 Arrivi dei DM10 su supporto magnetico di competenza aprile 99, maggio 99, giugno 99



²⁸ I DM10 compilati su supporto informatico arrivano alle sedi locali e sono caricati in locale sul sistema Informatico (AS400). Contestualmente a questa operazione vengono memorizzati i dati necessari alla trasmissione su un archivio dedicato all’ISTAT. Alla accensione successiva del sistema i dati sono trasmessi al centro su due tabelle del database (DB2). Alle informazioni del DM10 vengono aggiunte alcune informazioni dell’archivio anagrafico e trasmesse all’ISTAT su richiesta. I DM10 che sono trasmessi possono avere data di competenza diversa. A causa di questo rapido processo di trasmissione i DM10 subiscono solo alcuni controlli formali. I supporti informatici pervengono prevalentemente alle sedi locali dell’INPS dal 10 al 15 di ciascun mese e le trasmissioni di conseguenza riflettono tali date.

²⁹ L’opportunità di descrivere il trimestre aprile-giugno 1999 deriva dal fatto che esso è il primo trimestre per cui si può considerare opportunamente funzionante la procedura messa in atto dall’INPS.

- 1) La maggioranza dei moduli arriva entro un mese dalla fine del periodo di competenza, in particolare dalla metà alla fine del mese successivo: ad esempio per i DM di competenza di aprile per fine maggio sono arrivati ca. 172000 moduli di cui poco meno di 150000 nella seconda quindicina del mese.
- 2) A 60 giorni dalla chiusura del trimestre, cioè al 31/8, si può contare su un campione riferito a tre mesi, la cui numerosità è di almeno 220000 posizioni contributive per mese.
- 3) La numerosità complessiva dell'insieme di DM10 informatizzati (non riportato nel grafico) è per i tre mesi considerati ca 256500 per aprile, 260700 per maggio, 273200 per giugno.

Va inoltre ricordato che la numerosità del campione tenderà a crescere nel tempo man mano che nuove imprese (direttamente o indirettamente) aderiranno alla compilazione informatizzata del DM10. Al momento risulta difficile fare una stima sul tasso al quale aumenterà tale numerosità, e l'andamento che seguirà questo fenomeno.

7.3 Il trattamento preliminare dei dati

7.3.1 Campioni

Le seguenti analisi sono effettuate su tre campioni. Essi si riferiscono ai DM10 con competenza novembre 1998, dicembre 1998 e gennaio 1999. Sono stati ottenuti raggruppando per data di competenza i DM10 appartenenti a 4 scarichi diversi avvenuti: il 10-01-99, il 17-01-99, il 28-02-99 e ad aprile 99³⁰.

La codifica geografica e quella dell'attività economica sono ottenute effettuando un abbinamento per codice fiscale con ASIA96. In entrambe i casi, ai fini di minimizzare gli errori di classificazione, si attribuiscono ove possibile i codici di ASIA e solo in via residuale i codici INPS. Le ipotesi sotto cui si lavora è che i codici di ASIA siano più corretti. Tali codifiche vanno intese, quindi, come georeferenziazione e referenziazione di attività economica dell'impresa a cui appartiene la posizione contributiva. Per il campione di dicembre i due codici sono attribuiti quindi per 182141 posizioni attraverso ASIA e per 40666 attraverso la codifica INPS.

Il codice usato per referenziare geograficamente le posizioni che non si abbinano con ASIA è il codice sede provinciale. Esso indica il luogo in cui l'unità INPS denuncia e versa i contributi. Tale provincia può quindi non coincidere con la sede di produzione dell'unità. In particolare, va segnalato che alcune imprese sono autorizzate al cosiddetto accentramento contributivo per cui posizioni contributive differenti possono versare i contributi in una unica sede provinciale. Dato che il codice provincia è presente per ogni posizione INPS si riesce ad attribuire il codice regione a tutte le posizioni contributive. Per avere una misura di possibile misclassificazione di queste ultime calcoliamo il rapporto di posizioni classificate diversamente in ASIA e in INPS sul totale di 182141 posizioni per cui è possibile effettuare questa operazione. Sul campione di dicembre, a livello regionale questo indice è pari a 1.1%, mentre a livello di macro area (Nord-Ovest, Nord-Est, Centro, Sud, Isole) a meno dell'1%.

7.3.2 Attribuzione del codice di attività economica e misclassificazione

L'INPS di recente sta implementando il codice ATECO91 per le posizioni contributive. I file di cui stiamo trattando sono infatti forniti di tale codice. E' questo il codice che viene usato per attribuire

³⁰ Nel periodo a cui ci stiamo riferendo l'INPS ha avuto problemi relativi alle procedure di scarico dei dati che hanno implicato notevoli discontinuità negli scarichi. Tali problemi sono sorti anche in relazione al fatto che era la prima volta che si sperimentava questa procedura messa in piedi proprio per le esigenze dell'ISTAT. Questo genere di problemi dovrebbe essere ormai superati e i successivi campioni non dovrebbero risentire di tali discontinuità. Per questi motivi, comunque, si è preferito usare tutti i DM10 di competenza dei tre mesi in oggetto piuttosto che simulare la situazione a "regime" selezionando sulle date di trasmissione.

la divisione laddove non è possibile farlo attraverso ASIA³¹. Per avere una misura di possibile misclassificazione calcoliamo sul subset di posizioni per cui disponiamo sia dell'ATECO di ASIA che di quello INPS la percentuale di posizioni per le quali le due classificazioni non coincidono a livello di due cifre e a livello di una cifra. I due indicatori sono pari rispettivamente a 21% e 13.5%.

