

n. 9/2007

**Trattamenti monetari non pensionistici  
metodologia sperimentale per la stima degli assegni al  
nucleo familiare**

*C. Peperoni e F. Tuzi*

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT, del SISTAN, o da studiosi esterni.

La *Rivista di Statistica Ufficiale* accoglie lavori che hanno come oggetto la misurazione dei fenomeni economici, sociali, demografici e ambientali, la costruzione di sistemi informativi e di indicatori, le questioni di natura metodologica, tecnologica o istituzionale connesse al funzionamento dei sistemi statistici e al perseguimento dei fini della statistica ufficiale.

I lavori pubblicati in *Contributi ISTAT* sono diffusi allo scopo di stimolare il dibattito intorno ai risultati preliminari di ricerca in corso.

I *Documenti ISTAT* forniscono indicazioni su linee, progressi e miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei *Contributi ISTAT* e nei *Documenti ISTAT* sono valutati preventivamente dai dirigenti dell'Istituto, mentre i lavori pubblicati nella *Rivista di Statistica Ufficiale* sono subordinati al giudizio di referee esterni.

Direttore responsabile della Rivista di Statistica Ufficiale: Patrizia Cacioli

Comitato di Redazione delle Collane Scientifiche dell'Istituto Nazionale di Statistica

Coordinatore: Giulio Barcaroli

Membri:	Corrado C. Abbate	Rossana Balestrino	Giovanni A. Barbieri
	Giovanna Bellitti	Riccardo Carbini	Giuliana Coccia
	Fabio Crescenzi	Carla De Angelis	Carlo M. De Gregorio
	Gaetano Fazio	Saverio Gazzelloni	Antonio Lollobrigida
	Susanna Mantegazza	Luisa Picozzi	Valerio Terra Abrami
	Roberto Tomei	Leonello Tronti	Nereo Zamaro

Segreteria: Gabriella Centi, Carlo Deli e Antonio Trobia

Responsabili organizzativi per la *Rivista di Statistica Ufficiale*: Giovanni Seri e Carlo Deli

Responsabili organizzativi per i *Contributi ISTAT* e i *Documenti ISTAT*: Giovanni Seri e Antonio Trobia

# CONTRIBUTI ISTAT

**n. 9/2007**

**Trattamenti monetari non pensionistici  
metodologia sperimentale per la stima degli assegni al  
nucleo familiare**

*C. Peperoni(\*) e F. Tuzi(\*)*

(\*) ISTAT - Servizio Statistiche sulle istituzioni pubbliche e private

**Contributi e Documenti Istat 2007**

Istituto Nazionale di Statistica  
Servizio Produzione Editoriale

Produzione libraria e centro stampa:  
*Carla Pecorario*  
Via Tuscolana, 1788 - 00173 Roma

# Indice

## Premessa

### Capitolo 1 – Il quadro generale

1.1 - I trattamenti monetari non pensionistici (Tmnp). Il quadro generale.....	pag. 7
1.1.2 - Le fonti.....	“ 8
1.2 - La produzione di informazioni statistiche sui Tmnp: la situazione attuale.....	“ 8
1.2.1 - Le fonti disponibili.....	“ 8
1.2.2 - Modelli di dichiarazione 770 e problemi di gestione del dato amministrativo.....	“ 10

### Capitolo 2 – Utilizzo del modello 770 per la produzione di stime sugli Assegni al nucleo familiare

2.1 – Variabili rilevanti e problemi relativi alla determinazione degli Assegni al nucleo familiare..	“ 15
2.2 – Definizione ed individuazione dei titolari di Assegni al nucleo familiare.....	“ 16
2.3 – Individuazione delle coordinate Anf incomplete o non corrette e relative regole di imputazione.....	“ 17
2.4 – Un approccio longitudinale per il raffinamento delle coordinate e del numero dei titolari.....	“ 20
2.5 – Determinazione dell’importo dell’assegno.....	“ 23

### Capitolo 3 – Determinazione degli importi totali e valutazione delle stime

3.1 – Assegni al nucleo familiare percepiti: la determinazione degli importi totali.....	pag. 29
3.1.1 – Matricole aziendali nel 770 e nel Dm10: la popolazione degli archivi.....	“ 29
3.2 – Verifica delle stime Anf da archivio 770 rispetto ai dati di bilancio Inps ed al Dm10.....	“ 32
3.2.1 – Verifica delle stime rispetto alla classe dimensionale.....	“ 33
3.2.2 – Verifica delle stime rispetto alla classificazione Ateco.....	“ 34
3.3 – Le procedure di stima degli Anf dagli archivi M_770: alcune considerazioni conclusive.....	“ 36

### Capitolo 4- Aspetti metodologici della stima degli assegni per il nucleo familiare

4.1 – Campo di osservazione e unità d’analisi.....	“ 39
4.2 – Fonti di distorsione: non-sampling errors.....	“ 40
4.3 - Problemi di copertura dell’archivio 770.....	“ 41
4.4 – Unit e item non response.....	“ 43
4.5 – Osservazioni conclusive.....	“ 49
Appendice A: <i>La scelta dell’archivio di riferimento per la stima degli assegni al nucleo familiare: studio di compatibilità tra archivi 770 di provenienza Inps e Sogei</i> .....	“ 51

### Capitolo 5 – Validazione e controllo di qualità delle stime Anf

5.1 – Valutazione del livello di precisione delle stime.....	pag. 55
5.2 – Popolazione e parametri di interesse.....	“ 55
5.3 – Ipotesi distributive e metodo di stima.....	“ 56
5.4 – Valutazione empirica della fattibilità delle stime.....	“ 57
5.5 – Osservazioni conclusive.....	“ 59

### Capitolo 6 – Le prospettive future del progetto Tmnp: l’introduzione del modello Emens

6.1 – L’introduzione del modello <i>Emens</i> .....	“ 69
6.2 – Il modello <i>Emens</i> : struttura generale.....	“ 70
6.3 – Il contenuto informativo del modello <i>Emens</i> .....	“ 71
6.4 – Le dichiarazioni correttive e rettificative.....	“ 77
6.5 – L’introduzione del modello <i>Emens</i> e le modifiche apportate alla parte C del modello 770.....	“ 78

Bibliografia.....	“ 81
Glossario.....	“ 83



## Premessa

La produzione di informazioni statistiche sulle prestazioni monetarie non pensionistiche è compresa nelle attività programmate dall'Istituto, nell'ambito delle Statistiche della previdenza e dell'assistenza sociale. Partendo dalla struttura classificatoria fornita dal *Sistema di Classificazione delle Prestazioni Monetarie non pensionistiche* (Scpm) l'attività svolta è stata finalizzata all'individuazione delle fonti informative necessarie al fine di ricostruire una rappresentazione statistica il più possibile esaustiva del fenomeno indagato. Successivamente, di tali fonti informative si è concretamente verificata la disponibilità ed utilizzabilità.

Tale attività ha evidenziato l'intrinseca eterogeneità del fenomeno oggetto di indagine, declinabile essenzialmente in tre direttrici:

- concettuale: le diverse tipologie di trattamenti monetari non pensionistici assistono il cittadino in corrispondenza di diversi eventi nella sua vita professionale e/o sociale e differiscono in maniera sostanziale quanto a modalità di calcolo dell'importo spettante, periodicità, soggetti aventi diritto, ente erogatore...;
- normativa: le diverse tipologie di Tmnp si caratterizzano per una legislazione assai complessa e frammentata. Differenze rilevanti sussistono anche all'interno dello stesso tipo di prestazione, ad esempio quando erogata a differenti tipologie di lavoratori;
- fonti informative: allo stato attuale la produzione di statistiche complete sui Tmnp presuppone l'utilizzo di molteplici fonti informative fra loro disomogenee da numerosi punti di vista: quello della complessità, quello della dimensione fisica, quello del diverso livello di disaggregazione disponibile, quello della diversa tempestività, quello della diversa qualità.

A valle di questa fase esplorativa, l'attenzione è stata rivolta al modello 770, utile per la produzione di statistiche su due tipologie di trattamenti monetari non pensionistici – gli Assegni al Nucleo Familiare (Anf) ed il Trattamento di Fine Rapporto (Tfr) – ed al centro di un'attività sperimentale che coinvolge un gruppo di lavoro interistituzionale volto a verificarne l'utilizzabilità anche in altri settori, quali quelli delle statistiche sui redditi, sull'occupazione, sulla tassazione delle persone fisiche.

È infine opportuno evidenziare come dalla continua evoluzione normativa che caratterizza il settore previdenziale conseguano, spesso, mutamenti sostanziali degli orizzonti informativi in cui la statistica ufficiale è chiamata ad operare. Così, con il Dl 269/2003, art. 44, comma 9, convertito in legge n.326/2003, è stato introdotto il modello *Emens*, che da un lato, a partire dai redditi del 2005, svuota il modello 770 di gran parte del contenuto utile ai fini dell'indagine sui Tmnp, dall'altro fornisce, se adeguatamente compilato, una mole di informazioni che, per dettaglio e periodicità, costituiscono un notevole passo in avanti rispetto al 770.

Il contenuto di questo volume riprende ed amplia i risultati sinteticamente presentati nell'ambito del Seminario *Le fonti amministrative e la produzione di statistiche sui trattamenti monetari non pensionistici*, svoltosi a Roma, presso la sede centrale dell'Istat, il 6 ottobre del 2006.

Il primo capitolo è dedicato all'inquadramento generale del fenomeno Tmnp: la classificazione adottata (Scpm), le fonti informative di riferimento e quelle effettivamente disponibili presso l'Istituto. Nel secondo capitolo viene descritto l'impianto concettuale utilizzato per produrre statistiche sugli Anf a partire dal modello 770. Nel terzo capitolo vengono presentate le stime complessive sugli importi Anf erogati, confrontate con il benchmark da bilancio Inps e con le altre stime prodotte in Istituto a partire dal modello Dm10. Il quarto ed il quinto capitolo descrivono rispettivamente gli aspetti metodologici delle procedure di stima degli Anf adottate, ed i relativi controlli di qualità effettuati per validarle. Il sesto capitolo è infine dedicato al contenuto del modello Emens, ed alle prospettive che la sua introduzione apre nell'ambito dell'indagine sui Tmnp.



# Capitolo 1 – Il quadro generale

## 1.1 I trattamenti monetari non pensionistici (Tmnp). Il quadro generale

Il *Sistema di Classificazione delle Prestazioni Monetarie non pensionistiche* (Scpm)<sup>1</sup> individua, nel contesto istituzionale italiano, 26 tipologie o titoli di prestazioni monetarie non pensionistiche (Figura 1.1) che, nell'ottica del Sistema europeo delle statistiche integrate della protezione sociale (Sespros)<sup>2</sup>, rientrano essenzialmente nelle seguenti funzioni economiche:

- malattia e cure sanitarie;
- famiglia;
- disoccupazione;
- invalidità;
- vecchiaia;
- superstiti;
- esclusione sociale.

**Figura 1.1 - Classificazione per funzione Sespross dei Trattamenti Monetari non Pensionistici**

FUNZIONE (Sespross)	Prestazione Tmnp
Famiglia	Assegno per il nucleo familiare
	Assegni familiari (vecchio tipo)
	Assegni ai nuclei con almeno tre figli minori
	Assegno per il nucleo familiare di lavoratori parasubordinati
	Indennità di maternità
	Assegno di maternità per le donne non indennizzate
	Congedo matrimoniale
Disoccupazione	Indennità di disoccupazione
	Integrazioni salariali: CIGO e CIGS
	Indennità di mobilità
	Indennità sostitutiva preavviso ed incentivi esodo
Malattia	Indennità giornaliera per malattia (ed integrazioni CCNL)
	Indennità giornaliera per malattia (datori di lavoro)
	Indennità TBC
	Indennità giornaliera per inabilità temporanea assoluta (ed integrazioni CCNL)
	Indennità giornaliera per inabilità temporanea assoluta (datori di lavoro)
Invalidità	Equo indennizzo
	Liquidazione in conto capitale di rendite per infortuni sul lavoro e malattie professionali
Vecchiaia	Liquidazione di fine rapporto
	Liquidazioni in conto capitale
Superstiti	Assegno funerario
	Indennità di morte
Esclusione sociale	Minimo vitale
	Reddito minimo di inserimento
	Contributi spese una tantum
	Altre prestazioni monetarie non pensionistiche

<sup>1</sup> Cfr. Paolo Consolini, Rita De Carli, *Le prestazioni sociali monetarie non pensionistiche: unità di analisi, fonti e rappresentazione statistica dei dati*, Roma, Istat, 2002. (Documenti n.1).

<sup>2</sup> Eurostat, *Esspros Manual, Population and social conditions, Methods*, Luxembourg, 1996.

## 1.1.2 Le fonti

Il progetto d'indagine sui trattamenti monetari non pensionistici si avvale in maniera quasi esclusiva di fonti di natura amministrativa che, sebbene offrano notevoli vantaggi in termini di maggiore rapidità ed economia di raccolta del dato (soprattutto in considerazione dell'ampiezza e complessità del fenomeno osservato), comportano non poche difficoltà nel loro utilizzo a fini statistici.

Le principali fonti informative sui Tmnp fanno capo alle seguenti Istituzioni:

- a) Ministero dell'economia e delle finanze: archivio dei modelli 770 - dichiarazioni del sostituto d'imposta (M\_770);
- b) Inps
  - archivio dei modelli Dm10/2 – dichiarazione fiscale del datore di lavoro;
  - archivio Gestione fiscale delle prestazioni dirette (Apd);
  - archivio dei modelli 770 - dichiarazioni del sostituto d'imposta, solo frontespizio e sezione C;
  - archivio delle dichiarazioni *EMens*. La legge n. 326 del 24 novembre 2003 ha previsto che, a partire dalle retribuzioni del mese di gennaio 2005, i datori di lavoro<sup>3</sup> trasmettano mensilmente agli Enti previdenziali in via telematica (direttamente o tramite gli incaricati), entro l'ultimo giorno del mese successivo a quello di riferimento, i dati retributivi e le informazioni utili al calcolo dei contributi, all'implementazione delle posizioni assicurative individuali ed all'erogazione delle prestazioni. Quest'ultima fonte, al momento non disponibile, avrà un notevole impatto sulla rilevazione dei TMNP del settore privato in quanto conterrà informazioni più dettagliate e puntuali di quelle contenute negli archivi M\_770<sup>4</sup>.
- c) Inpdap: archivi sulle prestazioni dirette<sup>5</sup>.
- d) Inail: archivio delle prestazioni dirette;
- e) Ragioneria Generale dello Stato (Rgs): banca dati del pubblico impiego e conto annuale. Le informazioni ivi raccolte non consentono di elaborare statistiche sulle prestazioni erogate al singolo lavoratore dipendente. Queste sono invece reperibili presso le amministrazioni (centrali e/o locali) di appartenenza;
- f) Istat:
  - rilevazione sui Bilanci consuntivi degli enti previdenziali;
  - rilevazione sulla Struttura del costo del lavoro (cadenza biennale);
  - indagine sul Sistema dei conti economici delle imprese;
  - Conto satellite della protezione sociale.

Le basi di dati Istat sono funzionali alla verifica dell'affidabilità degli archivi amministrativi di cui al punto a) e b) e alla rilevazione del dato economico aggregato di alcune tipologie di Tmnp non altrimenti disponibili. Per specifiche categorie professionali (giornalisti, agenti di commercio, addetti al settore marittimo, ecc.) l'erogazione di alcune tipologie di Tmnp spetta ad enti previdenziali dedicati. In questi casi, ai fini della produzione di informazioni statistiche, si dispone solo di dati aggregati a livello nazionale, desumibili dalla citata indagine Istat sui Bilanci consuntivi degli enti previdenziali.

## 1.2 La produzione di informazioni statistiche sui Tmnp: la situazione attuale

### 1.2.1 Le fonti disponibili

Rispetto al sopradescritto universo informativo, necessario a ricostruire nella sua interezza il fenomeno in analisi, le fonti ad oggi disponibili ed in uso presso l'Istat permettono di riprodurre un quadro solo parziale dei Tmnp e dei loro percettori. Infatti, mentre alcune fonti informative, rispetto al contenuto teorico, sono risultate inutilizzabili perché non ancora alimentate in maniera stabile ed informatizzata (archivio Rgs), altre sono ad oggi disponibili solo con riferimento a specifiche funzioni (archivio Inpdap-Apd, acquisito solo con riferimento alla funzione disoccupazione). Infine, altre fonti non sono attualmente disponibili presso l'Istat (ad esempio

<sup>3</sup> Tra i soggetti tenuti all'invio dei dati mensilizzati rilevano, ai nostri fini, i datori di lavoro già tenuti alla compilazione della parte C, dati previdenziali ed assistenziali Inps o Inpdap, del modello 770 semplificato.

<sup>4</sup> I rilasci E-Mens già raccolti presso l'Inps riguardano tutte le mensilità del 2005. I dati sono in fase di consolidamento informatico presso il Ced, ma non ancora disponibili per il trattamento statistico. Considerato il dettaglio dell'informazione e l'elevata tempestività rispetto al periodo di riferimento, sarebbe opportuno porre in essere accordi istituzionali con l'Inps per ottenere la disponibilità degli archivi e/o avviare gruppi di lavoro congiunti.

<sup>5</sup> Disponibili presso l'Istat per gli anni dal 1999 al 2003, solo con informazioni sul Tfr a livello aggregato.

l'archivio Inps-Apd, nonostante sia stato ufficialmente e formalmente richiesto, ad oggi non è stato ancora rilasciato dall'Inps).

Alle difficoltà connesse alla disponibilità delle fonti informative, vanno aggiunte quelle che derivano dalla circostanza che molti degli archivi citati non forniscono dati con un livello di disaggregazione tale da consentire l'individuazione diretta dei beneficiari (Dm10, Rgs), rendendo necessario l'utilizzo di opportuni modelli statistici.

Allo stato attuale, l'archivio dei modelli 770 costituisce la principale fonte informativa utilizzata. L'archivio è disponibile per gli anni di imposta 2000-2004 nella versione Sogei (2000-2003 in quella Inps) ed è strutturato in modo da fornire informazioni sui singoli beneficiari delle prestazioni. Per il comparto privato, tale archivio consente di ottenere una copertura completa dei percettori di Tmnp. L'elaborazione e l'analisi dei dati connessi alle specifiche tipologie di prestazione richiedono la risoluzione di problemi legati alla mancanza, all'interno dell'archivio, di informazioni adeguatamente disaggregate. In estrema sintesi, si può affermare che il modello 770 è una buona fonte per quanto concerne l'individuazione dei soggetti beneficiari (chi?) mentre presenta rilevanti limiti quanto alla determinazione degli ammontari di spesa a livello aggregato ed, ancor più, a livello di singola prestazione (quanto?).