La variabile relativa ai dipendenti considerata è il numero dipendenti riportata nel quadro A (al momento in cui si scrive permane infatti qualche incertezza sulla variabile dipendenti ricostruita dal quadro B). Essa non risulta, però, completamente affidabile. I motivi sono:

- 1) I campioni non sono soggetti a controllo da parte dell'INPS, se non ad alcuni controlli puramente formali.
- 2) La variabile in questione è parte del quadro conoscitivo generale del modulo. E' probabile che le imprese abbiano una minore attenzione alla qualità dell'informazione introdotta.
- 3) Al momento in cui si scrive, inoltre, non è stata ancora completata una procedura di controllo e correzione dati.

Per ovviare a quest'ultimo problema i campioni sono sottoposti ad una sommaria operazione di pulizia dei dati. Lo scopo di tale operazione è solo quello di eliminare i probabili errori più evidenti. In pratica si è provveduto ad eliminare le osservazioni che presentavano scostamenti anomali tra un mese e l'altro. Il confronto longitudinale non è ovviamente possibile per le osservazioni presenti solo in uno dei mesi. In aggiunta, quindi, vengono inoltre eliminate le posizioni contributive che presentano scostamenti anomali tra la variabile di riferimento e la somma di operai ed impiegati ricavabili dal quadro B. Il numero di posizioni eliminate con queste operazioni è ridottissima, inferiore alle 30 unità per mese.

In questa fase è effettuata una riduzione dei campioni alle posizioni con sezione di attività economica in una delle seguenti: C,D,E,F,G,H,I,J,K. La numerosità dei campioni e la loro riduzione dopo l'esclusione delle posizioni anomale e la selezione sull'attività economica è riportata nella tabella 1.

Tabella 1 - La riduzione dei campioni

Mese di competenza del DM10	Numero di posizioni contributive	N° di posizioni C-K
Novembre 1998	224232	194413
Dicembre 1998	222807	193098
Gennaio 1999	209855	181819

D'ora in avanti si considerano i campioni relativi al dominio di interesse (sezioni da C a K).

7.3.3 L'Universo INPS

L'universo dei DM10 relativi al dicembre 1997 è ottenuto selezionando dall'universo Anagrafico aggiornato al febbraio 1999 le posizioni in vita nel periodo di riferimento.

Le procedure di assegnazione dei Codici Regione e Codici di Attività Economica ricalcano quelle effettuate per il campione. L'unica, sostanziale, differenza riguarda il codice di attività economica dell'INPS usato residualmente rispetto a quello di ASIA. Il file dell'INPS in questione, infatti, rispetto al campione non contiene l'ATECO91. Il codice in questione viene allora ottenuto trascodificando il Codice Statistico Contributivo e il codice ATECO81.

Come per il campione la selezione dell'universo è effettuata considerando tutte le posizioni tra C e K con almeno un dipendente³².

³¹ Ricordiamo inoltre che alle posizioni contributive è assegnato anche dall'INPS il Codice statistico Contributivo (CSC) che individua il settore di attività economica. Tuttavia la trascodifica di questo codice con l'ATECO91 non è univoca ed è spesso fallace. In quanto segue questo codice non è usato per i campioni, ma è usato per l'universo INPS.

³² Va però notato che nel set di dati considerato sono presenti oltre 200000 posizioni con zero dipendenti, pur

7.4 Confronti tra le strutture

Nel seguito viene presentato il confronto tra le strutture dei set di dati di riferimento. Ai fini di fornire una ulteriore variabile di confronto viene presentata anche la struttura di ASIA. Si tratta del sottoinsieme di imprese che in ASIA 96 hanno un numero di dipendenti a dicembre strettamente maggiore di 0 e che appartengono al dominio C-K.

Va quindi tenuto in considerazione che la struttura riguarda posizioni contributive per le fonti INPS e imprese per quanto riguarda ASIA.

Il campione considerato è quello riferito a Dicembre 98. Esso contiene 193086 posizioni con attività economica da C a K per un totale di 1574712 dipendenti. Esso rappresenta il 20.42% delle posizioni presenti nella stesso dominio un anno prima e il 18.88% in termini di dipendenti.

Tabella 2 – Struttura per sezione ATECO91: posizioni e imprese

Sezione	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. posizioni (a)	%	N. posizioni (b)	%	N. imprese	%	
C	430	0.22	2870	0.30	2661	0.27	14.98
D	64005	33.15	272219	28.79	284163	28.68	23.51
E	128	0.07	1277	0.14	987	0.10	10.02
F	30889	16.00	140579	14.87	158780	16.02	21.97
G	49324	25.55	253246	26.78	279660	28.22	19.48
H	11143	5.77	66101	6.99	89095	8.99	16.86
I	8284	4.29	38904	4.11	37420	3.78	21.29
J	3738	1.94	19618	2.07	19172	1.93	19.05
K	25145	13.02	150805	15.95	118927	12.00	16.67
Totale	193086	100	945619	100	990865	100	20.42

Tabella 3 – Struttura per sezione ATECO91: dipendenti

Sezione	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. dipendenti (a)	%	N. dipendenti (b)	%	N. dipendenti	%	
C	3128	0.20	25644	0.31	32125	0.37	12.20
D	667985	42.42	3919242	46.98	4041671	46.28	17.04
E	33666	2.14	151188	1.81	160677	1.84	22.27
F	136044	8.64	728710	8.73	758796	8.69	18.67
G	269450	17.11	1252786	15.02	1252290	14.34	21.51
H	52925	3.36	330751	3.96	384868	4.41	16.00
I	69233	4.40	545157	6.53	891683	10.21	12.70
J	117969	7.49	507184	6.08	481348	5.51	23.26
K	224312	14.24	882128	10.57	729249	8.35	25.43
Totale	1574712	100	8342790	100	8732707	100	18.88

Le tabelle 2 e 3 presentano la struttura in termini di sezioni dei tre set di dati. Nell'ultima colonna delle tabelle è mostrato la quota rispettivamente di posizioni e dipendenti del campione rispetto all'Universo. Al minimo sono presenti ca. il 10% delle posizioni e oltre il 12.20 dei dipendenti. Le sezioni relativamente meno rappresentate sono le sezioni C ed E in termini di posizioni e le sezioni C e I in termini di dipendenti.

presentando un codice di stato di attività che non segnala né la sospensione né la cessazione della posizione. Si tratta di due insiemi di casi: quello delle posizioni iscritte da un tempo limitato e che ancora devono cominciare l'attività con dipendenti; quello delle posizioni per cui esiste un ritardo nella registrazione della sospensione o della cessazione.

Tabella 4 – Dimensioni medie a confronto

Sezione	N. dipendenti per posizione	
	Campione INPS Dic98	Universo INPS Dic97
C	7.27	8.94
D	10.44	14.40
E	263.02	118.39
F	4.40	5.18
G	5.46	4.95
H	4.75	5.00
I	8.36	14.01
J	31.56	25.85
K	8.92	5.85
Totale	8.16	8.82

La tabella 4 presenta il confronto delle dimensioni medie in termini di dipendenti delle posizioni presenti nel campione e di quelle presenti nell'universo. In totale la dimensione media delle posizioni nel campione è di 8.16 contro i 8.82 dell'universo INPS.

Tabella 5 – Struttura per sottosezione ATECO91: posizioni e imprese

Sottosez.	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. posizioni (a)	%	N. posizioni (b)	%	N. imprese	%	
CA	2	0.00	27	0.00	20	0.00	7.41
CB	428	0.22	2843	0.30	2641	0.27	15.05
DA	6540	3.39	35260	3.73	36048	3.64	18.55
DB	9271	4.80	39657	4.19	43050	4.34	23.38
DC	3827	1.98	14347	1.52	15410	1.56	26.67
DD	3323	1.72	15403	1.63	17174	1.73	21.57
DE	3651	1.89	15433	1.63	16072	1.62	23.66
DF	52	0.03	461	0.05	353	0.04	11.28
DG	947	0.49	4384	0.46	4059	0.41	21.60
DH	2203	1.14	9229	0.98	9335	0.94	23.87
DI	2918	1.51	14844	1.57	15162	1.53	19.66
DJ	13682	7.09	50597	5.35	53778	5.43	27.04
DK	5939	3.08	22887	2.42	22999	2.32	25.95
DL	5593	2.90	21957	2.32	22086	2.23	25.47
DM	774	0.40	3844	0.41	3645	0.37	20.14
DN	5285	2.74	23916	2.53	24992	2.52	22.10
E	128	0.07	1277	0.14	987	0.10	10.02
F	30889	16.00	140579	14.87	158780	16.02	21.97
G	49324	25.55	253246	26.78	279660	28.22	19.48
H	11143	5.77	66101	6.99	89095	8.99	16.86
I	8284	4.29	38904	4.11	37420	3.78	21.29
J	3738	1.94	19618	2.07	19172	1.93	19.05
K	25145	13.02	150805	15.95	118927	12.00	16.67
Totale	193086	100	945619	100	990865	100	20.42

Tabella 6 – Struttura per sottosezione ATECO91: dipendenti

Sottosez.	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. dipendenti (a)	%	N. dipendenti (b)	%	N. imprese	%	
CA	13	0.00	1414	0.02	7686	0.09	0.92
CB	3115	0.20	24230	0.29	24439	0.28	12.86
DA	49574	3.15	308856	3.70	335004	3.84	16.05
DB	90196	5.73	544577	6.53	574458	6.58	16.56
DC	41127	2.61	183036	2.19	192122	2.20	22.47
DD	17643	1.12	98107	1.18	101007	1.16	17.98
DE	38003	2.41	201448	2.41	212595	2.43	18.86
DF	597	0.04	25758	0.31	26964	0.31	2.32
DG	27540	1.75	200688	2.41	206478	2.36	13.72
DH	29752	1.89	181102	2.17	177377	2.03	16.43
DI	25252	1.60	203658	2.44	209200	2.40	12.40
DJ	125560	7.97	604737	7.25	609684	6.98	20.76
DK	76678	4.87	478682	5.74	493839	5.66	16.02
DL	69303	4.40	379366	4.55	382380	4.38	18.27
DM	31705	2.01	275031	3.30	289223	3.31	11.53
DN	45055	2.86	234196	2.81	231340	2.65	19.24
E	33666	2.14	151188	1.81	160677	1.84	22.27
F	136044	8.64	728710	8.73	758796	8.69	18.67
G	269450	17.11	1252786	15.02	1252290	14.34	21.51
H	52925	3.36	330751	3.96	384868	4.41	16.00
I	69233	4.40	545157	6.53	891683	10.21	12.70
J	117969	7.49	507184	6.08	481348	5.51	23.26
K	224312	14.24	882128	10.57	729249	8.35	25.43
Totale	1574712	100	8342790	100	8732707	100	18.88

Per le specifiche esigenze del Regolamento Congiunturale le tabelle 5 e 6 riportano la struttura per la disaggregazione settoriale per sottosezione. Risultano particolarmente sguarnite la sottosezione CA in termini di posizioni e dipendenti e la DF in termini di dipendenti.