L'Istituto dispone di due versioni della parte C, *Dati previdenziali ed assistenziali Inps*, dell'archivio M\_770 per gli anni di imposta 2000, 2001, 2002, 2003 e 2004, e anni di riferimento, rispettivamente, 1999, 2000, 2001, 2002 e 2003. Il primo archivio, rilasciato dal Progetto Sistemi Informativi Integrati (Dcar/A), della Direzione centrale degli archivi amministrativi e registri statistici (Dcar) dell'Istat, copre tutte le sezioni del modello ufficiale ad esclusione di quella dedicata al lavoro autonomo, ed è direttamente disponibile in versione grezza, in assenza, cioè, di controlli o correzioni. Il secondo archivio, in versione semplificata e 'depurata', è fornito dall'Inps. In tale archivio la sezione previdenziale ed assistenziale è integrata con alcune informazioni anagrafiche sul percettore. Inoltre i dati amministrativi originari sono sottoposti ad un insieme di controlli ed elaborazioni che rendono tale archivio, rispetto al primo, più attendibile in termini di qualità<sup>6</sup> del dato ma meno completo, in quanto privo dell'informazione relativa alla sezione fiscale, di cui andrebbe comunque indagata l'eventuale utilità ai fini della ricostruzione di alcune poste relative ai Tmnp.

La migliore qualità del dato ha motivato la decisione di utilizzare la base dati fornita dall'Inps e di procedere ad una sua successiva integrazione con l'archivio fornito dal Dcar/A. L'utilizzo dell'archivio Inps consente di migliorare la tempistica del progetto, in quanto il dato non presenta problemi legati a dichiarazioni ripetute, integrate e rettifiche<sup>7</sup> ma rinvia al futuro le problematiche relative a un eventuale aggancio con l'archivio fornito dal Dcar/A, che richiederà un consolidamento preliminare dei record. In vista della necessità di tale integrazione sono stati effettuati alcuni controlli relativi alla 'consistenza' dei due archivi (Prospetto 1.1) e in generale è emerso un maggior popolamento dell'archivio fornito dall'Inps rispetto a quello fornito dal servizio Dcar/A (dovuto, però, a una diversa data di acquisizione dei dati), con tassi di caduta che rendono comunque ammissibile la confrontabilità dei due archivi.

#### Prospetto 1.1 - Archivio M\_770 (parte C): confronto tra le basi di dati Istat e Inps

Anno di esercizio	Numero imprese			Numero dichiarazioni			Numero codici fiscali distinti		
	Inps	Istat	tasso di caduta	Inps	Istat	tasso di caduta	Inps	Istat	tasso di caduta
2001	1.482.472	1.501.962	1,3	16.125.595	17.897.242	9,9	12.918.509	13.092.946	1,33
2002	1.560.035	1.378.099	-11,66	16.676.344	16.162.814	-3,08	13.342.876	12.182.646	-8,7

Le problematiche connesse all'effettiva disponibilità delle fonti informative ha reso necessario ritardare l'ambizioso obiettivo con il quale è stato avviato il progetto d'indagine sui Tmnp. Relativamente alla funzione *famiglia*, l'elevato grado di copertura dell'universo dei beneficiari, che si unisce però all'assenza di un dato adeguatamente disaggregato per le singole prestazioni appartenenti alla funzione in analisi, hanno indotto, finché nuove basi dati non saranno disponibili, a focalizzare l'attenzione sul settore privato. Relativamente alla funzione *malattia e cure sanitarie*, le fonti informative attualmente disponibili non consentono di elaborare statistiche complete ed attendibili: gli unici archivi che forniscono indizi sulla presenza di eventi associati a tale

<sup>6</sup> Per la comparazione qualitativa della versione Inps dell'archivio M\_770 con quella Sogei cfr. Appendice A, pag 41.

<sup>7</sup> Un lavoratore dipendente può apparire in più di una dichiarazione 770, tipicamente in presenza di particolari eventi legati alla vita dell'impresa, a variazioni di qualifica o anche in seguito a dichiarazioni compilate in maniera erronea od incompleta e successivamente rettifiche mediante reinvio delle dichiarazioni. Ciò incrementa in maniera significativa le difficoltà connesse al trattamento informatico dell'archivio 770.

funzione sono l'archivio 770 e l'archivio delle prestazioni dirette Inail. Il primo fornisce informazioni che consentono di risalire con un buon grado di approssimazione solo alle durate degli eventi associati a malattia ed infortuni, per quanto riguarda il secondo, invece, occorre valutarne l'utilizzabilità in quanto è strutturato secondo il criterio di cassa, contrariamente agli altri archivi in uso che contengono, invece, dati di competenza. Per quanto attiene la funzione *Disoccupazione*, l'archivio 770 consente l'individuazione dei soggetti a cui sia stata erogata retribuzione ridotta a titolo di cassa integrazione guadagni (Cig). Quanto all'individuazione dei relativi ammontari erogati, un primo problema deriva dalla circostanza che eventuali periodi in cui la Cig sia stata pagata direttamente dall'Inps non vengono rilevati nel modello 770. Una seconda difficoltà deriva dal contenuto della voce 39 (Differenza retribuzione da accreditare per Cig) in cui deve essere indicato «per i periodi di sospensione e/o riduzione dell'orario di lavoro che cadono nell'anno di riferimento e per i quali sia stata autorizzata la corresponsione della integrazione salariale, l'ammontare complessivo delle retribuzioni che sarebbero spettate al lavoratore se nello stesso periodo avesse lavorato normalmente, escludendo le somme corrisposte dal datore di lavoro nei periodi anzidetti e assoggettate a contribuzione obbligatoria». Quindi nella voce in questione non viene indicato l'ammontare Cig erogato al lavoratore, ma la retribuzione standard che gli sarebbe spettata o quella ridotta comunque erogata dal datore di lavoro. Deve essere quindi indagata la possibilità di ricostruire l'effettivo ammontare del Tmnp erogato, a partire da un dato che potrebbe, forse, esserne il complemento rispetto ad un totale costituito dalla retribuzione totale ordinaria.

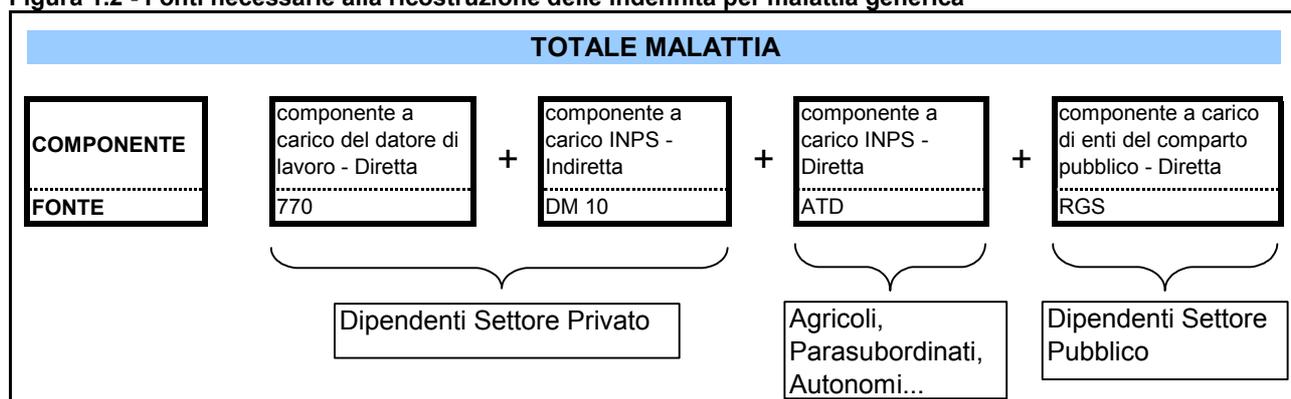
### 1.2.2 Modelli di dichiarazione 770 e problemi di gestione del dato amministrativo

L'utilizzo dell'archivio M\_770 per la produzione di informazioni statistiche sui trattamenti monetari non pensionistici presenta problemi di:

- copertura; da un punto di vista dei beneficiari, l'archivio M\_770 copre l'universo dei potenziali percettori di Tmnp con esclusivo riferimento ai dipendenti del settore privato. Restano quindi esclusi i lavoratori autonomi e quelli pubblici (ad eccezione di particolari categorie, quali quelle legate ai contratti cosiddetti 'atipici', la cui gestione previdenziale spetta all'Inps). La figura 1.2, pur riferendosi ad una specifica tipologia di prestazione monetaria non pensionistica è esemplificativa delle difficoltà connesse alla ricostruzione di una prestazione, operazione possibile solo attraverso la complessa integrazione di dati provenienti da fonti numerose ed eterogenee. Nella figura si può osservare, infatti, che per il settore privato, oltre alla componente direttamente erogata dal datore di lavoro (fonte 770) dovrà essere considerata la componente da questi anticipata, ma a carico dell'Inps (fonte Dm10). Per i lavoratori autonomi, agricoli e parasubordinati i dati sono invece forniti dall'archivio delle prestazioni dirette dell'Inps. Infine, per il settore pubblico l'archivio di riferimento è quello della Ragioneria Generale dello Stato (Rgs). Mentre per le due prime categorie di percettori l'informazione è formalmente disponibile, anche se caratterizzata da limitazioni di copertura e contenuto, per il settore pubblico l'archivio Rgs (non disponibile in Istat) presenta rilevanti problemi di aggiornamento e qualità dei dati.
- incompletezza dell'informazione; le informazioni contenute nel modello 770 consentono di determinare il numero dei beneficiari ma non l'importo delle prestazioni percepite, fornendo solo alcuni indizi utili alla ricostruzione di valori stimati. Per quanto riguarda le procedure di stima adottate, e la conseguente attendibilità dei risultati, sono stati definiti degli ordini di priorità che tengono conto sia dei requisiti statistici di precisione delle stime sia di motivazioni legate alle tempistiche del progetto.

Le informazioni maggiormente utili ai fini dell'indagine sono contenute nella parte C del modello 770, nelle sezioni volte alla rilevazione dei *Dati previdenziali ed assistenziali Inps e Inpdai* (Figura 1.3). Queste riportano la situazione previdenziale/assistenziale della persona fisica cui si riferisce il modello. Tali informazioni, tuttavia, non sono sufficienti alla produzione delle informazioni statistiche di interesse, consentendo di ipotizzare, ma non di stabilire con certezza, la durata e l'importo dei Tmnp.

**Figura 1.2 - Fonti necessarie alla ricostruzione delle indennità per malattia generica**



Con specifico riferimento agli Assegni al Nucleo Familiare (Anf), nel modello 770 i nuclei titolari sono individuati da apposite voci che ne sintetizzano la composizione familiare e reddituale, attraverso le quali è possibile ricondursi a tabelle Inps per il calcolo del valore dell'assegno mensile spettante. Quindi il modello 770 non fornisce direttamente l'informazione relativa all'ammontare dell'assegno percepito, ma soltanto alcune coordinate che permettono di ricostruire indirettamente l'informazione, mediante il rimando a tabelle attraverso le quali ricavarlo. La stima dell'ammontare degli Anf sulla base dell'archivio M\_770 presenta quindi una serie di problemi:

- il reddito presente sulla dichiarazione 770 non è quello su cui andrebbero calcolati gli Anf, sia in termini di riferimento temporale che di composizione: l'assegno familiare spettante per il primo semestre dell'anno t viene infatti stabilito in base al reddito percepito nell'anno t-2, mentre quello spettante per il secondo semestre viene stabilito in base al reddito dell'anno t-1. Il reddito presente nel modello 770 è invece, ovviamente, quello dell'anno t. Quanto invece alla composizione del reddito, il diritto all'assegno è riconosciuto in base al reddito complessivo del nucleo familiare, mentre il reddito desumibile dal modello 770 è unicamente quello del richiedente, titolare della prestazione;
- non esiste alcun riferimento certo che permetta di stabilire per quanti mesi l'Anf sia stato percepito;
- le regole di compilazione del modello 770 prevedono che le voci determinanti ai fini Anf debbano essere compilate solo se il dipendente ha percepito l'assegno nel mese di dicembre. Quindi un soggetto che percepisse Anf da gennaio a novembre, ma non a dicembre, verrebbe totalmente 'mancato' da una rilevazione esclusivamente basata sui dati del modello 770<sup>8</sup>.

Per quanto concerne le altre prestazioni riconducibili alle funzioni malattia, disoccupazione (ad esclusione della Cig) e famiglia (relativamente alle voci 'maternità' e 'congedi parentali'), all'interno del modello 770 le voci di maggiore rilievo sono quelle relative agli importi delle retribuzioni ridotte (voce 38) e alle durate complessive degli eventi che hanno originato la riduzione delle retribuzioni. Non essendo disponibile l'informazione relativa alla singola prestazione, è prevedibile che non sarà possibile risalire ai singoli importi, se non con un elevato grado di approssimazione. Solo quando i dati delle dichiarazioni *EMens* si renderanno disponibili sarà possibile ricostruire con maggiore precisione le voci disaggregate delle prestazioni. Per quanto riguarda invece la Cig, pur essendo presente una voce specificamente destinata a raccogliere dati monetari connessi a questa prestazione (voce 39), la determinazione dell'ammontare erogato pare comunque problematica, come già sintetizzato nel paragrafo 1.2.1, al quale si rimanda.

<sup>8</sup> Per una descrizione della procedura adottata per ovviare a questo problema cfr. cap.2, par.2.4, pag. 23.

Figura 1.3 - Modello 770 – parte C, anno 2003 (redditi 2002)

Codice fiscale del percipiente											Progressivo Comunicazione													
<b>PARTE C</b>																								
<b>DATI PREVIDENZIALI ED ASSISTENZIALI INPS E INPDAI</b>																								
Qualifica			Esito		Matricola azienda			Prov.Lav.		F/S		DS		altre		FG		Competenze comuni						
1 2 3			4		5			6		7 8		9		10		11								
Altre competenze											Settimane retrib.		Giorni retrib.		MESI RETRIBUITI NELL'ANNO									
12											13		14		15 16									
															Mesi con esclusione di									
															T G F M A M G L A S O N D									
Livello inquadramento											Data assunzione		Tipo rapporto		Trasf. rapporto		Settimane utili		Accantonamento TFR spettante		COORD. ASSEgni FAMILIARI			
19											20		21		22		23		24		25 26 27			
<b>SEZIONE 2</b>																								
<b>RETRIBUZIONI PARTICOLARI</b>																								
Tipo		Data inizio			Data fine			Retribuzione		Settimane retrib.		GG retrib.		GG utili		GG non retrib.		Retribuzione pensionabile						
1)		29			30			31		32		33		34		35		36						
2)		29			30			31		32		33		34		35		36						
3)		29			30			31		32		33		34		35		36						
4)		29			30			31		32		33		34		35		36						
<b>SEZIONE 3</b>																								
<b>CONTRIBUZIONE FIGURATIVA</b>																								
Isolamento dalla contribuzione			Retribuzioni ridotte			Differenza retribuzione da accreditare per C/2			MALATTIA E/O INFORTUNA		MALATTIA E/O INFORTUNA		MALATTIA E/O INFORTUNA		MALATTIA E/O INFORTUNA		MALATTIA E/O INFORTUNA		MALATTIA E/O INFORTUNA					
37			38			39			40		41		42		43		44		45					
L. 80/97			C/2			Congedi art. 42 c. 5 D.Lgs. 151/01			D. 165/99, art. 4		D. 165/99, art. 4		D. 165/99, art. 4		D. 165/99, art. 4		D. 165/99, art. 4		D. 165/99, art. 4					
46			47			48			49		50		51		52		53		54					
<b>SEZIONE 4</b>																								
<b>COLLAB. COORDINATE E CONTINUATIVE</b>																								
Data inizio			Data fine			Compensi corrisposti			Contributi dovuti		Contributi trattenuti		Contributi versati		Contributi dovuti all'INPS o all'INPDAI									
57			58			59			60		61		62		63 64 65									
<b>DATI PREVIDENZIALI ED ASSISTENZIALI INPDAP</b>																								
Codice fiscale Annuale				Decorrenza giuridica				Data inizio		Data fine		N. giorni utili ai fini del TFR		Causa cessazione										
66				67				68		69		70		71										
Genitori assicurativi		Tipo impiego		Tipo servizio		Cassa/ Corp		Giorni utili		Maggiorazioni														
72		73		74		75		76		77 78 n. giorni 79 80 n. giorni 81 82 n. giorni 83 84 n. giorni														
Retribuzioni fisse e contributive				Retribuzioni accessorie				Retribuzioni di base per il 18%		Retribuzioni ai fini TFS		Retribuzioni ai fini TFR												
85				86				87		88		89												
Premio di produzione				Indennità non annualizzabili				D.Lgs. 165/99, art. 4		Tredicesima mensilità														
90				91				92		93														
Data opzione per il TFR				Codice fiscale Annuale versate in tribuzione fissa e contributiva				Codice fiscale Annuale versate in tribuzione accessoria																
94				95				96																
<b>SEZIONE 5</b>																								
<b>DATI ASSICURATIVI INAIL</b>																								
Qualifica				Posizione assicurativa territoriale				C. C.		Data inizio		Data fine		Codice comune		Personale viaggiante								
101				102				103		104		105		106		107								