Tabella 7 – Struttura per classe dimensionale: posizioni e imprese

Classe di dipendenti	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. posizioni (a)	%	N. posizioni (b)	%	N. imprese	%	
Fino a 19	181993	94.25	884906	93.58	931950	94.05	20.57
20-99	10219	5.29	52448	5.55	50828	5.13	19.48
100-499	766	0.40	7247	0.77	7046	0.71	10.57
500 e oltre	108	0.06	1018	0.11	1041	0.11	10.61
Totale	193086	100	945619	100	990865	100	20.42

Tabella 8 – Struttura per classe dimensionale: dipendenti

Classe di dipendenti	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. dipendenti (a)	%	N. dipendenti (b)	%	N. imprese	%	
Fino a 19	719499	45.69	3257732	39.05	3293368	37.71	22.09
20-99	362397	23.01	1965357	23.56	1895094	21.70	18.44
100-499	132734	8.43	1390502	16.67	1351119	15.47	9.55
500 e oltre	360082	22.87	1729199	20.73	2193126	25.11	20.82
Totale	1574712	100	8342790	100	8732707	100	18.88

Nelle tabelle 7 e 8 la struttura è riportata in termini di classi di dipendenti con la modalità consigliata dal manuale del Regolamento Congiunturale. Appare sottorappresentata rispetto

all'Universo la classe 100-499 in termini di dipendenti.

Tabella 9 – Struttura per ripartizione: posizioni e imprese

Area	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. posizioni (a)	%	N. posizioni (b)	%	N. imprese	%	
Nord-Ovest	78532	40.67	315686	33.38	315284	31.82	24.88
Nord-Est	56378	29.20	231679	24.50	246700	24.90	24.33
Centro	25050	12.97	190412	20.14	203651	20.55	13.16
Sud	30986	16.05	141799	15.00	154046	15.55	21.85
Isole	2140	1.11	66043	6.98	71184	7.18	3.24
Totale	193086	100	945619	100	990865	100	20.42

Tabella 10 – Struttura per ripartizione: dipendenti

Area	Campione INPS Dic98		Universo INPS Dic97		ASIA 96 (Dicembre)		(a/b)*100
	N. dipendenti (a)	%	N. dipendenti (b)	%	N. dipendenti	%	
Nord-Ovest	770878	48.95	3394697	40.69	3306068	37.86	22.71
Nord-Est	416973	26.48	2080686	24.94	2086749	23.90	20.04
Centro	197202	12.52	1633658	19.58	2063270	23.63	12.07
Sud	176615	11.22	882491	10.58	911644	10.44	20.01
Isole	13044	0.83	351258	4.21	364976	4.18	3.71
Totale	1574712	100	8342790	100	8732707	100	18.88

Le tabelle 9 e 10 presentano la struttura per area geografica. La nota più importante riguarda la presenza molto ridotta nel campione delle Isole rispetto all'universo. Anche il centro presenta percentuali di copertura notevolmente inferiori a quelle medie.

7.5 La permanenza delle posizioni nei campioni

La situazione di permanenza delle posizioni contributive nei campioni di mesi differenti è illustrato nella tabella 10.

Tabella 11 – Il panel delle posizioni contributive nel campione

Presente a Novembre98	Presente a Dicembre98	Presente a Gennaio99	N. Posizioni contributive	% sulle posizioni presenti in almeno un mese
Si	Si	Si	152479	68.3
Si	Si	No	34800	15.6
Si	No	Si	1520	0.7
Si	No	No	5614	2.5
No	Si	Si	4718	2.1
No	Si	No	1101	0.5
No	No	Si	23102	10.3
Totale			223334	100.0

152479 posizioni contributive pari al 68.3% delle posizioni presenti in almeno uno dei tre mesi sono presenti in tutti e tre i campioni. Esiste così un'elevata persistenza dei DM10. Tra dicembre e gennaio scompaiono 34800 posizioni. Rispetto a dicembre entrano a gennaio 24622 posizioni.

Le date di costituzione delle posizioni entrate a gennaio sono riportate nella tabella 11. Come si nota solo l'8.5% di esse si possono considerare "nuove nate" nel mercato in base alla data di costituzione. Il rimanente 90% si riferisce invece a due possibilità: 1) accesso alla modalità telematica di posizioni precedentemente esistenti ma non presenti nel campione o 2) riemersione di

posizioni nel campione che avevano sospeso l'attività a dicembre 98. Ad esempio 1520 posizioni sono presenti nel campione di novembre, "scompaiono" a dicembre e ricompaiono a gennaio.

Tabella 12 – Nascite ed entrate nel campione

Data di costituzione	N. di posizioni	%
Entro il 31/12/1995	16685	67.8
1996	1454	5.9
1997	1836	7.5
I Trimestre 1998	667	2.7
II Trimestre 1998	613	2.5
III Trimestre 1998	457	1.9
IV Trimestre 1998	826	3.4
Dal 1/1/1999	2084	8.5
Totale	24622	100

7.6 Conclusioni

Lo scopo di questo capitolo è quello di presentare le dimensioni e le strutture dei set di dati che costituiscono la base per la rilevazione trimestrale su Occupazione Retribuzione e Costo del Lavoro basate sui dati INPS.