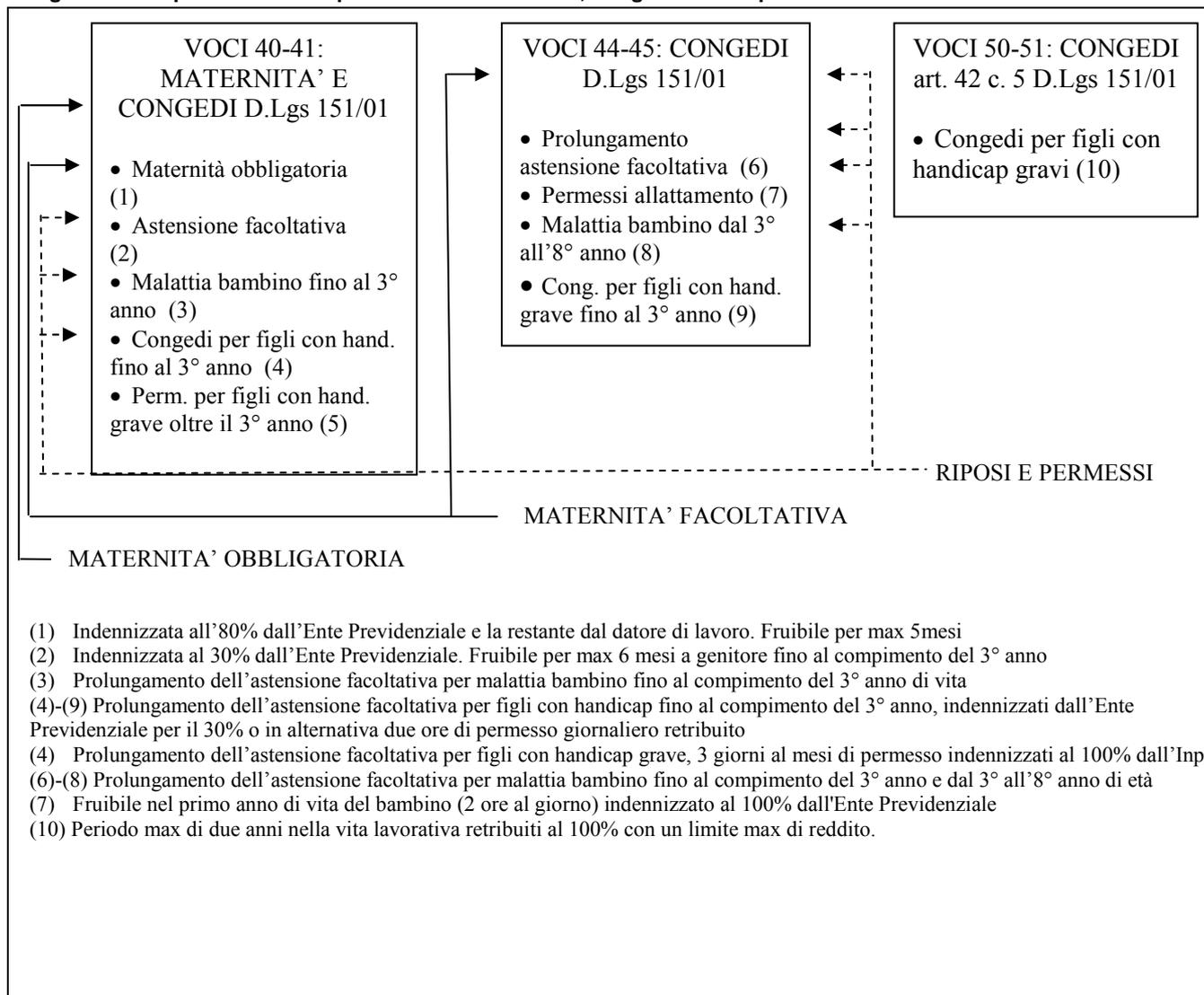
Oltre che per le ragioni sin qui illustrate, allo stato attuale, l'implementazione delle procedure di stima è limitata in quanto:

- il modello 770 rileva esclusivamente gli eventi Tmpn ai quali si associa una riduzione della retribuzione, mentre non consente di rilevare quelli in corrispondenza dei quali la retribuzione rimane completa. Si considerino, ad esempio, i primi tre giorni di malattia. Questi vengono retribuiti dal datore di lavoro, solitamente in misura pari al 100% del reddito spettante per i giorni lavorativi 'ordinari'. In questo caso all'evento 'primi tre giorni di malattia' (cosiddetta 'carenza') non si associa alcuna riduzione di retribuzione, e quindi l'evento stesso non è rilevabile attraverso le informazioni presenti nel 770.
- non tutti gli eventi sono conteggiati nelle durate, in quanto i periodi inferiori alla settimana non rientrano nel computo. Ciò porta a sottostimare il dato e in alcuni casi, come ad esempio in periodi di

- malattia molto brevi, costituisce un problema non trascurabile;
- c) nel caso di maternità e congedi, l'informazione non è resa nel dettaglio necessario. Ad esempio, i periodi di maternità obbligatoria e facoltativa vengono conteggiati congiuntamente a forme di congedo parentale, per legge soggette a una diversa disciplina contributiva e retributiva (Figura 1.4).

Ne deriva che la ripartizione delle durate dei singoli eventi che originano Tmnp tra i vari periodi dovrebbe basarsi su ipotesi discrezionali, il cui sforzo computazionale potrebbe essere vanificato quando il dato mensilizzato *EMens* sarà reso disponibile.

**Figura 1.4 - Ripartizione delle prestazioni di maternità, congedi ed altri permessi all'interno del modello 770**





## Capitolo 2 – Utilizzo del modello 770 per la produzione di stime sugli Anf

### 2.1 Variabili rilevanti e problemi relativi alla determinazione degli Assegni al nucleo familiare

Ai fini della determinazione degli Anf, le variabili rilevanti sono rappresentate da un set di coordinate (voci 25-26-27 del modello 770, anno 2003) che rinviano a tabelle definite di anno in anno da circolari Inps. Tali coordinate contengono rispettivamente riferimenti a tabelle (voce 25), numero di componenti familiari (voce 26), classi di reddito (voce 27) e sono distinte a seconda si riferiscano a nuclei con figli (tab. 11, 12,13, 14, 15, 16, 17, 18, 19) o senza figli (20A, 20B, 21A, 21B, 21C, 21D). Quindi, in linea generale, determinata la tabella di riferimento con la prima coordinata, l'incrocio delle due successive coordinate (riga e colonna della tabella di riferimento) individua l'assegno mensile spettante.

Tuttavia, la determinazione degli importi e del numero di titolari di Anf risulta limitata sia dall'incompletezza e parziale correttezza delle informazioni sulle coordinate familiari che da una serie di aspetti tecnici connessi alla compilazione del modello. Relativamente a quest'ultimo punto, i problemi principali riscontrati per il calcolo dell'ammontare degli Anf sulla base dell'archivio M\_770 sono essenzialmente tre:

- a) i redditi di riferimento per il calcolo degli Anf spettanti sono quelli dichiarati nei due anni precedenti quello della richiesta. In altri termini, per il computo degli Anf relativi all'anno 2002, si tiene conto del reddito 2000 per il periodo di erogazione che va dal primo gennaio 2002 al 30 giugno 2002 e del reddito del 2001 per il periodo di erogazione che va dal primo luglio 2002 al 31 dicembre 2002. Questo comporta problemi:
  - nella verifica della correttezza della classe di reddito (voce 27) che non può avvenire sulla base del reddito indicato nella dichiarazione, in quanto non è quello su cui andrebbero calcolati gli Anf;
  - nella definizione delle coordinate atte ad individuare l'Anf spettante. Coordinate che, riferendosi a redditi di periodi distinti, potrebbero essere diverse nel corso dei due semestri. Ad esempio, per il calcolo degli Anf 2002 è possibile che nell'anno 2000 (riferimento al periodo gennaio-giugno 2002) l'individuo appartenga ad una classe di reddito diversa da quella dell'anno 2001 (riferimento al periodo luglio-dicembre 2002)<sup>9</sup>. Quanto al reddito va evidenziato l'ulteriore problema connesso al fatto che quello indicato nella dichiarazione è il reddito del solo titolare del trattamento, non quello dell'intero nucleo familiare beneficiario, sulla base del quale viene calcolato l'assegno spettante;
- b) Le regole di compilazione del modello 770 prevedono che le voci 25-26-27 debbano essere compilate solo se il lavoratore ha percepito Anf nel mese di dicembre. Questo comporta la perdita totale dell'informazione relativa a chi ha percepito Anf negli altri mesi dell'anno ma non in dicembre, con notevoli implicazioni sulla definizione dell'universo dei percettori di Anf, che verrebbe a coincidere con i soli titolari al mese di dicembre (sottostima Anf);
- c) nessun dato desumibile nell'archivio M\_770 permette di stabilire direttamente per quanti mesi l'Anf sia stato percepito. La soluzione adottata è quella di ricavare il periodo in cui sono stati percepiti gli Anf a partire da quello di retribuzione (sovrastima Anf). In altri termini, l'ipotesi è che il titolare abbia percepito Anf tutti i mesi in cui è stato retribuito. Con opportune e preventive procedure di consolidamento del dato, è possibile evincere l'informazione sul periodo di retribuzione dal numero di settimane, giorni e mesi retribuiti (modello 770, parte C,-voci 13-16). La soluzione più 'promettente' sotto il profilo della precisione delle stime, ma non esente da problemi di implementabilità, è quella relativa all'integrazione temporale degli archivi. Questo approccio consentirebbe di raffinare la stima Anf, recuperando l'informazione sul possibile cambiamento di coordinate nei due sottoperiodi dell'anno, e di allargare l'universo dei potenziali titolari, recuperando l'informazione sui percettori in corso d'anno. La prevalenza dell'effetto sovrastima o di quello sottostima (di cui al punto b) potrà essere valutata in sede di comparazione del dato complessivo Anf con il *benchmark* desumibile dal bilancio consuntivo Inps.
- d) le coordinate Anf non sempre individuano univocamente l'importo dell'assegno mensile spettante. Esistono infatti delle ulteriori, significative specifiche. Ad esempio, per quanto riguarda le tabelle 11, 12, 14, 15, 17 e 18, l'importo individuato nella maniera sopradescritta va ridotto «in presenza di un solo figlio di € 10,33 per il primo fratello, sorella o nipote presente nel nucleo e di €53,71 per ciascuno degli altri eventuali fratelli, sorelle o nipoti; - in presenza di almeno due figli di €53,71 per ogni fratello, sorella o nipote presente nel nucleo. In caso di nuclei composti da più di sette componenti, l'importo dell'assegno

<sup>9</sup> L'aggancio di archivi di anni consecutivi potrebbe essere una via percorribile per tener conto di questo aspetto.

previsto va maggiorato di un ulteriore 10%, nonché di €53,71 per ogni componente oltre il settimo». In altri termini, a rilevare per la determinazione dell'importo dell'assegno spettante non è solo il numero di componenti della famiglia, ma anche il grado ed il tipo di parentela degli stessi. È evidente come in queste circostanze la corretta determinazione dell'assegno richiederebbe informazioni sullo stato di famiglia, in nessun modo desumibili dal 770. Viene così introdotto un ulteriore elemento di indeterminazione.

In sintesi ai fini della produzione di informazioni statistiche, dopo aver individuato, all'interno del modello 770, le variabili utili alla determinazione degli Anf, occorre:

- definire ed individuare la platea dei titolari (chi?);
- correggere le coordinate;
- affinare le coordinate Anf ed il numero dei titolari;
- determinare l'importo (quanto?)

## 2.2 Definizione ed individuazione dei titolari di Assegni al nucleo familiare

Come descritto nel paragrafo precedente, in base alle regole di compilazione del modello 770, è titolare di Anf solo chi percepisce l'assegno nel mese di dicembre dell'anno di interesse. Nell'ambito dell'indagine si è deciso di adottare una definizione più ampia in quanto è stata riscontrata la presenza di coordinate Anf anche in dichiarazioni che facevano riferimento a periodi retributivi non comprendenti il mese di dicembre. Questi dati, pur non coerenti con le regole di compilazione, forniscono informazioni aggiuntive utili agli scopi dell'indagine. Per poterle utilizzare si è quindi deciso di prescindere, nella definizione di titolare, dal riferimento al mese di dicembre.

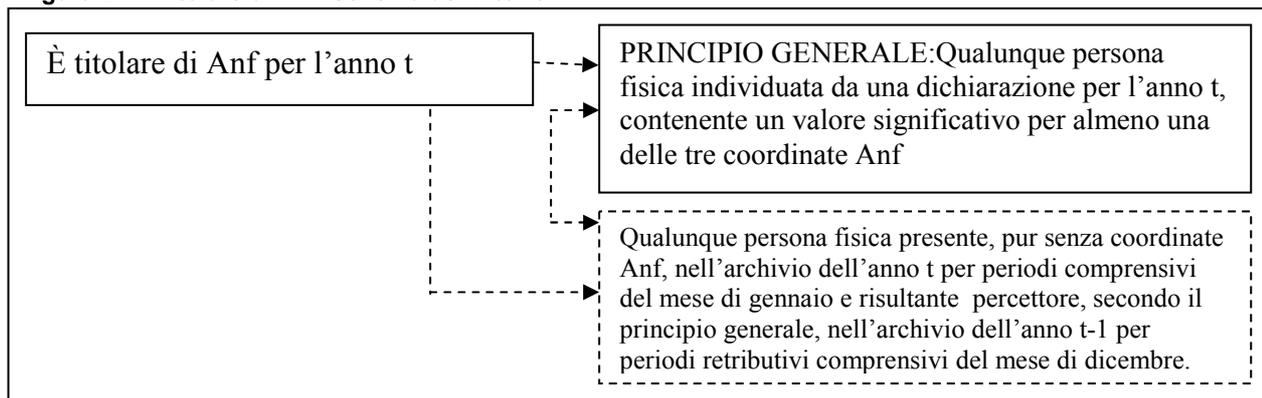
Sulla base di questi assunti, nella fase operativa sono state definite una serie di regole per individuare in maniera univoca i titolari di Anf. Il principio generale è che il titolare, indipendentemente dal riferimento o meno ad un periodo retributivo comprensivo del mese di dicembre, è colui la cui dichiarazione contenga un valore 'significativo' per almeno una delle tre coordinate di rilievo: riferimento tabellare, classe di reddito, numero di componenti familiari. Sono ritenute significative tutte quelle coordinate che contengono un carattere alfanumerico entro i seguenti intervalli:

[a,z], [A,Z], [0,9]

indipendentemente dalla presenza di qualsiasi altro carattere alfanumerico (\*,/,...) la cui eliminazione viene effettuata tramite apposito filtraggio di natura deterministica.

Come verrà meglio illustrato nel paragrafo 2.4, nel caso venisse adottato l'approccio basato sull'integrazione temporale degli archivi M\_770 (anno t ed anno t-1) sarebbero considerati titolari per l'anno t anche coloro i quali pur non risultando percettori nell'anno in questione (ma comunque presenti nell'archivio M\_770 in quanto lavoratori dipendenti), sono presenti con coordinate Anf a dicembre dell'anno t-1 (Figura 2.1). L'ipotesi sottesa a questa scelta è che alla continuità nel ruolo di lavoratore dipendente corrisponda una parallela continuità in quello di percettore Anf. Si suppone cioè che chi ha percepito Anf a dicembre dell'anno t-1, e che nell'anno t risulta presente nel 770 come lavoratore dipendente, continui a percepire Anf anche in assenza di indicazione esplicita delle coordinate (che, come da regole di compilazione, potrebbero non essere state indicate in quanto gli Anf non sono stati percepiti nel mese di dicembre dell'anno t).

**Figura 2.1 - Titolare di Anf: schema definitorio**



### 2.3 Individuazione delle coordinate Anf incomplete o non corrette e relative regole di imputazione

La percentuale delle dichiarazioni parziali (con dati mancanti) o non corrette (con coordinate errate) sul totale delle dichiarazioni significative ai fini del computo degli Anf assume un valore che si attesta intorno al 6% per gli anni 2000 e 2001 e scende al 4,6% per il 2002 (Prospetto 2.1). Ciò suggerisce l'opportunità, da un punto di vista statistico, di implementare tecniche di imputazione per la ricostruzione del dato mancante. Per questo motivo l'archivio fornito dall'Inps è stato scisso in due sotto archivi, il primo riferito alle dichiarazioni 'valide', a cui si fa ricorso per la ricostruzione del dato *missing* o non corretto, il secondo relativo alle dichiarazioni 'non valide', da sottoporre a procedure di correzione. Queste ultime sono state pensate seguendo criteri di uniformità per tutti gli anni di dichiarazione e si basano su approcci di tipo misto deterministico-stocastico, variabili in funzione della natura del dato da correggere o da imputare.

**Prospetto 2.1 - Incidenza delle 'dichiarazioni' non corrette e non complete (Archivio 770 Inps)**

Percettori Anf	2000	2001	2002
Dichiarazioni corrette	1.902.561	1.906.855	1.963.787
Dichiarazioni non corrette	123.842	122.286	94.405
Percentuale di incidenza	(6.1)	(6.0)	(4.6)
<b>Totale dichiarazioni</b>	<b>2.026.403</b>	<b>2.029.141</b>	<b>2.058.192</b>

Relativamente alle coordinate Anf le prime elaborazioni hanno consentito di individuare una casistica sui dati mancanti e non corretti, sintetizzata nella seguente classificazione:

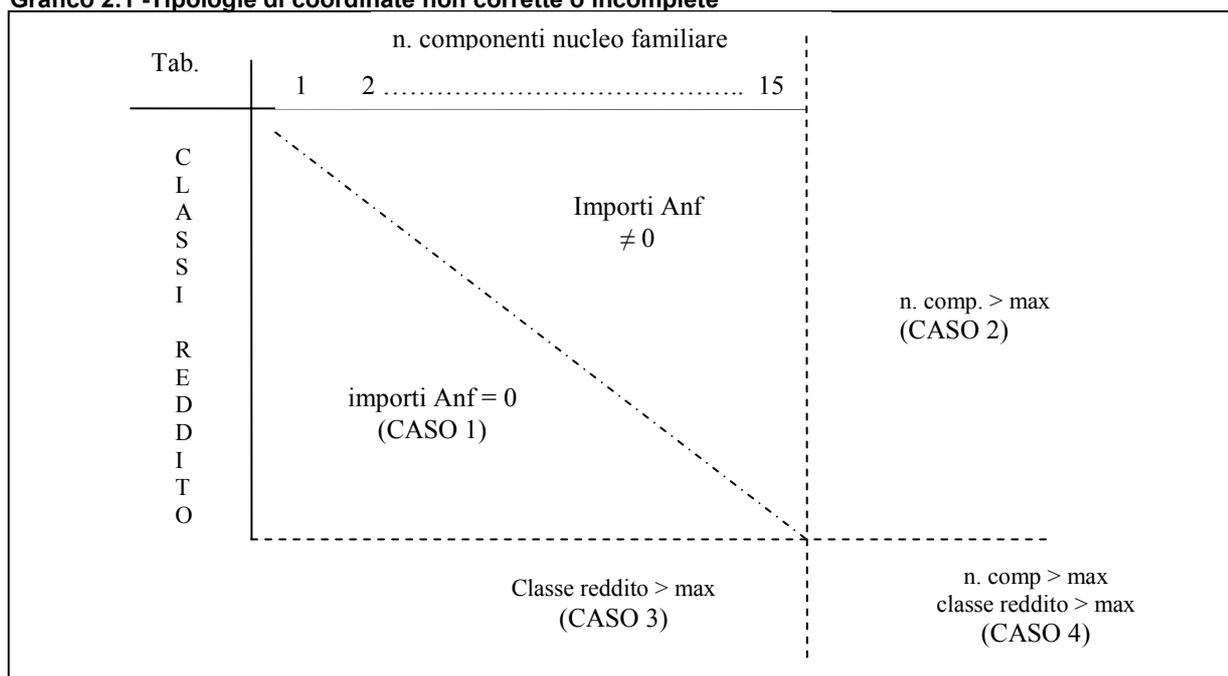
- coordinate incomplete (almeno un campo non compilato) (CASO 0).
- coordinate complete la cui combinazione sulle tre dimensioni genera un importo Anf nullo (CASO 1);
- coordinate complete il cui numero di componenti risulta superiore ad un valore definito ammissibile (CASO 2);
- coordinate complete il cui riferimento è ad una classe reddituale superiore al valore massimo ammissibile (16 per le tabelle da 11 a 16 e 11 per le tabelle da 17 a 21D) (CASO 3);
- coordinate complete con classe di reddito e numero componenti oltre i valori ammissibili (CASO 4).