I principali dati presentati possono essere così schematizzati:

- 1) Attualmente la dimensione del campione a 30 giorni dalla chiusura del mese di competenza è di oltre 170000 posizioni contributive. A 60 giorni dalla chiusura di trimestre si può contare su un set di tre campioni mensili la cui dimensione minima è di circa 220000 posizioni.
- 2) Per il dominio di interesse (sezioni da C a K dell'ATECO91) gli invii informatizzati dovrebbero consentire di contare a 60 giorni dalla chiusura del trimestre su almeno 180000 posizioni per mese.
- 3) La struttura per sezione del campione in termini sia di posizioni che di dipendenti pur presentando distorsioni rispetto a quella dell'Universo di DM10 consente una copertura di almeno il 10%. Al livello di sottosezione la struttura del campione denuncia relativamente all'universo alcune coperture scarse. Il loro numero appare tuttavia molto limitato.
- 4) Un problema di numerosità relativa è presente per le zone del centro e delle isole. Per queste aree, infatti, il campione contiene una percentuale di posizioni e dipendenti sostanzialmente inferiore a quella dell'universo.
- 5) La componente panel del campione nei tre mesi successivi, presi in considerazione, è pari ad oltre 152000 posizioni pari a circa il 68% del totale delle posizioni presenti in almeno uno dei mesi. Il 10% circa di questo totale rappresenta le posizioni entrate solo nell'ultimo campione a disposizione. Solo però il 10% circa di queste nuove entrate possono essere considerate nuove nate.

I risultati riportati in questo lavoro indicano quindi che la rilevazione può contare su un subset di dati di dimensioni ragguardevoli e sufficientemente ben strutturato dal punto di vista della ripartizione settoriale. La analisi descrittiva effettuata rimane limitata da almeno due punti di vista. Il primo riguarda le diverse procedure di codifica dell'attività economica che si sono dovute seguire per il campione e l'universo dei dati. La seconda riflette l'attuale situazione informativa ed è relativa al fatto che i dati del campione sono rapportati a quelli dell'universo di un anno prima, non ad un universo contemporaneo.

Due questioni principali rimangono aperte. La prima riguarda la capacità del campione di rappresentare l'universo in termini di variazione delle variabili di interesse. In altri termini occorre rispondere alla domanda: quanto distorta risulterebbe una stima basata sul campione come se questo

fosse estratto con un disegno campionario semplice? La seconda questione riguarda la capacità del campione di contenere informazioni sui movimenti demografici. A entrambe queste domande la attuale situazione informativa non consente di rispondere.

8 CONCLUSIONI PROVVISORIE: LE PRINCIPALI CARATTERISTICHE DELLA RILEVAZIONE E I PROBLEMI APERTI

8.1 Le difficoltà nella progettazione del disegno della rilevazione

E' importante ricordare che per l'Istat questa è la prima esperienza nel campo dell'uso dei dati amministrativi come fonte per la produzione di statistiche congiunturali. Infatti sino ad oggi l'uso dei dati amministrativi per le imprese ha trovato applicazione soltanto nella costruzione dell'archivio ASIA. Inoltre ci si è trovati di fronte a un problema più generale: la scarsa esperienza dell'Istat sullo sviluppo e l'applicazione di metodologie adatte a rilevazioni congiunturali sulle imprese. Infatti, tutte le rilevazioni congiunturali sulle imprese condotte dall'Istat (produzione industriale, fatturato, vendite, ecc.) sono caratterizzate da limitazioni e problemi metodologici di varia natura:

- si riferiscono a sottoinsiemi della popolazione, in genere vengono trascurate completamente le piccole e medie imprese;
- vengono quasi sempre effettuate su panel chiusi di imprese di durata fino 5 anni, trascurando le nuove imprese;
- rilevano soltanto variazioni e mai valori assoluti.

In passato soltanto una indagine è stata progettata e realizzata per superare alcuni di questi problemi: la "Rilevazione su occupazione, ore lavorate e retribuzioni" denominata LPS001 gestita dall'Istat in collaborazione con il Ministero del Lavoro. Questa indagine, condotta per mezzo postale su oltre 18.000 unità locali, tuttavia, non ha condotto ai risultati attesi. Tale esperienza, dopo un lungo periodo di progettazione(1990-1992) e tre anni di rilevazione (1993-1994-1995), con una forte molestia statistica per le imprese, è stata interrotta senza alcun rilascio dei dati.

L'utilizzo dei DM10 come fonte di informazioni corrente su occupazione e retribuzioni nasce proprio dal fallimento di (e in un certo senso come alternativa) una rilevazione diretta di tipo postale come la LPS001.

In sostanza quando è stata intrapresa la progettazione della rilevazione non ci si è scontrati soltanto con le novità e le difficoltà dell'uso dei dati amministrativi INPS, ma anche con tipici problemi di rilevazioni congiunturali sulle imprese mai affrontati e risolti in precedenza nell'Istituto.

La difficoltà maggiore è stata, forse, quella di dover coprire tutte le classi dimensionali d'impresa, elemento essenziale per le stime sull'occupazione.

Altri tre elementi hanno condizionato fortemente la progettazione della rilevazione:

1. sin dall'inizio si è posta la necessità di costruire uno strumento che avesse numerosi e diversificati obiettivi;
2. i tempi di progettazione e realizzazione sono stati piuttosto stretti (in relazione alle risorse umane investite) poiché limitati da obiettivi comunitari da raggiungere entro date prestabilite;
3. il disegno iniziale è stato progettato senza l'ausilio concreto dei dati INPS. La disponibilità dei dati nelle forme effettive disponibili a regime (in termini di variabili, tracciati record standard, qualità dei dati, ecc.) sono a disposizione del gruppo di lavoro solo da pochi mesi. Infatti alcune sperimentazioni sono state svolte su dati e file che non rappresentavano effettivamente la situazione informativa possibile a regime.

In definitiva al contrario di una indagine tradizionale in cui si sceglie a priori uno strumento a seconda dell'obiettivo da raggiungere, nel caso in questione soltanto attraverso l'uso e l'apprendimento graduale delle potenzialità, dei difetti e dei costi di trattamento dei dati INPS è possibile pervenire a scelte strategiche in termini di disegno dell'indagine. A tal proposito non è stato possibile progettare un disegno completo dell'indagine a priori ma soltanto una ipotesi di

lavoro che viene rivista in corso d'opera sulla base dell'esperienza concreta e dei risultati attraverso un processo di tipo *trial and error*.

Tutte le scelte metodologiche proposte vanno discusse e verificate alla luce dell'esperienza empirica, della qualità dei risultati e dell'analisi costi-benefici sugli strumenti utilizzati (inclusi i possibili costi di elaborazione, gestione, ecc.).

8.2 Sintesi delle principali caratteristiche della rilevazione

La rilevazione congiunturale sull'occupazione dipendente, sulle retribuzioni e sugli oneri sociali è stata progettata al fine di colmare un gap fondamentale nelle statistiche nazionali e soddisfare esigenze comunitarie, in primo luogo il LCI. L'obiettivo è quello di fornire informazioni congiunturali sulle variazioni (e se possibile sui livelli) trimestrali dell'occupazione dipendente, delle retribuzioni lorde e degli oneri sociali (costo del lavoro). Nella prima fase si tralascerà il progetto di stimare le ore effettivamente lavorate. Le variabili da rilevare sono, quindi, l'occupazione dipendente, le retribuzioni lorde, gli altri oneri sociali per settori di attività economica, dimensione d'impresa e, possibilmente, ripartizioni territoriali. Gli strumenti su cui si basa la rilevazione sono due: per le imprese sotto i 500 addetti si fa riferimento ai dati amministrativi tratti dai modelli DM10 INPS, per le imprese oltre i 500 addetti le informazioni derivate da fonte INPS vengono integrate dalla rilevazione mensile sulle grandi imprese (GI). Inoltre, come per tutte le rilevazioni Istat sulle imprese, il frame di riferimento è l'archivio ASIA.

Popolazione obiettivo (Target population)

La popolazione obiettivo della rilevazione è costituita dall'insieme dei datori di lavoro (imprese, cooperative, ecc.) operanti sul territorio nazionale con l'eccezione di quelli che svolgono attività in agricoltura, caccia e pesca (sezioni A e B), nei servizi sociali e riservati alle famiglie (sezioni M, N, O e P), nella Pubblica Amministrazione e difesa (sezione L).

Popolazione di riferimento della rilevazione (Survey population)

Per ottenere le stime delle variabili si fa innanzitutto riferimento all'archivio statistico delle imprese attive (ASIA). ASIA viene aggiornato ogni anno, in teoria con 18 mesi di ritardo rispetto alla data di riferimento. Attualmente è disponibile la versione relativa all'anno 1996. In aggiunta si ricorre all'archivio delle imprese INPS desunto dai moduli DM10 con cui le imprese dichiarano l'attività svolta con dipendenti all'istituto di previdenza.

Sottocopertura

I dati di fonte INPS non consentono di rilevare alcune tipologie di imprese (aziende pubbliche o di origine pubblica) che per definizione non hanno obblighi contributivi con l'INPS (esempio FFSS, ecc.). Tali imprese sono tuttavia rilevate dalla indagine GI.

Sovracopertura

Nei dati di fonte INPS, per errore, può essere rilevata qualche impresa che svolge attività in settori non inclusi nel campo di osservazione della rilevazione. Ad esempio imprese che effettivamente operano nel settore M,N,O e che vengono erroneamente classificate nei settori da C a K.

Unità statistica di analisi

L'unità statistica di analisi è l'impresa definita secondo l'archivio ASIA. In teoria le statistiche in questione dovrebbero riferirsi, in accordo con le definizioni e le metodologie prevalenti indicate da EUROSTAT, all'unità funzionale, tuttavia l'assenza di informazioni su tali unità nei dati INPS non lascia altra scelta che ricorrere, quale unità di analisi all'impresa alla cui individuazione si giunge al seguito di una operazione di ricostruzione delle posizioni contributive l'unità statistica di rilevazione per la parte INPS. Ovviamente utilizzare l'impresa invece che l'unità funzionale comporta di

allocare tutta l'attività dell'impresa esclusivamente nel settore di attività economica principale. La rilevazione mensile GI, tranne alcune eccezioni, consente di riferirsi alle unità funzionali.

Unità statistica di rilevazione

Ogni impresa che svolge la propria attività utilizzando lavoro dipendente deve aprire presso l'INPS almeno una "posizione contributiva" caratterizzata da un numero di matricola. Tale posizione è necessaria per effettuare i versamenti e le dichiarazioni mensili degli obblighi contributivi tramite il modulo DM10. La posizione contributiva non necessariamente costituisce una unità funzionale, né una unità locale. La posizione nella gran parte dei casi (circa il 97%) corrisponde all'impresa; negli altri casi a parti di impresa, Per giungere all'unità di analisi "impresa" è necessario effettuare una operazione di ricostruzione attraverso una complessa procedura. Nella rilevazione mensile GI l'unità di rilevazione è l'unità funzionale delle imprese con oltre 500 addetti.