Con riferimento al CASO 1, che è anche il più complesso da trattare, le tabelle presentano strutture più o meno simili a matrici di tipo 'triangolare superiore', in cui i riferimenti al di sotto della diagonale principale coincidono con valori Anf nulli. È poco probabile pensare che individui che presentano coordinate Anf complete siano non percettori; è invece presumibile che si tratti di compilazioni non corrette legate, ad esempio, ad una cattiva interpretazione delle istruzioni (es. con riferimento ad un nucleo familiare composto di tre persone vengono indicati solo i componenti diversi dall'effettivo dichiarante quindi, in questo esempio 2). La gestione di questo tipo di dichiarazione è molto problematica, in quanto si tratterebbe di scegliere quale tra le tre coordinate sottoporre a modifica. Occorre tener conto che operare sulla tabella implicherebbe introdurre modifiche sulla composizione familiare, mentre agire sulle fasce di reddito e/o sul numero di componenti familiari richiederebbe un movimento lungo colonne o righe caratterizzate da importi Anf fortemente variabili. La dimensione del dato errato, come mostrato nel prospetto 2.2, non consente del resto di prescindere dal problema.

Relativamente al CASO 2 (componenti del nucleo familiare), la normativa potenzialmente ammette un qualsiasi valore che rientri nel range  $[1, +\infty)$ . Esigenze di ordine pratico (ingestibilità di tabelle eccessivamente allargate) e motivazioni di ordine 'biologico' (es. 40+ componenti) ci inducono a rigettare alcuni valori presenti nell'archivio ma ragionevolmente non ammissibili, fissando una soglia massima per la composizione del nucleo familiare. Tale soglia è stata stabilita pari a 15 componenti<sup>10</sup> e i valori che oltrepassano tale limite non vengono imputati direttamente al margine ma sono opportunamente trattati con procedure di tipo deterministico-stocastico. Questo consente di evitare un addensamento dei valori sulla soglia massima e di minimizzare la distorsione sulla variabilità del fenomeno.

<sup>10</sup> Il limite è stato prescelto tenendo conto dell'esiguo peso (0,35%) che le dichiarazioni con la coordinata 'numero di componenti del nucleo familiare' >15 avevano sul totale delle dichiarazioni, oltreché sulla base di evidenti considerazioni demografiche.

**Grafico 2.1 -Tipologie di coordinate non corrette o incomplete**



Le regole di imputazione sono state pensate per evitare di utilizzare procedure *ad hoc* definite sulla base dell'osservazione diretta del dato. L'obiettivo è stato quello di definire un approccio che fosse in linea con gli standard d'Istituto, uniforme nel tempo e valido per tutti o quasi tutti i tipi di errore. Ne è emerso un impianto di imputazione piuttosto complesso, basato su tecniche di tipo donatore, precedute da regole di filtraggio deterministico. L'idea di fondo è generare in un primo step un *pool* di potenziali donatori da cui stratificare secondo comuni caratteristiche (sesso e qualifica professionale). In un secondo step, un principio di distanza minima (basato sull'età) tra donatore e ricevente consentirà di individuare il donatore più simile in termini di vicinanza rispetto all'età del ricevente. In questo senso, le variabili utilizzate per la stratificazione e la definizione della distanza minima, rispettivamente sesso, qualifica professionale ed età, sono state scelte in funzione di una plausibile correlazione con le variabili da ricostruire.

**Prospetto 2.2 - Dichiarazioni con coordinate Anf non corrette per tipologia di errore (Archivio 770 Inps)**

Casistica errori	2000(*)	2001(*)	2002
caso 0	88.641 (71,1)	82.168 (66,7)	57.033 (60,4)
caso 1	32.140 (25,8)	30.124 (24,5)	29.896 (31,7)
caso 2	3.427 (2,7)	10.607 (8,6)	7.199 (7,6)
caso 3	549 (0,4)	284 (0,2)	277 (0,3)
caso 4	0 (0,00)	1 (0,0)	0 (0,00)
Tot. dichiarazioni da correggere	124.757 (100)	123.184 (100)	94.405 (100)

(\*) dati provvisori

Da un punto di vista statistico, rispetto al meccanismo di generazione del dato mancante, la procedura soffre del vincolo dell'ipotesi Mar (*missing at random*). L'assunzione Mar implica che per ricostruire l'informazione mancante sia sufficiente utilizzare l'informazione dei soli rispondenti. Tale ipotesi, seppure criticabile da un punto di vista statistico, è tuttavia difendibile in questo contesto, in quanto è ragionevole assumere che il dato mancante dalle coordinate non sia generato da un sorta di meccanismo di autoselezione, derivando piuttosto da errori non direttamente legati alle caratteristiche del rispondente (es. errori nella compilazione del modello 770 da parte del commercialista).

Uno dei vantaggi della procedura proposta è che consente di caratterizzare i gruppi di donatori in funzione sia dell'informazione nota sulle coordinate Anf, sia di caratteristiche individuali osservabili e simili per gruppi di rispondenti (età, sesso, qualifica professionale). In questo senso, la procedura tenderebbe ad operare un'ulteriore selezione all'interno del gruppo dei donatori, restringendo l'attenzione, per ogni ricevente, ai donatori più probabili non solo a livello di distribuzioni empiriche ma anche secondo caratteristiche osservabili simili. Rispetto a procedure più generali di tipo esclusivamente bayesiano, la tecnica di imputazione proposta possiede dunque il vantaggio di non relegare la scelta del donatore a criteri di natura casuale. In questo senso il *pool* dei donatori tende a restringersi per ogni sequenza di coordinate da imputare, ma la scelta del dato corretto è anche più attendibile in termini di similitudine socio-economica tra rispondente e donatore.

Di seguito, sono descritte in termini generali le linee di sviluppo della procedura di imputazione:

- CASO 0 (almeno una coordinata mancante), affinché la sequenza di coordinate possa essere sottoposta ad azione correttiva occorre che almeno una delle tre coordinate sia caratterizzata da valori ammissibili secondo riferimento normativo (cfr. circolari annuali Inps, contenenti le tabelle relative alla determinazione degli Anf) e/o secondo le procedure deterministico - stocastico già descritte:

[11,...21D]	per il riferimento tabellare
[1,.....15]	per il numero dei componenti familiari
[1,.....16]	per la classe di reddito (nelle tabelle dalla 11 alla 16)
[1,.....11]	per la classe di reddito (nelle tabelle dalla 17 alla 21D)

le coordinate che non rispondono a tali caratteristiche sono sottoposte a procedure d'imputazione. La selezione del *pool* dei donatori più simili viene effettuata lungo il sentiero delle coordinate pareto-ottimali;

- CASO 1 (coordinate formalmente corrette ma caratterizzata da Anf nullo), la combinazione di coordinate potenzialmente attribuibili è selezionata nel range dei valori più attendibili secondo un criterio di distanza minima rispetto alle righe e colonne (diagonale superiore ovvero sentiero delle coordinate pareto-ottimali). La scelta del donatore è successivamente definita in base al criterio generale di stratificazione ed estrazione secondo caratteristiche comuni osservabili;
- CASO 2 (numero di componenti il nucleo familiare superiore al valore massimo ammissibile): poiché sono state riscontrate alcune ricorrenze nei valori eccedenti i 15 componenti (maggiore addensamento nei valori che terminano con 0, ad es. 20, 30, 40...), è plausibile pensare che si tratti di errori di digitazione che con un certo margine di attendibilità possono essere ricondotti ai rispettivi valori omettendo lo '0'. Il riallineamento dei valori viene quindi effettuato secondo criteri deterministici; le restanti coordinate non imputabili entreranno a far parte del *pool* dei dati da imputare con procedure da *donor*;
- CASO 3 e CASO 4: appurato che almeno una delle tre coordinate sia significativa secondo i criteri sopra enunciati, si adotta la procedura generale da donatore.

Alcuni aspetti della procedura necessitano di essere puntualizzati:

- a) una potenziale complicazione della procedura di imputazione proposta deriva dal fatto che alcune selezioni potrebbero generare *pool* di donatori di dimensione inferiore alle 5 unità e, come noto, la teoria suggerisce che un campione di donatori inferiori a questa soglia non garantisce le buone proprietà di uno stimatore;
- b) esigenze di ordine statistico (*pool* di donatori inferiore alle 5 unità) ci impongono di ottimizzare la dimensione delle caratteristiche in funzione delle quali viene effettuata la stratificazione. In questo senso le variabili età e qualifica professionale sono state opportunamente riclassificate, riducendo il numero delle classi di età ed accorpando alcune tipologie di qualifica. Questa operazione è stata effettuata tenendo conto della compatibilità con variabili che si ritiene abbiano una relazione con l'età e la qualifica professionale. Quest'ultima, ad esempio, è stata riclassificata tenendo conto delle retribuzioni lavorative.
- c) la costruzione del *pool* dei donatori, richiede che la stratificazione avvenga secondo caratteristiche che ben approssimino il fenomeno in funzione del quale si effettua l'imputazione. Nello specifico caso, la scelta di età, sesso e qualifica professionale come variabili di stratificazione è principalmente dettata dalla quantità e qualità di informazione fornita dal mod. 770, che non ci consente di ricorrere a variabili che potrebbero essere più significative in termini di rappresentabilità del fenomeno (es. dato sulle competenze correnti). Motivazioni di ordine socio-economico ci consentono, tuttavia, di giustificare con discreto margine di attendibilità un legame tra qualifiche professionali e posizione reddituale (coordinata sulla classe di reddito) e tra sesso, età e composizione familiare.

## 2.4 Un approccio longitudinale per il raffinamento delle coordinate e del numero dei titolari

Corrette le coordinate Anf secondo le procedure appena descritte, una decisione essenziale riguarda la scelta di: a) produrre le statistiche utilizzando l'informazione corrente; b) produrre le statistiche sfruttando l'informazione longitudinale degli archivi. Nel primo caso verrebbe utilizzata solo l'informazione contenuta nell'anno di dichiarazione, nel secondo verrebbe sfruttato anche il patrimonio informativo aggiuntivo generato dall'integrazione di archivi di anni consecutivi (Figura 2.2).

Pur nella complessità della sua implementazione, la seconda opzione consente sensibili raffinamenti nell'individuazione dell'universo dei titolari. Nello specifico essa consente di ovviare almeno in parte alla perdita di informazione derivante dal fatto che nel 770 devono essere riportate le coordinate solo dei lavoratori che percepiscono Anf nel mese di dicembre.

La logica generale che sottende la procedura di individuazione del titolare di Anf è schematizzata nella figura 2.2. Per motivi computazionali si è deciso di tener distinte le dichiarazioni multiple (caso A) dalle dichiarazioni singole (caso B). Poiché l'incidenza delle prime sul totale è limitata, si è deciso di effettuare l'integrazione temporale solo per le dichiarazioni singole. Inoltre, all'interno della categoria delle dichiarazioni multiple è stata riscontrata la presenza di coordinate Anf anche per i record che non fanno riferimento a periodi retributivi comprensivi di dicembre. Pur contravvenendo alle norme di compilazione del modello 770, questa 'erronea' compilazione consente di ricostruire in modo articolato sia le variazioni contrattuali, sia le eventuali variazioni di coordinate Anf in corso d'anno, rendendo non necessaria l'integrazione con i dati dell'anno precedente. Nel caso invece di dichiarazioni singole, viene effettuata l'integrazione per codice fiscale con i dati dell'anno precedente. Se l'integrazione ha luogo (Caso B.1), si confrontano le coordinate Anf per i due anni. Nel caso di coincidenza (caso B.1.1) si applicano, ovviamente, a tutti i mesi retribuiti (MesiR) dell'anno. Se invece le coordinate Anf sono diverse (caso B.1.2), per i MesiR del primo semestre si applicano le coordinate dell'anno t-1 mentre per i MesiR del secondo semestre si applicano le coordinate dell'anno t. La logica sottesa a questa scelta deriva dal fatto che nel caso di dichiarazione singola le coordinate Anf presenti nel modello si riferiscono all'ultimo dei MesiR, ma non è detto che siano rimaste invariate per tutti i MesiR precedenti.

Tale supposizione deriva dalla stessa normativa che regola l'erogazione degli Anf. Come già detto, questa prevede che l'assegno spettante per l'anno t sia calcolato sul reddito dell'anno t-2 fino a giugno dell'anno corrente, e sul reddito dell'anno t-1 da luglio a dicembre.

Questo rende possibile una variazione della coordinata 'classe di reddito' in corso di anno, ipotesi avvalorata dai risultati riportati nel prospetto 2.3, che evidenziano come, per i percettori Anf presenti sia nel 2001 che nel 2002 (65,1% del totale dei percettori al 2002), le coordinate 'classe di reddito' variano da un anno all'altro nell'87% circa dei casi<sup>11</sup>.

**Prospetto 2.3 - Integrazione archivi percettori Anf anni 2002-2001**

Integrazione per tutte le coordinate				No Integrazione per Coordinata 1: Tabella		No Integrazione per Coordinata 2: Classe di reddito		No Integrazione per coordinata 3: Componenti del nucleo familiare	
Sì		No							
Numero	%	Numero	%	Numero	%	Numero	%	Numero	%
782.500	50,05	768.384	49,5	65.981	8,6	668.961	87,1	184.838	24,1

Il caso di mancata integrazione degli archivi riferiti a due anni (Caso B.2) può derivare dal fatto che un soggetto ha percepito Anf nell'anno t (ma non nell'anno t-1), o viceversa. Nel caso in cui il lavoratore ha percepito Anf solo nell'anno t (Caso B.2.a), per il calcolo dell'importo annuo erogato si applicano a tutti i mesi retribuiti (MesiR) dell'anno t le coordinate risultanti dal modello.

Nel caso in cui il lavoratore ha percepito Anf nell'anno t-1 (Caso B.2.b) e non risulta abbia beneficiato della prestazione nel mese di dicembre dell'anno t, la procedura è più articolata in quanto introduce ipotesi mirate al recupero di quei percettori che non sono direttamente desumibili dall'archivio M\_770 riferito all'anno t. A tale scopo ci si avvale dell'informazione proveniente dall'archivio generale dei lavoratori dipendenti per l'anno t,

<sup>11</sup> È però necessario chiarire che la circostanza che il reddito su cui calcolare l'assegno sia quello dell'anno t-2 fino a giugno e dell'anno t-1 da luglio in poi, non necessariamente genera una variazione della coordinata "classe di reddito". In altri termini non è detto che la differenza esistente tra i due redditi di riferimento comporti il trasferimento da una classe di reddito a quella successiva, potendo la variazione essere ricompresa entro gli estremi di una stessa classe.

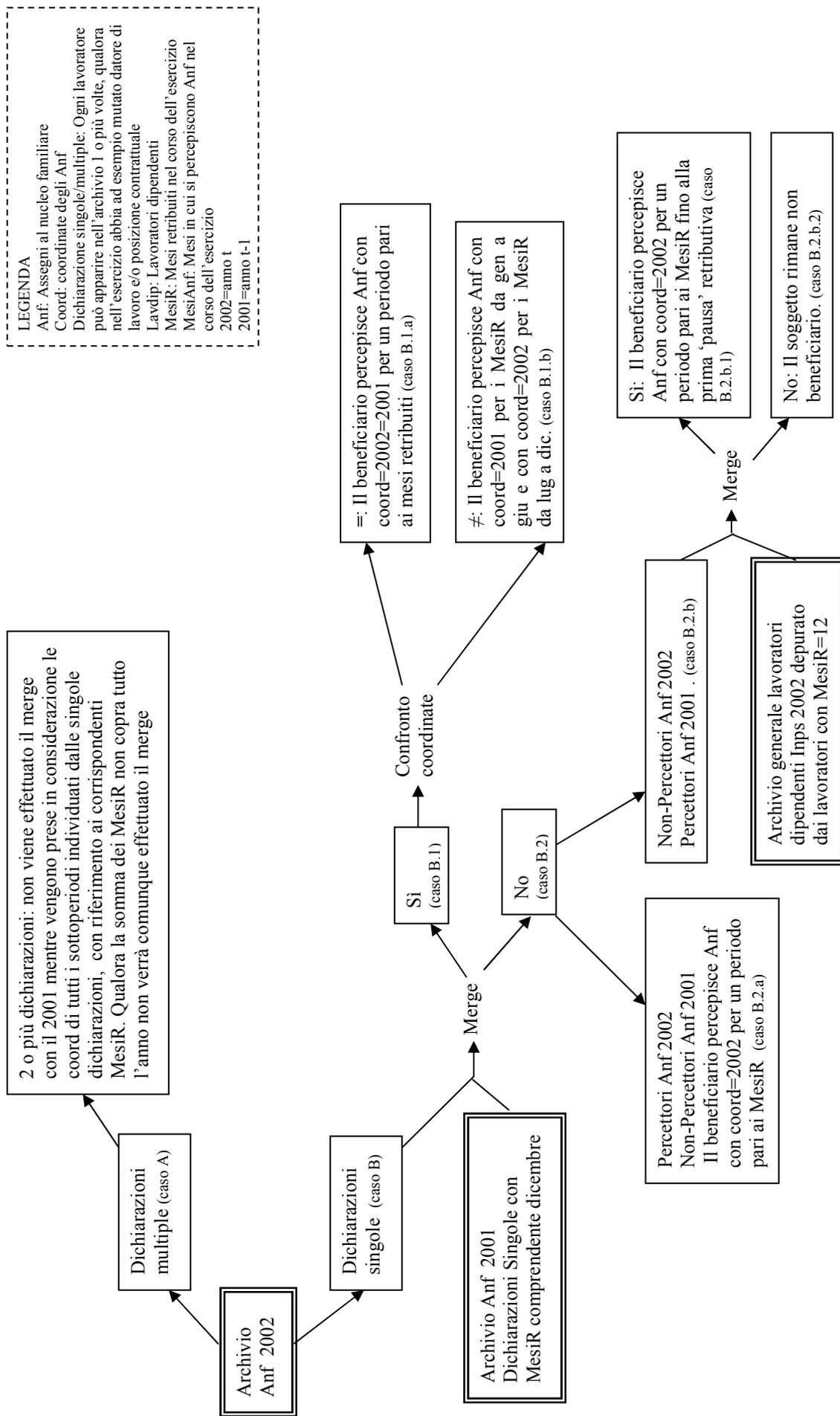
fornito dall'Inps. Più in dettaglio, si procede all'integrazione dell'archivio dei percettori di Anf nell'anno t-1, ma non nell'anno t, con quello dei lavoratori dipendenti dell'anno t, filtrato dai soggetti che nell'anno t non hanno lavorato nel mese di gennaio o che hanno lavorato per tutte le dodici mensilità. Il primo filtro deriva dal fatto che alla base della procedura c'è l'ipotesi che la continuità temporale da un anno al successivo nel ruolo di lavoratore dipendente sia indizio di una parallela continuità nel ruolo di percettore di Anf. Per questo motivo i percettori dell'anno t-1 devono esserlo al mese di dicembre, ed i lavoratori dell'anno t devono aver lavorato almeno a gennaio, in modo appunto che non ci sia alcun vuoto temporale nel passaggio da un anno al successivo.