Periodo di riferimento

Il riferimento temporale di base dell'indagine è il mese. I dati INPS disponibili sono infatti mensili e le variabili forniscono informazioni sul mese di riferimento. La rilevazione GI riporta informazioni mensili. Tali informazioni mensili devono successivamente essere aggregate per il rilascio pubblico dei dati espressi in trimestri.

Metodologia

Le statistiche da pubblicare dovrebbero essere prodotte sulla base dell'integrazione della stima anticipata (del censimento) dei dati amministrativi INPS-DM10 e dalla rilevazione postale mensile sulle grandi imprese (GI).

Il metodo suggerito, che utilizza sia l'informazione disponibile negli archivi di fonte INPS sia di fonte ASIA costruisce le stime dei parametri d'interesse attraverso una stime aggiornate al tempo corrente (o stime anticipate) utilizzando la fonte INPS riferita all'universo a t-12 e allo pseudo-campione al tempo t. Le stime sono calcolate sulla base di un modello statistico di tipo regressivo che lega le variabili d'interesse (gli occupati, la retribuzione e il costo del lavoro) a variabili ausiliarie, di cui si può conoscere la distribuzione nell'universo delle imprese dagli archivi INPS; Per la stima del mese t si utilizza uno "pseudo-campione" non rappresentativo di DM10 che è costituito dall'insieme dei modelli che sono compilati dalle imprese su supporti magnetici (circa 160.000) riferito al mese t e l'universo dei DM10 INPS, disponibile sempre al tempo t ma riferito a t-12.

I dati Amministrativi INPS

Il ricorso all'uso dei dati amministrativi contribuisce alla riduzione dei costi di una eventuale indagine alternativa o all'estensione di quella riferita alle grandi imprese e soprattutto all'aggravio della "molestia statistica" per le imprese. Una convenzione quadro fra Istat e INPS permette all'istituto di statistica di accedere alle informazioni grezze sui DM10 mensili che riportano dati sul numero totale di occupati a libro paga nel mese di riferimento, sul monte retributivo lordo, sul totale dei contributi (debito a carico del datore di lavoro) del mese di riferimento. Le imprese possono effettuare le dichiarazioni contributive seguendo modalità diverse il cui ricorso dipende esclusivamente dalle scelte di convenienza da parte delle imprese: compilando un solo modulo per tutti i dipendenti (posizione contributiva singola) ovvero utilizzando più moduli ((posizione contributive multiple).

La rilevazione sulle grandi imprese

La rilevazione mensile sull'occupazione, gli orari di lavoro e le retribuzioni nelle grandi imprese (in sintesi GI) si presenta come una tipica rilevazione esaustiva di tipo longitudinale (panel), cioè le stesse unità statistiche vengono rilevate ogni mese. Il campo di osservazione è definito dalle imprese

che, in un qualsiasi momento, acquisiscono o mantengono lo status di imprese di "grandi dimensioni" cioè hanno un numero di addetti superiore a 500 unità. Si tratta, quindi, di un panel per così dire "aperto". Sono interessate alla rilevazione tutte le imprese dell'industria (esclusa l'edilizia) e dei servizi che svolgono la loro attività nei settori C-K. Le informazioni rilevate riguardano distintamente per le qualifiche impiegate e non impiegate:

- i dipendenti alla fine del mese di riferimento e del mese precedente;
- le retribuzioni lorde (a carattere continuativo e saltuario o occasionale);
- gli oneri sociali a carico del datore di lavoro (contributi sociali e provvidenze al personale);
- gli accantonamenti per il trattamento di fine rapporto (TFR).

Raccolta dei dati

Attualmente i dati amministrativi INPS cioè i moduli DM10 vengono trasferiti su cassette dall'INPS in formato grezzo originale. Ciò significa che non sono sostanzialmente trattati dall'INPS ma trasferiti all'Istat come se provenissero direttamente dalle imprese. Essi vengono quindi trattati ed elaborati all'interno dell'Istat. I modelli della rilevazione GI, invece sono acquisiti dall'Istat direttamente dalle imprese, tramite fax o posta e i dati vengono registrati con revisione controllata da software.

Controllo e correzione dei dati

Una volta acquisiti, i dati INPS vengono sottoposti a delle procedure di controllo e correzione che valutano l'esattezza e la consistenza. Le procedure di editing includono il confronto dei valori delle caratteristiche chiave (ad esempio numero di dipendenti, i monti retributivi, ecc.) per 3 mesi successivi, per rilevare eventuali grandi variazioni che potrebbero rappresentare degli errori.

Nella rilevazione GI attualmente il check viene effettuato sui singoli modelli e per le singole variabili oggetto di studio. Il controllo viene effettuato confrontando i dati relativi alla retribuzione pro capite continuativa e non, alle ore lavorate pro capite, alla percentuale di oneri sociali pagati dall'impresa sia con quelli relativi all'anno precedente nello stesso mese della rilevazione, sia con quelli del mese precedente. Viene, inoltre, effettuato un check sui singoli modelli per controllare che gli addetti del mese di rilevazione, ad inizio mese coincidano con gli addetti rilevati nel mese precedente.