Quanto invece al filtro volto ad escludere dall'archivio dei lavoratori dipendenti dell'anno t coloro che hanno lavorato per tutti i dodici mesi, questo viene utilizzato perché, se da un lato la procedura infrannuale consente di adottare una definizione di "titolare di Anf" più adatta a descrivere la reale dinamica del fenomeno in analisi, dall'altro lato espone al rischio di una sovrastima del numero dei percettori e dell'ammontare complessivo erogato. Proprio in considerazione di questo rischio si è deciso quindi di escludere dal "serbatoio" dal quale attingere titolari recuperati dall'anno t-1 coloro che nell'anno t hanno lavorato per dodici mesi e che quindi per quel medesimo periodo verrebbero recuperati come percettori. In termini molto semplici si è ritenuto che se si può ragionevolmente ipotizzare un parallelismo tra la continuità da un anno al successivo della condizione di lavoratore dipendente con quella di titolare di Anf, non sarebbe tuttavia prudente spingere le conseguenze di questa supposizione fino al punto di far apparire percettori per 12 mesi soggetti che senza la procedura infrannuale sarebbero totalmente esclusi dal computo dell'erogazione degli Anf. Tornando all'integrazione dei due archivi così filtrati, i soggetti per i quali la stessa non sia verificata (caso B.2.b.2) rientrano nella categoria dei non-percettori<sup>12</sup>.

---

<sup>12</sup> Ovviamente alla base dell'integrazione con l'archivio dei lavoratori dipendenti c'è l'ipotesi che, nell'ambito dell'archivio 770, non può essere beneficiario di Anf chi non risulti lavoratore dipendente nel periodo rispetto al quale si intende rilevare la prestazione.

**Figura 2.2 - Percorso logico di individuazione dei percettori di Anf e di determinazione dell'ammontare dell'assegno**



Nei casi in cui l'integrazione è verificata (caso B.2.b.1) si ipotizza che il soggetto sia percettore di Anf per un periodo pari al numero di MesiR continuativi fino alla prima pausa retributiva. I risultati della simulazione della procedura appena descritta sui dati del 2001 e del 2002 sono riportati nel prospetto 2.4, che indica un recupero di percettori pari al 7,15% del totale.

**Prospetto 2.4 - Distribuzione del tipo di dichiarazione per natura dei percettori**

Tipo di percettori	Anno 2002 (t)	
	Numero	%
Percettori presenti anno t e t-1	1.547.386	70,22
Percettori presenti solo anno t	445.810	20,23
Percettori con dichiarazioni multiple nell'anno t	52.834	2,40
Percettori recuperati dall' anno t-1 non continuativamente retribuiti nell'anno t	157.439	7,15
<b>Totale</b>	<b>2.203.469</b>	<b>100,00</b>

Nella figura 2.3 viene sintetizzato il risultato ottenuto nel popolamento dell'archivio seguendo i due diversi approcci. Il primo è quello che limita l'analisi ai dati relativi ad un unico anno (Approccio A); il secondo invece fa riferimento alla procedura di integrazione di archivi di due anni consecutivi (Approccio B). Quest'ultimo approccio consente: 1) il recupero di percettori che in base alla regola di compilazione "solo a dicembre" non verrebbero rilevati e 2) il controllo dell'evoluzione delle coordinate da un anno all'altro.

**2.5 Determinazione dell'importo dell'assegno**

L'ipotesi alla base della determinazione dell'importo annualmente erogato per il pagamento di Anf, già citata nei paragrafi precedenti, è che il periodo di percepimento degli Anf coincida con il periodo retribuito (MesiAnf=MesiR). Tuttavia, tale ipotesi va raffinata in sede di stima degli importi Anf, in quanto l'effettivo diritto alla corresponsione dell'assegno familiare sorge al raggiungimento di un monte minimo mensile di ore lavorate. In altri termini, perché in un dato mese abbia diritto all'Anf non è sufficiente che il soggetto abbia lavorato: occorre quantificare il numero effettivo di ore lavorative prestate.

Per fornire una stima del numero di mesi in cui gli assegni familiari sono stati effettivamente percepiti, ovvero per effettuare quel raffinamento che permette di convertire i MesiR in MesiAnf, è stato calcolato un coefficiente di correzione che tiene conto, oltre che del vettore dei mesi retribuiti, di informazioni aggiuntive derivanti da varie fonti: modello 770 (settimane e giornate retribuite), fonte normativa (limite minimo di ore lavorative mensili ai fini dell'erogazione degli Anf, orario settimanale contrattuale per categorie lavorative). Il coefficiente di correzione è calcolato come segue:

$$Coeff\ corr\ MR = \frac{\text{GiorniR}}{\text{MesiR}} \cdot hgn \cdot (hmens\_Inps)^{-1}$$

*Coeff corr MR > 1*    MesiAnf=MesiR  
*Coeff corr MR < 1*    MesiAnf=MesiR\* *Coeff corr MR*

*GiorniR* = Giorni retribuiti

*MesiR* = Mesi retribuiti

*hgn* = numero medio di ore lavorate giornaliere calcolate secondo normativa (6 per impiegati e dirigenti, 6.6 per operai)

*hmens\_Inps* = limite minimo di ore lavorative mensili per il diritto agli Anf

Il coefficiente sopra descritto permette di calcolare i MesiAnf riponderando i MesiR in base al numero di ore mediamente lavorate in un mese rispetto al limite minimo di ore mensili necessarie affinché si abbia diritto alla corresponsione dell'assegno<sup>13</sup>.

In tutti i casi in cui il coefficiente assume valori uguali o maggiori di 1, si stabilisce che il soggetto ha diritto all'assegno Anf per tutti i MesiR (MesiAnf=MesiR), in quanto per tutti i mesi retribuiti il numero delle ore

<sup>13</sup> Tale limite è stabilito dall'Inps in 130 ore mensili per gli operai e 104 ore mensili per gli impiegati. Cfr. Dpr.797/55, art.59.

lavorate è stato uguale o superiore al limite imposto dall'Inps perché venga riconosciuto il diritto alla corresponsione di Anf. Se il coefficiente assume invece un valore inferiore ad 1, ossia se il numero di ore lavorate non è uguale o superiore per tutti i mesi retribuiti al limite minimo di ore lavorative mensili per il diritto all'Anf, il numero di MesiAnf viene ridotto in proporzione ( $MesiAnf < MesiR$ ). Il numero dei mesi nei quali si ha effettivo diritto agli Anf viene infine moltiplicato per l'importo dell'assegno mensile spettante, individuato tramite i riferimenti tabellari. In questo modo si arriva a determinare l'importo Anf annuale percepito. Nella figura 2.4 è illustrata schematicamente la logica sottostante al calcolo dell'importo degli Anf percepiti annualmente. Di seguito vengono, invece, descritte le procedure di calcolo con riferimento alle diverse casistiche descritte<sup>14</sup>.

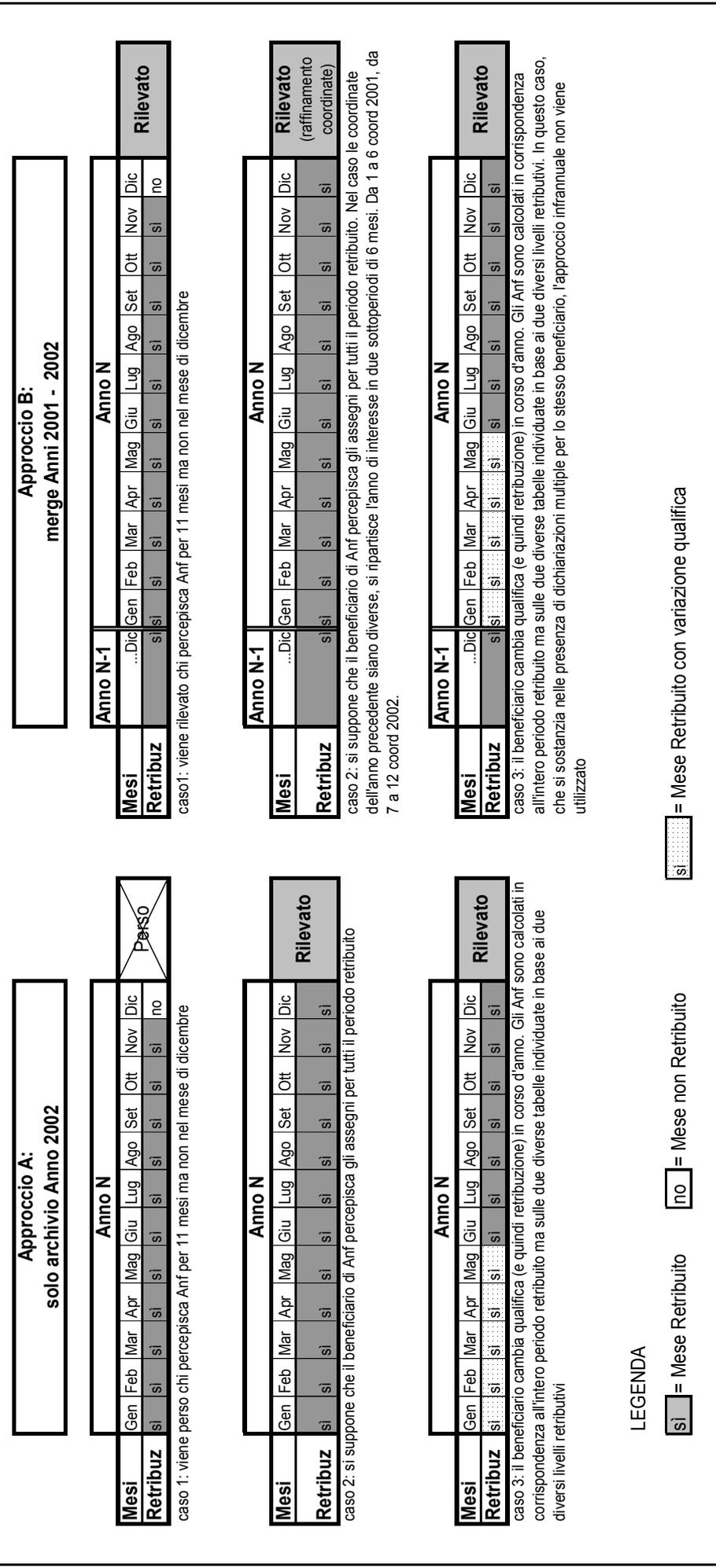
*Calcolo dell'assegno nel caso di dichiarazione singola* (Figura 2.5): Nel caso di dichiarazione singola si procede all'integrazione dei dati dell'anno t con quello dei dati dell'anno t-1. Il procedimento per il calcolo dell'importo percepito risulta differente a seconda che siano presenti o meno coordinate Anf nell'anno t-1 e, in caso affermativo, in base alla loro invarianza rispetto alle coordinate dell'anno t:

- *Coordinate invarianti* (esempio 1): Le coordinate dell'anno t-1 sono uguali a quelle dell'anno t e si applicano a tutti i MesiAnf. È il caso più semplice e ricalca il procedimento generale appena descritto.
- *Coordinate diverse* (esempio 2): in questo caso ai MesiAnf del primo semestre si applicano le coordinate dell'anno t-1 e ai MesiAnf del secondo semestre quelle dell'anno t. Dal punto di vista operativo si procede alla determinazione dei MesiAnf totali con il metodo già illustrato. Il risultato viene poi ripartito tra primo e secondo semestre usando come pesi i MesiR del primo e secondo semestre. Così, nell'esempio 2, i MesiR sono 11, di cui 5 mesi nel primo semestre e 6 nel secondo. Quindi per il primo semestre i MesiAnf saranno ottenuti moltiplicando il coefficiente correttivo (se  $< 1$ ) per i MesiR compresi nel primo semestre (in questo caso 5). I MesiAnf del secondo semestre saranno ottenuti, ovviamente, moltiplicando il coefficiente correttivo (se  $< 1$ ) per il numero di MesiR nel secondo semestre (in questo caso 6). Definite in questo modo le durate dei due periodi caratterizzati da coordinate diverse, si procederà come di consueto a moltiplicare il numero dei MesiAnf per l'assegno spettante in base alle diverse coordAnf dei due semestri.

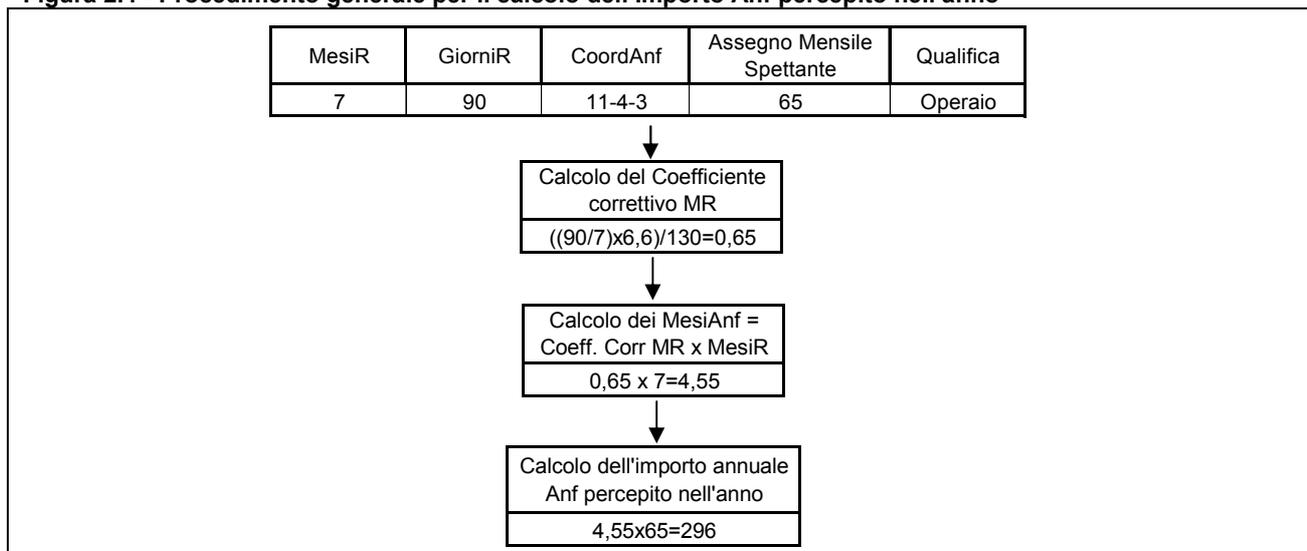
---

<sup>14</sup> Cfr. Figura 3.

Figura 2.3 - Confronto tra metodologie per la determinazione dell'importo dell'assegno Anf



**Figura 2.4 - Procedimento generale per il calcolo dell'importo Anf percepito nell'anno**



- *Coordinate assenti nell'anno t-1* (esempio 3): è il caso in cui l'integrazione tra i due anni non è verificata perché il soggetto è percettore nell'anno t ma non nell'anno t-1. Come nel caso di coordinate invarianti, le coordinate presenti nell'anno t vengono applicate a tutti MesiAnf.
- *Coordinate assenti nell'anno t* (esempio 4): è il caso in cui non c'è integrazione tra i due anni perché il soggetto è percettore nell'anno t-1 ma non a dicembre dell'anno t. Si suppone allora che il soggetto percepisca Anf con coordinate t-1 per tutti i mesi continuativi fino alla prima interruzione, nel caso di MesiR < 12, l'unico ammissibile.

*Calcolo dell'assegno nel caso di dichiarazione multipla:* nel caso di dichiarazione multipla (cfr Figura 2.6) la situazione lavorativa dell'individuo è ripartita in più dichiarazioni che fanno riferimento a sottoperiodi lavorativi distinti all'interno del medesimo esercizio amministrativo. Come già detto, questa situazione si verifica ad esempio nel caso di variazioni di qualifica, di posizione contrattuale o di azienda di appartenenza. Questi eventi non necessariamente generano variazioni nelle coordAnf.

- *Coordinate invarianti nei sottoperiodi* (esempio 5): il calcolo sarà identico a quello descritto nell'esempio 1 relativo alle dichiarazioni singole, in quanto i sottoperiodi descritti nelle diverse dichiarazioni saranno trattati congiuntamente;
- *Coordinate diverse nei sottoperiodi* (esempio 6): Ogni sottoperiodo viene trattato separatamente, calcolando il numero dei MesiAnf e moltiplicandolo per l'assegno mensile spettante in base alle coordinate del periodo. L'importo totale percepito nell'anno risulterà quindi dalla somma dei totali percepiti nei singoli sottoperiodi.

Il procedimento di calcolo dei MesiAnf consente anche di ovviare al problema connesso alla sovrapposizione dei MesiR che può verificarsi agli estremi dei diversi sottoperiodi. Nella figura 2.6, esempio 5, il mese di marzo appare due volte, alla fine del primo sottoperiodo ed all'inizio del secondo. La somma dei MesiR risulta quindi pari a 12, anche se i mesi solari retribuiti sono 11. La differenza è appunto causata dal duplice conteggio del mese di Marzo. Tuttavia questa circostanza (che comunque riguarda una casistica quantitativamente piuttosto esigua) non disturba in alcun modo il calcolo dell'Anf annuale, in quanto questo si basa, in ultima analisi, non sul numero di mesi retribuiti, ma su quello dei giorni retribuiti diviso quello dei mesi retribuiti. Nel caso di 'mese sovrapposto' questo garantisce che in ogni sottoperiodo il mese in questione non 'peserà' per la sua intera durata, ma solo per i giorni effettivamente retribuiti nel sottoperiodo. In altri termini la procedura adottata risolve il problema del 'mese sovrapposto', grazie all'utilizzo dei giorni retribuiti, per i quali non esisterà sovrapposizione alcuna (a meno che non si voglia supporre che un individuo lavori nel medesimo giorno per due diversi datori di lavoro, ipotesi giuridicamente plausibile ma statisticamente non significativa).