Mancate risposte e metodi di imputazione per i dati INPS

A causa di ritardi di aggiornamento l'universo dei dati INPS disponibili a 12 mesi non è effettivamente completo a causa di mancate risposte totali. A tal proposito in primo luogo è necessario identificare le effettive mancate risposte dalle assenze (corrette) dei DM10 dovute a inattività temporanea dell'impresa (stagionalità) o cessazione non registrata negli archivi INPS. Occorre distinguere in sostanza le imprese attive da quelle non attive nel periodo di riferimento. In secondo luogo imputare a livello micro i dati. L'individuazione delle imprese (posizioni contributive) attive avviene attraverso il calcolo di una probabilità di essere impresa attiva stimata sulla base di un modello probit (o logit) applicato sui dati nei periodi precedenti. L'imputazione dei dati avviene con dei modelli che utilizzano nello stesso tempo informazioni in serie storica e cross-section.

Errori campionari

L'errore campionario, che riguarda solo le stime preliminari, è dovuto alla natura stessa delle osservazioni INPS utilizzate a questo livello di analisi, poiché ottenute da un campione piuttosto che dall'intero universo. Le stime basate su un campione possono differire da quelle che sarebbero state ottenute da un censimento completo (che è disponibile solo a 12 mesi per la stima definitiva). L'errore campionario della stima, può essere valutato sul campione utilizzando l'errore standard o meglio il coefficiente di variazione relativo soltanto alle caratteristiche osservabili (cioè le variabili

su cui si calcolano le celle di stima per la regressione). Il fatto che il campione INPS non sia estratto casualmente implica un ulteriore problema. E' necessario valutare la possibilità che il campione sia distorto per caratteristiche non osservate della popolazione di riferimento. Ad esempio che le imprese che inviano i dati su supporto magnetico all'INPS siano più dinamiche in termini di crescita dell'occupazione perché più informatizzate e innovative. L'ampiezza dell'errore campionario per sé e della "non rappresentatività" del campione possono essere valutati completamente ex-post sulla base delle revisioni quando sono disponibili i dati censuari.

Errori non campionari

I problemi non campionari più importanti dei dati INPS riguardano:

1. la definizione dell'unità di rilevazione (posizione contributiva);
2. la classificazione dell'attività economica (CSC e non ATECO 91);
3. i contenuti delle variabili (occupazione, retribuzione, oneri sociali).

Indici a base fissa delle retribuzioni e degli oneri sociali

Oltre a produrre dei valori assoluti per alcune variabili si intende rilasciare degli indici con base fissa e ponderazione fissa all'anno base (tipo Laspeyres). Tali indici vanno utilizzati per tenere conto dell'effetto dei cambiamenti nella composizione di occupazione fra i settori di attività economica, le classi dimensionali (eventualmente le qualifiche e le ripartizioni) Questi indici forniscono certamente migliori indicazioni sulle tendenze sottostanti le semplici variazioni nei valori assoluti delle retribuzioni e del costo del lavoro pro capite. L'anno base, coerentemente con la rilevazione GI, potrebbe essere il 2000.

Diffusione delle informazioni

Attualmente la rilevazione GI rilascia comunicati stampa mensili. I dati sulle GI vengono temporaneamente utilizzati per produrre per l'Eurostat delle medie trimestrali per l'indice europeo del costo del lavoro (LCI). Si deve prevedere la sostituzione dei dati delle GI con degli indici integrati con i dati INPS.

Sul piano nazionale, invece, si può prevedere dei rilasci a fine trimestre, coincidenti con il comunicato stampa delle GI relativo al terzo mese del trimestre (aprile, luglio, ottobre, gennaio).

Revisioni

Ogni trimestre sarà rilasciata una stima "preliminare" per il trimestre di riferimento a circa 80 gg. La stima definitiva potrà essere rilasciata 12 mesi dopo quando sarà effettivamente disponibile l'universo dei dati INPS. La prima provvisoria è caratterizzata da un errore campionario + errori non campionari; la seconda definitiva rilasciata a 12 mesi di distanza dall'altra e basata sul censimento completo dei DM10, è caratterizzata soltanto dagli errori non campionari.

RIFERIMENTI BIBLIOGRAFICI

ANIV, (1999), Agenda Aniv Guida pratica per una corretta contribuzione

Eurostat (1998a) Regolamento Congiunturale

Eurostat (1998b) Data Quality Report.

Eurostat (1998a) Regolamento Congiunturale

Eurostat (1998b) Methodological Manual

Eurostat, (1995), Developing a Eurostat Indicator of Average Annual Working Time Doc. OS/E3/WT/03-95

Freedman, D. A., and W.C., Navidi, (1986), Regression Models for adjusting the 1980 Census, *Statistical Science*, Vol.1 N.1, 3-39

Istat (1999) "Indice di costo del lavoro: l'uso dei dati della rilevazione sulle grandi imprese per la stima provvisoria", nota interna, settembre 1999.

Ginzburg, A., M. Scaltriti, G. Solinas, e R.Zoboli, (1998), Un nuovo autunno caldo nel Mezzogiorno? Note in margine al dibattito sui differenziali salariali territoriali, *mimeo*

Lucifora, C., e F., Origo, (1997), Evoluzione della struttura retributiva 1990-1994: un'analisi dei micro-dati tratti dagli archivi INPS, in CNEL Rapporto sulle retribuzioni e sul costo del lavoro parte prima (versione provvisoria) 249-263

Revelli, R., (1995), Potenzialità degli archivi INPS ai fini della stima degli aggregati dei conti economici territoriali, *Quaderni di Ricerca ISTAT*, n.1, 138-184

R&P (1990) Il CSC, *Lavoro della ricerca commissionata dall'INPS*

R&P (1990) Questioni interpretative in materia di retribuzioni poste a DM10M; riscontri con altre fonti *Lavoro della ricerca commissionata dall'INPS*.