Figura 2.5 - Esempi di calcolo dell'importo annuale Anf per dichiarazioni singole

Esempio 1: coordinate invarianti	Anno N												Coord Anno N - 1	Coord Anno N	Importo mensile Anno N - 1	Importo mensile Anno N	MesiR	GiorniR	Qualifica	Coefficiente di correzione	MesiAnf	Calcolo Anf Annuale
	Anno N - 1			Anno N			Anno N															
	Gen	Feb	Mar	Apr	Mag	Giu	Lug	Ago	Set	Ott	Nov	Dic										
...	si	si	si	no	no	si	si	si	si	si	si	si	11 - 4 - 4	158	158	10	140	Operai	$((140/10) \times 8) / 130 = 0,86$	$11 \times 0,86 = 9,47$	$158 \times 9,47 = 1738$	
...	si	si	si	no	no	si	si	si	si	si	si	si	11 - 4 - 3	65	158	11	135	Impiegato	$((135/11) \times 8) / 104 = 0,94$	$5 \times 0,94 = 4,7$ $6 \times 0,94 = 5,6$	$4,7 \times 65 + 5,6 \times 158 = 1190$	
...	no	no	no	no	no	no	no	no	no	no	no	no	Assenti	Assenti	Assenti	12	260	Operai	$((260/12) \times 8) / 130 = 1,33$	12	$158 \times 12 = 1896$	
...	si	si	si	no	no	no	no	no	no	no	no	no	11 - 4 - 3	65	Assenti	8	95	Operai	$((95/8) \times 8) / 130 = 0,73$	$8 \times 0,73 = 5,85$	$65 \times 5,85 = 380$	

Figura 2.6 - Esempi di calcolo dell'importo annuale Anf per dichiarazioni multiple

Esempio 5: Coordinate invarianti nei sottoperiodi	Anno N												Coord	Importo mensile	MesiR	GiorniR	Qualifica	Coefficiente di correzione	MesiAnf (a)	Calcolo Anf Annuale	
	Anno N - 1			Anno N			Anno N			Anno N											
	Gen	Feb	Mar	Apr	Mag	Giu	Lug	Ago	Set	Ott	Nov	Dic									
Dichiarazione 1	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 4	158	3	50	Operai	$((223/11) \times 6,6) / 130 = 1,03$	11	$11 \times 158 = 1738$	
Dichiarazione 2	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 4	158	5	98	Operai				
Dichiarazione 3	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 4	158	2	40	Operai				
Dichiarazione 4	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 4	158	2	35	Operai				
...	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 3	65	Assenti	Assenti	Assenti	Assenti	Assenti	Assenti	Assenti
...	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 4	158	4	70	Operai				
...	no	no	no	no	no	no	no	no	no	no	no	no	11 - 4 - 4	158	5	67	Operai				



## Capitolo 3 – Determinazione degli importi totali e valutazione delle stime

### 3.1 Assegni al nucleo familiare percepiti: la determinazione degli importi totali

Dopo aver seguito le procedure sopradescritte per individuare i titolari di Anf, definite le coordinate, i “periodi di corresponsione” e l’importo dell’assegno annualmente percepito, si può procedere alla stima dell’ammontare complessivo degli Anf.

Prima di effettuare questa operazione è però necessario ‘ripulire’ l’archivio 770 dalle matricole aziendali riferibili a datori di lavoro pubblici. Infatti i dati raccolti attraverso la parte C, sezione *Dati previdenziali ed assistenziali Inps ed Inpdai* del modello 770, ineriscono anche particolari tipi di dipendenti pubblici (tipicamente i contrattisti a progetto, ma anche altre categorie speciali, quali ad esempio i detenuti dipendenti dalle amministrazioni penitenziarie), la cui gestione previdenziale afferisce all’Inps ed i cui dati previdenziali sono quindi registrati nella sezione citata del modello 770. Al fine di escludere dall’archivio 770 le dichiarazioni riferite a questi soggetti, ottenendone una versione esclusivamente riferita a lavoratori dipendenti del settore privato, i soli sui quali si possono produrre statistiche Tmpn basate sull’archivio in questione, si è utilizzato il codice Csc (Codice statistico contributivo)<sup>15</sup> associato a ciascuna matricola aziendale<sup>16</sup>. In pratica l’archivio è stato filtrato escludendo le matricole aziendali caratterizzate da un valore del primo carattere del codice Csc pari a 2 (Enti Pubblici) o 3 (Amministrazioni Statali non soggette alla disciplina degli Af). Il peso delle matricole aziendali così individuate è pari allo 0,26% (1.385 matricole) del totale archivio M\_770<sup>17</sup> (532.379 matricole).

Ottenuto in questo modo l’archivio dei percettori Anf appartenenti al settore privato, si può procedere alla stima dell’ammontare complessivo degli Anf, sommando tra loro gli importi annuali erogati ai singoli titolari del trattamento, ed al confronto con i dati dell’archivio Dm10, utilizzato per la verifica della coerenza di stime del medesimo fenomeno prodotte a partire da differenti fonti amministrative disponibili presso l’Istituto.

#### 3.1.1 Matricole aziendali nel 770 e nel Dm10: la popolazione degli archivi

La valutazione delle stime prodotte utilizzando i dati dell’archivio 770 rispetto agli ammontari risultanti da Dm10 ha senso se effettuata su popolazioni speculari, prendendo cioè in considerazione solo le matricole

---

<sup>15</sup> Cfr. *Manuale di istruzioni in materia di iscrizione e classificazione delle aziende* allegato alla Circ.Inps n.65 del 28 marzo 1996. Il codice, composto di cinque cifre, è utilizzato dall’Inps per classificare le aziende, consentendo «...le opportune rilevazioni statistiche [individuando altresì] quale deve essere il comportamento contributivo dell’azienda». Il primo carattere del codice individua il ramo economico, il secondo la classe economica, il terzo la categoria, secondo una logica analoga a quella della classificazione Ateco.

<sup>16</sup> La matricola aziendale è un codice numerico assegnato dall’Inps per individuare la posizione contributiva di un’impresa. Ogni impresa può avere una o più matricole aziendali ma un solo codice fiscale. Questo deriva dal fatto che, ad esempio, a stabilimenti diversi (ad esempio dislocati in diverse regioni italiane) vengono assegnati diverse matricole aziendali riconducibili però al medesimo codice fiscale aziendale.

<sup>17</sup> Come evidenziano i prospetti (dal 3.6 al 3.8), la successiva disaggregazione delle stime per classificazione Ateco ha evidenziato la presenza di 21 matricole aziendali riferibili a datori di lavoro pubblici. Per completezza di informazione si è scelto di presentare i dati tenendo conto anche di questo esiguo numero di matricole aziendali appartenenti alla sezione L, che però non verranno utilizzate in sede di eventuale pubblicazione delle stime sugli importi degli assegni al nucleo familiare erogati. La ripulitura dell’archivio dai percettori Anf dipendenti pubblici a valle del processo di imputazione delle coordinate Anf e di stima dell’importo annuale spettante al singolo beneficiario non inficia la validità dei risultati ottenuti per le seguenti ragioni:

- sesso, qualifica professionale ed età, le variabili utilizzate per stratificare i soggetti donatori nel processo di imputazione delle coordinate Anf, sono caratteristiche che si possono ragionevolmente ritenere indipendenti rispetto alla natura pubblica o privata del datore di lavoro;
- i dipendenti pubblici presenti nell’archivio, oltre ad essere quantitativamente marginali rispetto al totale dei percettori Anf, sono dipendenti pubblici ‘particolari’, che, soprattutto per il contratto di lavoro e per la qualifica, sono in certo qual modo equiparabili, dal punto di vista della gestione previdenziale, a lavoratori del settore privato.

aziendali presenti in entrambi gli archivi. In merito è necessario precisare che la matricola aziendale è la sola chiave utilizzabile in quanto l'unica presente in entrambi gli archivi<sup>18</sup>.

Il 96% delle matricole aziendali (Prospetto 3.1) presenti nell'archivio M\_770 si integrano con quelle dell'archivio Dm10, mentre il restante 3,57% (19.085 matricole aziendali) non trova corrispondenza nell'archivio Dm10, nel quale sono invece presenti 71.441 matricole aziendali assenti nell'archivio M\_770.

### Prospetto 3.1 - Matricole aziendali - confronto dati 770 e Dm10

Archivio	Matricole Aziendali presenti			
	Cod 0035		Cod "Tutti"	
	Numero	Percentuale su totale 770 Inps	Numero	Percentuale su totale 770 Inps
Totale 770 Inps	532.379	100,00%	532.379	100,00%
Totale Dm10	586.253		589.895	
<b>Integr 770-Dm10</b>	<b>513.294</b>	<b>96,42%</b>	<b>515.092</b>	<b>96,75%</b>
Solo 770 Inps	19.085	<b>3,58%</b>	18.805	<b>3,53%</b>
Solo Dm10	71.441		74.803	

La comparazione delle popolazioni dei due archivi viene realizzata con esclusivo riferimento al sottouniverso rappresentato da datori di lavoro che erogano Anf. Per l'archivio Dm10 tale "filtro" viene realizzato selezionando solo le matricole aziendali alle quali si associa la presenza del codice 0035, che individua l'erogazione di Anf correnti. Il codice citato non è l'unico del Dm10 legato alla prestazione Anf. Si è tuttavia ritenuto, tenendo conto della procedura di stima adottata e dell'universo di riferimento costituito dall'archivio 770, che il codice 0035 fosse l'unico da considerare ai fini del raffronto.<sup>19</sup>

Operativamente si è quindi proceduto come segue: 1) vengono estratte dall'archivio Dm10 le matricole aziendali associate ad un valore del codice 0035 diverso da zero; 2) l'elenco delle matricole aziendali così selezionate viene confrontato con quello delle matricole presenti sull'archivio M\_770. La precisione delle stime verrà verificata sulla base delle matricole aziendali comuni ai due archivi. Le matricole aziendali presenti solo nell'archivio dei modelli 770 (da questo punto in poi, matricole aziendali 'solo770') sono invece state analizzate per valutare l'opportunità di ricomprenderle nella popolazione complessiva in riferimento alla quale elaborare le stime Anf definitive, verificandone l'eventuale effetto distorsivo sulle stesse. Nel prospetto 3.1 sono riportati anche i risultati dell'integrazione effettuata prendendo in considerazione tutti i codici<sup>20</sup>. I risultati dell'operazione non si discostano in maniera significativa rispetto a quelli ottenuti considerando solo il codice 0035.

Venendo alle matricole aziendali presenti nell'archivio M\_770 e non nell'archivio Dm10, tale fenomeno può avere molteplici spiegazioni. Innanzitutto pur avendo un eguale significato originario, tali codici identificativi vengono sottoposti a procedure di controllo che nei due archivi considerati sono totalmente indipendenti. Nell'archivio M\_770 viene verificata la coerenza della matricola aziendale con il periodo contributivo di riferimento. Un errore ricorrente deriva infatti dalla circostanza che datori di lavoro ai quali è stata assegnata una nuova matricola aziendale, attivata nell'anno t, usino la matricola in questione anche per dichiarazioni riferite all'anno t-1 o antecedente. In questo caso l'Inps provvede all'imputazione della matricola aziendale corretta, operazione che può generare incoerenze con l'archivio Dm10, nel quale non viene effettuato lo stesso tipo di controllo.

L'altra fondamentale causa della non coincidenza delle popolazioni delle matricole aziendali dei due archivi è legata alla diversa data di rilascio degli stessi.

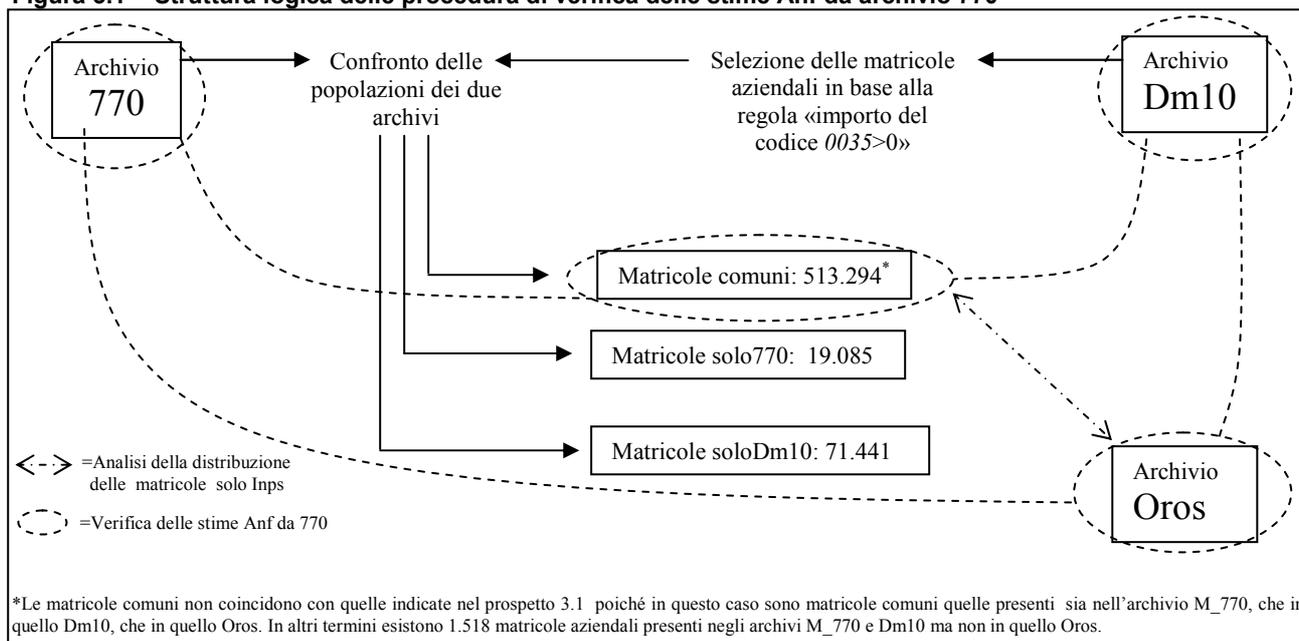
Per indagare ulteriormente le ragioni sottese alla diversa popolazione degli archivi 770 e Dm10 e per verificarne eventuali conseguenze sull'affidabilità delle stime realizzate, si è deciso di utilizzare l'archivio delle aziende 2002, elaborato dall'Istat nell'ambito dell'indagine su Occupazione, Retribuzioni e Oneri Sociali (Oros) (cfr Figura 3.1).

<sup>18</sup>A differenza del codice fiscale aziendale, presente solo nell'archivio Dm10

<sup>19</sup>Infatti gli altri codici che nel Dm10 individuano prestazioni Anf sono: G821 (Anf di cui CIGO) ed H301 (Anf indennità TBC), che fanno riferimento a prestazioni Anf erogate direttamente dall'Inps, e quindi non presenti nell'archivio M\_770; L036 (Arretrati Anf) che fa riferimento a prestazioni che non possono essere considerate parte degli ammontari stimabili a partire dall'archivio M\_770, nei quali sono presenti le coordinate in base alle quali la persona fisica ha diritto all'erogazione di Anf nel periodo di riferimento (prestazioni correnti).

<sup>20</sup> Cfr. nota precedente.

**Figura 3.1 - Struttura logica delle procedura di verifica delle stime Anf da archivio 770**



Il confronto effettuato per le 19.085 matricole presenti solo nell'archivio 770 ha evidenziato come il 57,2% di queste si associno ad un numero di dipendenti<sup>21</sup> pari a 0 (Prospetto 3.2). L'ulteriore disaggregazione di queste matricole aziendali con 0 dipendenti in base ad eventi rilevanti della vita aziendale (anche questi desumibili dai dati Oros) permette di spiegare ancor meglio gli effetti che le diverse date di rilascio dei due archivi hanno sulla popolazione degli stessi. Nel 35,7% dei casi si tratta di aziende attive, e quindi la presenza di 0 dipendenti può spiegarsi quasi esclusivamente con il ritardato invio della dichiarazione Dm10 all'Inps.

Per il 56,9% dei casi si tratta di matricole aziendali a cui corrispondono eventi di sospensione o riattivazione, eventi aziendali straordinari che lasciano intuire come, anche in questo caso, l'assenza di dipendenti derivi da un ritardato invio della dichiarazione. Il restante 7,3% riguarda invece imprese che risultano cessate per l'archivio Oros (che elabora dati da fonte Dm10), ma non altrettanto per quello 770 (nel quale, se così fossero, non avrebbero ragione d'essere presenti). Queste matricole (795) sono quelle per le quali esiste un'incoerenza tra dati 770 e dati Dm10 non spiegabile con l'ipotesi delle diverse date di scarico degli archivi, ma piuttosto con le procedure amministrative di controllo e modifica delle matricole aziendali nell'archivio 770 in uso presso l'Inps.

**Prospetto 3.2 - Matricole aziendali presenti esclusivamente nell'archivio 770**

Matricole Aziendali	n	%
Zero dipendenti	10.919	57,20%
- attiva	3.904	35,70%
- cessata	795	7,30%
- sospesa	4.339	39,70%
- riattivata	1.881	17,20%
1 o più dipendenti	8.166	42,80%
- attiva	5.183	63,50%
- cessata	74	0,90%
- sospesa	711	8,70%
- riattivata	2.198	26,90%
<b>Totale</b>	<b>19.085</b>	<b>100,00%</b>

<sup>21</sup> Il numero dei dipendenti viene stimato sulla base dell'archivio Oros.

Le matricole aziendali con 1 o più dipendenti (42,8%) riguardano per il 63,5% realtà aziendali attive, per il 35,6% sospese o riattivate nell'esercizio di interesse, per le quali valgono le osservazioni appena esposte per le matricole aziendali con 0 dipendenti.

Più in generale la classificazione delle 19.085 matricole per classe dimensionale, evidenzia come il 79,8% afferisca ad imprese fino a 10 dipendenti, il 90% imprese fino a 20 dipendenti. Questo fornisce un'ulteriore chiave interpretativa circa la presenza di matricole aziendali *solo770*, che si rivelano per lo più legate a situazioni che sovente complicano la rilevazione statistica: aziende di ridotte dimensioni e caratterizzate da eventi vitali rilevanti quali cessazione, sospensione o riavvio.

Delle matricole *solo770* si è altresì analizzata la distribuzione rispetto alle sezioni Ateco (Prospetto 3.3). Questa distribuzione mostra sensibili scostamenti rispetto a quella che caratterizza la popolazione delle 527.945 matricole comuni solo per due sezioni Ateco: la D (attività manifatturiere) che nella distribuzione

**Prospetto 3.3 - Matricole aziendali per sezione Ateco, confronto matricole comuni e matricole solo archivio 770**

SEZIONI ATECO	Matricole aziendali					Differenza delle distribuzioni in punti percentuali
	Comuni		Solo 770			
	n	% sul totale	n	% sul totale		
<b>A</b> Agricoltura, caccia e silvicoltura	1.801	0,4	174	0,9	-0,6	
<b>B</b> Pesca, piscicoltura e servizi connessi	2.613	0,5	601	3,1	-2,6	
<b>C</b> Estrazione di minerali	2.315	0,5	50	0,3	0,2	
<b>D</b> Attività manifatturiere	155.041	30,2	2.334	12,2	18,0	
<b>E</b> Produzione e distribuzione di energia elettrica, gas e acqua	1.034	0,2	22	0,1	0,1	
<b>F</b> Costruzioni	100.948	19,7	6.225	32,6	-13,0	
<b>G</b> Commercio all'ingrosso e al dettaglio; riparazione di autoveicoli, motocicli e di beni personali e per la casa	104.991	20,5	4.103	21,5	-1,0	
<b>H</b> Alberghi e ristoranti	31.516	6,1	1.095	5,7	0,4	
<b>I</b> Trasporti, magazzinaggio e comunicazioni	28.588	5,6	546	2,9	2,7	
<b>J</b> Attività finanziarie	4.587	0,9	183	1,0	-0,1	
<b>K</b> Attività immobiliari, noleggio, informatica, ricerca, servizi alle imprese	52.112	10,2	1.316	6,9	3,3	
<b>L</b> Amministrazione pubblica	21	0,0	233	1,2	-1,2	
<b>M</b> Istruzione	2.921	0,6	490	2,6	-2,0	
<b>N</b> Sanità e assistenza sociale	10.325	2,0	791	4,1	-2,1	
<b>O</b> Altri servizi pubblici, sociali e personali	14.324	2,8	809	4,2	-1,4	
<b>Q</b> Organizzazioni ed organismi extraterritoriali	29	0,0	1	0,0	0,0	
<b>X</b> Non ripartibili	128	0,0	112	0,6	-0,6	
<b>Totale</b>	<b>513.294</b>	<b>100,0</b>	<b>19.085</b>	<b>100,0</b>	<b>0,0</b>	

delle matricole *solo770* pesa 18 punti percentuali in meno rispetto a quanto avvenga per la distribuzione della matricole comuni, e la F (costruzioni), che pesa 13 punti percentuali in più. Per tutte le altre sezioni Ateco, pari al 50,1% del totale delle matricole aziendali comuni, la differenza è risultata inferiore ai 3,3 punti percentuali. E, come verrà approfondito più avanti, questo consente di comprendere le 19.085 matricole *solo770* nella popolazione complessiva sulla quale elaborare le stime definitive, senza produrre effetti distorsivi sulle stesse.

### 3.2 Verifica delle stime Anf da archivio 770 rispetto ai dati di bilancio Inps ed al Dm10

Per valutare la precisione delle stime e quindi la loro affidabilità in termini statistici, gli importi Anf stimati tramite gli archivi 770 sono stati posti a raffronto con i dati da bilancio Inps e con quelli da archivio Dm10. Il dato di benchmark fondamentale, è quello derivante dal bilancio consuntivo 2002 dell'Inps. La gestione Inps deputata all'erogazione degli Anf ai lavoratori dipendenti del settore privato è la *Gestione prestazioni temporanee ai lavoratori dipendenti*. Come evidenziato nel prospetto 3.4, rispetto al valore di bilancio

dell'Inps<sup>22</sup>, quello elaborato a partire dall'archivio 770 risulta inferiore dell'8,3%, una sottostima accettabile che conferma la validità del processo di stima adottato.

Il confronto con il Dm10 è invece utile soprattutto per due motivi:

- a differenza di quanto avviene per il dato di bilancio Inps, è possibile effettuarlo anche in forma disaggregata, rispetto ad esempio alla classificazione Ateco o alla dimensione aziendale;
- quella da Dm10 è una stima derivante da un altro degli archivi amministrativi disponibili presso l'Istat; il confronto è quindi utile anche nell'ottica di una verifica di coerenza delle stime ottenibili, sul medesimo fenomeno, a partire da diverse fonti disponibili presso l'Istituto.

Rispetto a quanto avveniva nei confronti del dato Inps (che rimane, va ribadito, il dato di benchmark), la prospetto 3.4 evidenzia tra stima da 770 e stima da Dm10 uno scostamento leggermente più ampio, pari all'8,6% e di segno opposto.

Come già sottolineato, tale ultimo raffronto è stato effettuato prendendo in considerazione, per l'archivio Dm10, solo i record con matricole aziendali (e relativi ammontari) associate a valori della variabile 'codice 0035' diversi da zero. Il confronto tra le stime è stato ulteriormente approfondito disaggregando gli importi Anf per classe dimensionale e classificazione Ateco.

**Prospetto 3.4 - Stime Anf da archivio 770 Inps, da bilancio Inps e da archivio Dm10 (dati in migliaia di euro)**

Stima da 770	Bilancio Consuntivo Inps ( <i>benchmark</i> )		Stima da Dm10	
	Ammontare	Differenza percentuale	Ammontare	Differenza percentuale
2.336.090	2.548.731	-8,3	2.151.865	8,6

**3.2.1 Verifica delle stime rispetto alla classe dimensionale**

Come evidenziato nel prospetto 3.5, i risultati denotano uno scostamento significativo solo per le matricole aziendali con meno di 10 dipendenti. Più specificamente lo scostamento è del 18,23% per le matricole aziendali con meno di 6 dipendenti, del 12,36% per quelle da 6 a 9 dipendenti. Queste matricole rappresentano, nel complesso, il 61,7% del totale della popolazione d'interesse e coprono il 26,8% degli importi Anf stimati.

Per l'insieme delle matricole con più di 10 dipendenti (38,29% del totale matricole e 73,2% del totale degli importi Anf), la stima da 770 si discosta dal dato Dm10 non oltre il 9,55%, con performance migliori al crescere delle dimensioni aziendali: per le matricole sopra i 20 dipendenti, (18,1% del totale matricole e 56,1% del totale Anf stimati), lo scostamento diventa inferiore 6,71%, per quelle sopra i 100 dipendenti (2,6% del totale matricole e 28,1% del totale Anf stimati) lo scostamento scende ancora, attestandosi entro il 4,20%.

**Prospetto 3.5 - Confronto stime Anf da archivio 770 Inps con dati Dm10 per classe dimensionale (dati in euro)**

Classe dimensionale	Matricole aziendali (a)		Importi			
	n	%	770Inps	Dm10	Differenza percentuale	% (a)
solo cod 0035	513.294	100,00	2.336.090.143	2.151.865.364	8,56	100,00
- da 0 a 5	223.780	43,60	379.621.220	321.085.565	18,23	16,25
- da 6 a 9	92.962	18,11	246.215.091	219.132.073	12,36	10,54
- da 10 a 19	103.772	20,22	399.341.087	364.515.055	9,55	17,09
- da 20 a 99	79.199	15,43	653.665.838	612.573.787	6,71	27,98
- da 100 a 249	9.458	1,84	243.442.380	233.630.421	4,20	10,42
- oltre 250	4.123	0,80	413.804.527	400.928.463	3,21	17,71
tutti i codici	515.092		2.338.274.005	2.370.058.400	-1,34	

(a) statistiche calcolate solo su archivio 770

<sup>22</sup>Il dato di bilancio Inps da considerare è quello presente nel volume *Gestione Prestazioni Temporanee ai lavoratori dipendenti*, allegato n. 3 (tabella 16a), al Bilancio Consuntivo dell'Inps, Anno 2002.

Per completezza d'informazione, l'ultima riga del prospetto 3.5 riporta il risultato del confronto tra stime da 770 e dati da Dm10 nel caso in cui, oltre lo 0035, vengano presi in considerazione tutti gli altri codici Anf del Dm10. Come era prevedibile attendersi, dato che il modello 770 rileva solo prestazioni correnti (relativamente agli Anf), il dato stimato da 770 risulta in aggregato inferiore (del 1,34%) a quello desumibile da Dm10. I codici 'altri' rispetto allo 0035 descrivono infatti erogazioni Anf registrate nel Dm10 ma non nel 770, e quindi non confrontabili tra loro<sup>23</sup>. Ciò conferma pertanto la correttezza della scelta effettuata, e cioè realizzare (e confrontare) le stime con esclusivo riferimento agli ammontari legati al codice 0035. Quantitativamente questo è ulteriormente confermato dall'analisi della variazione della stima Anf calcolata sul solo codice 0035 o su tutti i codici. Le stime da archivio 770, se calcolate rispetto a tutti i codici Anf, subiscono un incremento pari a solo lo 0,09%, interamente imputabile alle ulteriori 1.798 matricole comuni che il filtro 'tutti i codici' permette di estrarre. La stima da Dm10 sale invece del 10,13%, poiché l'effetto incrementativo è quasi esclusivamente imputabile al peso che i codici 'altri' hanno in questo archivio.

### 3.2.2 Verifica delle stime rispetto alla classificazione Ateco

Quanto invece alla verifica della precisione delle stime rispetto alla classificazione Ateco, i risultati sono presentati nel prospetto 3.6. Le sezioni Ateco per le quali le stime da 770 si discostano da quelle da Dm10 per meno del 10% coprono il 71,4% del totale Anf erogati (stimati su 770).

**Prospetto 3.6 - Confronto stime Anf da archivio 770 Inps con dati Dm10 per sezione Ateco**

SEZIONI ATECO	Stime Anf erogati				
	770Inps		Dm10		Diff %
	Importi	% sul totale	Importi	% sul totale	
A Agricoltura, caccia e silvicoltura	3.986.293	0,2	3.725.711	0,2	7,0%
B Pesca, piscicoltura e servizi connessi	21.367.204	0,9	17.617.294	0,8	21,3%
C Estrazione di minerali	13.472.796	0,6	12.857.585	0,6	4,8%
D Attività manifatturiere	867.904.623	37,2	814.213.277	37,8	6,6%
E Produzione e distribuzione di energia elettrica, gas e acqua	16.254.703	0,7	14.460.579	0,7	12,4%
F Costruzioni	427.807.131	18,3	386.756.049	18,0	10,6%
G Commercio all'ingrosso e al dettaglio; riparazione di autoveicoli, motocicli e di beni personali e per la casa	299.365.416	12,8	273.181.386	12,7	9,6%
H Alberghi e ristoranti	105.733.945	4,5	100.434.110	4,7	5,3%
I Trasporti, magazzinaggio e comunicazioni	191.035.209	8,2	173.449.711	8,1	10,1%
J Attività finanziarie	19.343.263	0,8	16.420.455	0,8	17,8%
K Attività immobiliari, noleggio, informatica, ricerca, servizi alle imprese	240.447.272	10,3	220.858.644	10,3	8,9%
L Amministrazione pubblica	37.253	0,0	93.581	0,0	-60,2%
M Istruzione	8.678.602	0,4	7.605.776	0,4	14,1%
N Sanità e assistenza sociale	52.999.638	2,3	48.598.743	2,3	9,1%
O Altri servizi pubblici, sociali e personali	67.059.503	2,9	61.130.820	2,8	9,7%
Q Organizzazioni ed organismi extraterritoriali	93.264	0,0	77.302	0,0	20,6%
X Non ripartibili	504.030	0,0	384.341	0,0	31,1%
<b>Totale</b>	<b>2.336.090.143</b>	<b>100,0</b>	<b>2.151.865.364</b>	<b>100,0</b>	<b>8,6%</b>

Nel prospetto 3.7, il raffronto in base alla classificazione Ateco tra le stime da archivio 770 e i dati da Dm10 viene ulteriormente disaggregato per sottosezione e divisione. A livello di sottosezione gli scostamenti tra 770 e Dm10 si mantengono entro il limite del 10% per le matricole aziendali i cui importi Anf pesano complessivamente il 71,3% del totale, con un lievissimo decremento rispetto a quanto avveniva a livello di sezione. Passando a livello di divisione rimangono all'interno di margini di tollerabilità del 10% matricole aziendali che coprono il 63,2% del totale degli Anf sul totale stimato.

<sup>23</sup> Cfr nota 20.

**Prospetto 3.7 - Confronto stime Anf da archivio 770 con dati Dm10 per sezione e sottosezione Ateco**

Sezione	Sotto sezione	Divisione	770Inps		Dm10		Differenza %
			Importi	% sul totale	Importi	% sul totale	
A		01	3.260.314	0,1	3.033.530	0,1	7,5
		02	725.978,30	0,0	692.181	0,0	4,9
B			21.367.204	0,9	17.617.294	0,8	21,3
C	A	10	553.694	0,0	723.127	0,0	-23,4
		11	366.182	0,0	340.521	0,0	7,5
	B	13	190.036	0,0	187.607	0,0	1,3
		14	12.362.884	0,5	11.606.330	0,5	6,5
	A	15	78.329.108	3,4	73.613.754	3,4	6,4
		16	480.515	0,0	673.868	0,0	-28,7
	B	17	38.800.941	1,7	37.779.016	1,8	2,7
		18	32.196.662	1,4	29.703.185	1,4	8,4
	C	19	43.448.090	1,9	40.453.752	1,9	7,4
	D	20	41.412.665	1,8	37.886.382	1,8	9,3
	E	21	15.892.383	0,7	15.270.367	0,7	4,1
		22	18.008.504	0,8	16.362.561	0,8	10,1
	F	23	4.217.737	0,2	3.984.032	0,2	5,9
	G	24	24.739.126	1,1	23.382.203	1,1	5,8
	H	25	42.634.346	1,8	40.125.498	1,9	6,3
	I	26	76.394.485	3,3	71.462.356	3,3	6,9
D	J	27	34.372.309	1,5	33.125.422	1,5	3,8
		28	147.610.967	6,3	138.211.974	6,4	6,8
	K	29	86.802.827	3,7	81.502.871	3,8	6,5
		30	1.212.499	0,1	1.113.165	0,1	8,9
	L	31	29.914.176	1,3	29.735.634	1,4	0,6
		32	11.851.940	0,5	11.269.776	0,5	5,2
		33	11.384.533	0,5	9.611.586	0,4	18,4
	M	34	44.900.911	1,9	42.098.229	2,0	6,7
		35	23.859.083	1,0	20.475.910	1,0	16,5
	N	36	54.719.130	2,3	51.954.341	2,4	5,3
E		37	4.721.686	0,2	4.417.396	0,2	6,9
		40	13.026.362	0,6	11.666.949	0,5	11,7
F		41	3.228.341	0,1	2.793.629	0,1	15,6
		45	427.807.131	18,3	386.756.049	18,0	10,6
G		50	75.588.504	3,2	68.422.720	3,2	10,5
		51	103.283.250	4,4	94.194.525	4,4	9,6
H		52	120.493.662	5,2	110.564.141	5,1	9,0
		55	105.733.945	4,5	100.434.110	4,7	5,3
		60	125.094.310	5,4	111.388.662	5,2	12,3
		61	3.515.252	0,2	3.853.233	0,2	-8,8
I		62	692.891	0,0	626.108	0,0	10,7
		63	53.605.494	2,3	48.220.178	2,2	11,2
		64	8.127.263	0,3	9.361.530	0,4	-13,2
		65	14.045.740	0,6	11.663.532	0,5	20,4
J		66	1.701.342	0,1	1.554.719	0,1	9,4
		67	3.596.180	0,2	3.202.204	0,1	12,3
		70	23.042.806	1,0	19.653.155	0,9	17,2
		71	3.160.726	0,1	2.688.466	0,1	17,6
K		72	14.246.301	0,6	12.706.770	0,6	12,1
		73	1.017.847	0,0	928.277	0,0	9,6
		74	198.979.593	8,5	184.881.975	8,6	7,6
	L	75	37.253	0,0	93.581	0,0	-60,2
M		80	8.678.602	0,4	7.605.776	0,4	14,1
N		85	52.999.638	2,3	48.598.743	2,3	9,1

**Prospetto 3.7 (segue): Confronto stime Anf da archivio 770 con dati Dm10 per sezione e sottosezione Ateco**

Sezione	Sotto sezione	Divisione	770Inps		Dm10		Differenza %
			Importi	% sul totale	Importi	% sul totale	
O		90	32.926.224	1,4	31.446.491	1,5	4,7
		91	3.282.280	0,1	2.802.780	0,1	17,1
		92	10.462.478	0,4	8.685.086	0,4	20,5
		93	20.388.522	0,9	18.196.464	0,8	12,0
Q		99	93.264	0,0	77.302	0,0	20,6
X			504.030	0,0	384.341	0,0	31,1
<b>Totale</b>			<b>2.336.090.143</b>	<b>100,0</b>	<b>2.151.865.364</b>	<b>100,0</b>	<b>8,6</b>

Oltre ai controlli appena descritti, il grado di similarità delle due distribuzioni è stato analizzato nel prospetto 3.8, mediante il confronto delle medie (calcolato sul totale degli importi erogati per singola matricola aziendale). Le medie da archivio 770 si discostano da quelle da archivio Dm10 in misura pari o inferiore al 10% per matricole cui corrisponde il 70,7% del totale Anf erogati.

**Prospetto 3.8 - Stime da archivio 770 e da archivio Dm10: valori medi e sugli indicatori di variabilità**

Sezione Ateco	Incidenza % su archivio 770		Media		Differenza %
	Sul totale matricole	Sul totale importi	770Inps	Dm10	
A Agricoltura, caccia e silvicoltura	0,35	0,17	2.213,38	2.068,69	6,99
B Pesca, piscicoltura e servizi connessi	0,51	0,91	8.177,27	6.742,17	21,29
C Estrazione di minerali	0,45	0,58	5.819,78	5.554,03	4,78
D Attività manifatturiere	30,21	37,15	5.597,90	5.251,60	6,59
E Produzione e distribuzione di energia elettrica, gas e acqua	0,20	0,70	15.720,22	13.985,09	12,41
F Costruzioni	19,67	18,31	4.237,90	3.831,24	10,61
G Commercio all'ingrosso e al dettaglio; riparazione di autoveicoli, motocicli e di beni personali e per la casa	20,45	12,81	2.851,34	2.601,95	9,58
H Alberghi e ristoranti	6,14	4,53	3.354,93	3.186,77	5,28
I Trasporti, magazzinaggio e comunicazioni	5,57	8,18	6.682,36	6.067,22	10,14
J Attività finanziarie	0,89	0,83	4.216,97	3.579,78	17,80
K Attività immobiliari, noleggio, informatica, ricerca, servizi alle imprese	10,15	10,29	4.614,05	4.238,15	8,87
L Amministrazione pubblica	0,00	0,00	1.773,95	4.456,24	60,19
M Istruzione	0,57	0,37	2.971,11	2.603,83	14,11
N Sanità e assistenza sociale	2,01	2,27	5.133,14	4.706,90	9,06
O Altri servizi pubblici, sociali e personali	2,79	2,87	4.681,62	4.267,72	9,70
Q Organizzazioni ed organismi extraterritoriali	0,01	0,00	3.216,00	2.665,59	20,65
X Non ripartibili	0,02	0,02	3.937,73	3.003	31

**3.3 Le procedure di stima degli Anf dagli archivi M\_770: alcune considerazioni conclusive**

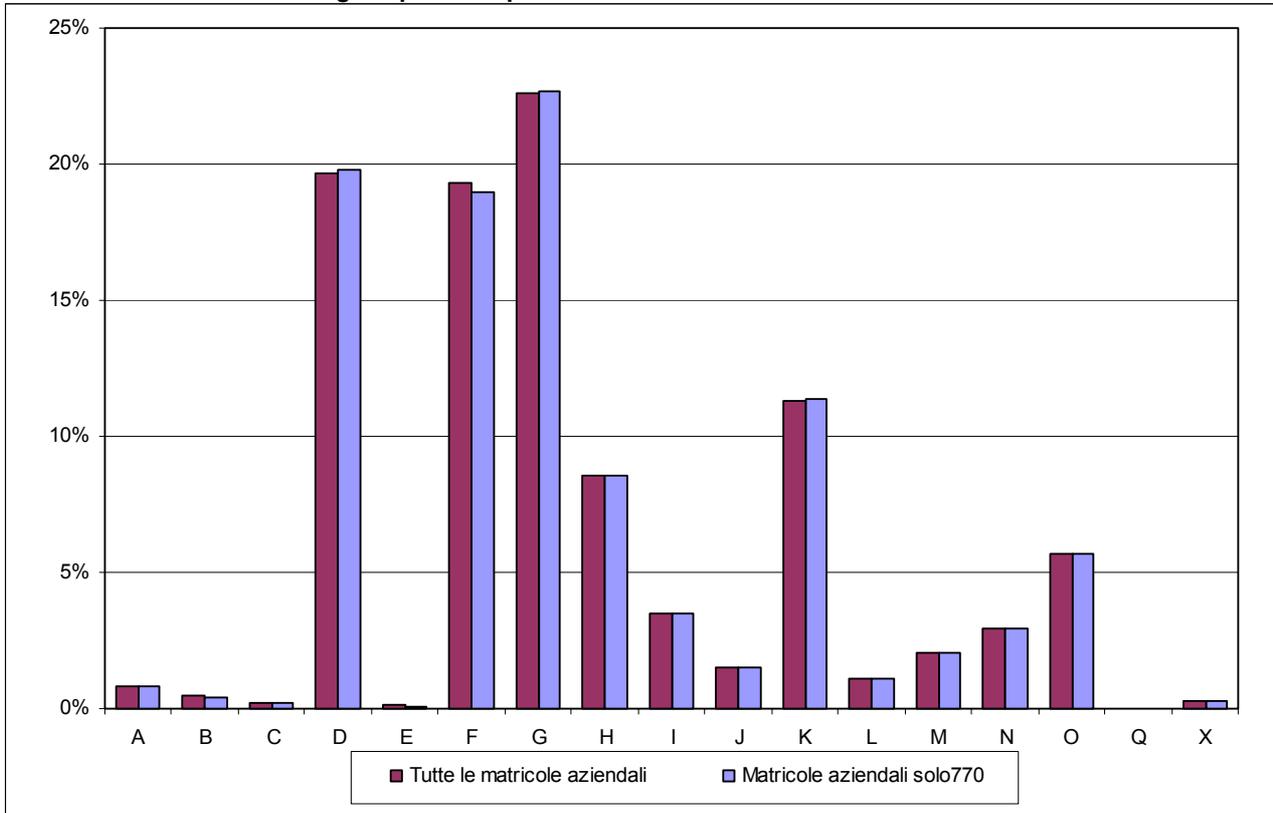
La sperimentazione ha dimostrato una buona approssimazione delle stime Anf da archivio 770 rispetto al dato di benchmark del bilancio consuntivo dell'Inps, come pure rispetto alle stime elaborate a partire dal modello Dm10, anche quando disaggregate per dimensione aziendale o per classificazione Ateco (fino al livello divisione).

Questo, unito alle analisi condotte sulla distribuzione delle 19.085 matricole *solo770*, porta a concludere, come già accennato, che la popolazione sulla quale è opportuno calcolare la stima definitiva degli Anf erogati è tutta quella del 770 (532.379), fin qui analizzata separatamente come somma delle matricole aziendali comuni (513.294) e delle matricole aziendali *solo770* (19.085). A sostegno di questa conclusione due considerazioni:

a) la procedura di stima fornisce risultati statisticamente robusti nel 71,4% delle matricole 770, ed è quindi assolutamente plausibile che lo stesso avvenga per le 19.085 matricole *solo770*, la cui distribuzione non presenta, tra l'altro, macroscopiche peculiarità rispetto a quella delle matricole comuni;

b) come dimostra il grafico 3.1, la distribuzione per classe Ateco degli importi stimati per le sole matricole comuni non risulta difforme da quella per il totale delle matricole 770. Pertanto l'inclusione delle 19.085 matricole *solo770* non produce effetti distorsivi sulla distribuzione finale nemmeno in base alla classificazione Ateco degli importi Anf stimati.

**Grafico 3.1 - Distribuzione degli importi Anf per sezione Ateco**





## Capitolo 4 - Aspetti metodologici della stima degli assegni per il nucleo familiare

Questo capitolo e il successivo sono dedicati all'approfondimento dei dettagli metodologici relativi al trattamento dei dati amministrativi dei modelli di dichiarazione fiscale 770, anno di dichiarazione 2003, per la produzione di informazioni statistiche sugli assegni al nucleo familiare (Anf)<sup>24</sup>. L'attenzione è posta sulle problematiche relative al trattamento degli errori non campionari e alle tecniche per la validazione e il controllo di qualità dei risultati.

I sostituti di imposta, comprese le Amministrazioni dello Stato, sono tenuti a comunicare in via telematica all'Agenzia delle Entrate i dati fiscali relativi alle ritenute operate nell'anno precedente nonché gli altri dati contributivi e assicurativi richiesti. Il modello di dichiarazione 770 contiene i dati relativi alle certificazioni rilasciate ai soggetti cui sono stati corrisposti redditi di lavoro dipendente ed assimilati, indennità di fine rapporto, prestazioni in forma di capitale erogate da fondi pensione, redditi da lavoro autonomo, provvigioni e redditi diversi, nonché i dati contributivi, previdenziali e assicurativi e quelli relativi all'assistenza fiscale prestata nell'anno di riferimento per il periodo d'imposta precedente.

Ai fini della produzione di statistiche sugli Anf, la sezione di interesse è quella relativa ai dati contributivi, previdenziali e assicurativi di competenza Inps, integrata da informazioni di tipo anagrafico riguardanti ciascun lavoratore dipendente<sup>25</sup>. Relativamente alla provenienza dell'archivio utilizzato per la produzione delle statistiche, occorre precisare che in Istat sono attualmente presenti due versioni della sezione 'dati previdenziali e assistenziali' dei modelli 770, la prima rilasciata dalla Sogei e la seconda dall'Inps<sup>26</sup>. Uno studio di compatibilità tra le due versioni dell'archivio ha evidenziato la superiorità della fonte Inps, ai fini del progetto Tmnp, sia in termini di possibilità di sfruttamento di un archivio più popolato, e quindi di minimizzazione della perdita di potenziali percettori, che di utilizzo di un dato già sottoposto a controlli di qualità<sup>27</sup>.

### 4.1 Campo di osservazione e unità d'analisi

L'universo di riferimento per la produzione delle statistiche sugli Anf è un sottoinsieme dell'universo dei soggetti che compilano i modelli 770. L'archivio rilevante ai fini statistici è infatti rappresentato da soggetti o enti privati che nell'anno d'imposta hanno corrisposto Anf per conto dell'Inps. Sono escluse dal campo di indagine le Amministrazioni dello Stato, comprese quelle con ordinamento autonomo, e tutti gli enti e soggetti che non rientrano nella sfera del privato. La scelta di limitare l'attenzione al solo comparto privato è motivata dal fatto che l'archivio amministrativo dei modelli 770 non rileva informazione utile alla produzione di statistiche Anf per il comparto pubblico. Gli Anf dei lavoratori dipendenti del settore pubblico non sono infatti a carico dell'ente previdenziale (Inpdap) bensì dell'amministrazione per cui prestano attività lavorativa. Trattandosi di una prestazione che viene erogata direttamente dall'amministrazione, la sezione dei dati previdenziali e assistenziali Inpdap non contiene alcuna informazione in merito.

---

<sup>24</sup> L'analisi è stata condotta sull'ultimo archivio disponibile in ordine temporale. Un'analisi longitudinale sugli archivi degli anni 2001, 2002 e 2003 non ha rilevato incongruenze fondamentali nei dati nel corso del tempo. È stato invece rilevato nel corso degli anni un diverso livello di popolamento degli archivi e una tendenza al miglioramento della qualità dell'informazione.

<sup>25</sup> L'archivio dei dati anagrafici è costruito in Istat come integrazione di informazioni provenienti da diverse fonti amministrative, allo stato attuale Sogei, Inps e Inail. Questo archivio, quando risulterà completo, consentirà di disporre di un'anagrafica comune a tutte le prestazioni monetarie non pensionistiche.

<sup>26</sup> Gli archivi disponibili sono quelli relativi agli anni di dichiarazione 2001, 2002, 2003 e si riferiscono rispettivamente alle posizioni amministrative degli anni 2000, 2001, 2002.

<sup>27</sup> La sezione dei dati previdenziali e assistenziali di competenza Inps è sottoposta a scrupolosi controlli amministrativi volti, da un lato a stabilire una compatibilità di contenuto tra le voci rilevanti ai fini della dichiarazione e dall'altro a garantire una completa copertura dell'universo dei soggetti tenuti alla presentazione delle dichiarazioni. Cfr. Cap.4, Appendice A, pag. 55.

L'unità di analisi è rappresentata dal lavoratore dipendente, la cui dichiarazione 770 compete al sostituto d'imposta. È opportuno precisare che il beneficiario di Anf non è in realtà il lavoratore dipendente ma il nucleo familiare cui fa capo. Le informazioni desumibili dalla sezione dei dati previdenziali e assistenziali consentono di risalire al beneficiario come qui inteso (nucleo familiare), solo attraverso la coordinata 'numero di componenti familiari', tuttavia, a parte le caratteristiche socio-demografiche del lavoratore dipendente, quelle dei suoi familiari non sono note<sup>28</sup>. Ne segue pertanto, che pur essendo la prestazione erogata a favore del nucleo familiare di appartenenza, tutte le statistiche sugli Anf saranno elaborate rispetto al lavoratore dipendente e si riferiranno ai suoi caratteri socio-demografici.

#### 4.2 Fonti di distorsione: non-sampling errors

L'archivio amministrativo dei modelli 770 copre l'intero universo dei lavoratori dipendenti del settore privato. Nello specifico, la sezione dei dati previdenziali ed assistenziali (parte C, sezioni 1, 2 e 3) consente una rappresentazione completa dell'universo dei percettori di assegni anticipati dal datore di lavoro e posti a conguaglio con i contributi dovuti all'Inps. Trattandosi di un'analisi sull'intera popolazione, sono superati tutti i problemi di distorsione e perdita di efficienza delle stime derivanti da errori campionari mentre rimangono tutti quelli generati da errori non campionari. L'inaccuratezza statistica delle stime potrebbe allora essere originata dalle seguenti fonti di errore:

- *problemi di copertura*, connessi: 1) al popolamento degli archivi con riferimento alla data di rilascio ai fini statistici e 2) all'impossibilità di individuare in maniera certa l'universo dei titolari di Anf. In entrambi i casi si generano effetti distorsivi e un aumento della variabilità delle stime sia del numero di percettori Anf che degli importi erogati.

Sul popolamento degli archivi di cui al punto 1) si discuterà in dettaglio nel paragrafo 3. Relativamente all'individuazione dei titolari di Anf di cui al punto 2), occorre notare che i datori di lavoro sono tenuti alla compilazione del riquadro relativo agli Anf solo se questi sono corrisposti al lavoratore anche nel mese di dicembre. Ciò implica una perdita di informazioni relative a tutti coloro che, per diverse motivazioni, hanno cessato il diritto alla corresponsione degli assegni nel corso dell'anno. La procedura implementata per la stima degli Anf tiene esplicitamente conto di questo problema e prevede il recupero dei percettori potenziali tramite l'integrazione longitudinale degli archivi. Ciò consente di limitare il problema di sottocopertura dell'universo dei titolari di assegni<sup>29</sup>;

- *mancata risposta*. La letteratura tradizionalmente contempla due tipi di mancata risposta, mancata risposta totale (*unit non response*) e mancata risposta parziale (*item non response*). Il rischio di mancata risposta totale è molto contenuto, in quanto, per le imprese e/o soggetti tenuti alla presentazione della dichiarazione, la sezione dei dati previdenziali e assistenziali di competenza Inps rappresenta un adempimento di natura amministrativa<sup>30</sup>. Inoltre, nel contesto dei modelli 770, la mancata risposta totale può essere originata sia a livello di matricola aziendale che di singolo dipendente. A quest'ultimo riguardo, essendo la singola dichiarazione composta dai fogli di tutti i dipendenti, la probabilità di mancata risposta totale a livello di dipendente è molto contenuta, in quanto, salvo il caso di imprese di grandi dimensioni, le trasmissioni avvengono mediante un unico invio telematico. In ogni caso, mentre a livello di matricola aziendale è possibile valutare l'impatto di questo tipo di errore sulle stime, attraverso l'aggancio con le matricole aziendali presenti nell'archivio Dm10, a livello di dipendente non è possibile in alcun modo quantificare l'incidenza della mancata risposta totale sul risultato finale.

Relativamente alla mancata risposta parziale, invece, sono state predisposte opportune tecniche che consentono di recuperare l'informazione incompleta tramite l'impiego di approcci di natura deterministico-stocastico, i cui aspetti metodologici sono discussi nel paragrafo 2.3 del capitolo 2. È

---

<sup>28</sup> Ai fini della determinazione degli ANF, le variabili rilevanti sono rappresentate da un set di coordinate, campi 25, 26, 27 del mod. 770 che rinviano a tabelle definite di anno in anno da circolari Inps. Tali coordinate contengono rispettivamente il riferimento alle tabelle (campo 25), numero di componenti familiari (campo 26) e classi di reddito (campo 27) e sono distinte a seconda si riferiscano a nuclei con figli (tab. 11, 12, 13, 14, 15, 16, 17, 18, 19) o senza figli (20A, 20B, 21A, 21B, 21C, 21D).

<sup>29</sup> Cfr. Cap.2, par.2.4, pag.19.

<sup>30</sup> Il modello 770 ha finalità esclusiva di 'comunicazione' e eventuali ritardi nella presentazione non sono sanzionabili da un punto di vista amministrativo. Tuttavia, la sezione rilevante ai fini Inps è soggetta a scrupolosi controlli amministrativi aventi l'obiettivo di limitare il problema di perdita delle dichiarazioni.

necessario precisare che le tecniche di recupero dell'informazione incompleta vengono applicate solo in presenza di informazione minima sulle variabili di interesse<sup>31</sup>.

Da un punto di vista statistico, mentre la mancata risposta totale genera sia effetti distorsivi sulle stime sia un aumento della variabilità delle stesse, la scelta della componente stocastica per l'imputazione parziale tende a limitare gli effetti di omogeneizzazione e deformazione della variabilità indotti da metodi di natura deterministica;

- *errori di misurazione*, costituiti dalla differenza fra il vero valore del carattere da misurare e il valore effettivamente osservato dall'indagine. Nella stima degli Anf, gli errori di misurazione sono introdotti sia in fase di compilazione del modello<sup>32</sup> (es. incomprensione delle istruzioni) sia nello svolgimento degli stadi di elaborazione successivi alla raccolta del dato. Analisi di coerenza interna e tra le variabili rilevanti ai fini della determinazione degli Anf, sono state condotte in fase preliminare e contestualmente sono state effettuate le necessarie correzioni. Tra le potenziali fonti di errore commesse in fase di elaborazione delle stime, è opportuno segnalare quelle derivanti: 1) dall'assunzione cruciale del periodo retributivo quale periodo di percepimento degli Anf<sup>33</sup> e 2) dall'invarianza delle coordinate Anf nel corso del periodo di percepimento<sup>34</sup>. In entrambi i casi non vi è alcuna ripercussione sul numero di percettori stimati mentre non è possibile valutare l'entità della sovrastima o sottostima sugli importi Anf<sup>35</sup>. È interessante rilevare che l'incidenza degli errori di misurazione è più alta rispetto a quella che si otterrebbe da campione, in quanto coinvolge un maggior numero di unità.

### 4.3 Problemi di copertura dell'archivio 770

In questa sede ci si occupa di fornire alcuni dettagli tecnici sulle conseguenze derivanti dalla non completa copertura dell'universo delle matricole aziendali erogatrici di Anf.

Per l'anno di dichiarazione 2003, l'integrazione dei due archivi amministrativi ha evidenziato un discreto numero di matricole aziendali 770 non presenti nel Dm10 e viceversa. Questo fenomeno potrebbe essere associato a tre problemi:

- 1) non coincidenza dell'universo delle matricole aziendali dichiaranti, per effetto di fattori normativi (es. è possibile che non tutte le aziende che devono adempiere all'obbligo di presentazione dei modelli 770 siano tenute alla presentazione del Dm10<sup>36</sup>);
- 2) problemi di copertura dell'universo delle matricole aziendali. Mentre per il Dm10 è stato condotto uno studio sul grado di popolamento degli archivi a diverse scadenze di rilascio del dato, non esiste ad oggi documentazione formale sul livello di popolamento dell'archivio dei modelli 770<sup>37</sup>;
- 3) potenziale sottocopertura dell'universo dei beneficiari per effetto della perdita dei percettori infra-annuali<sup>38</sup>.

---

<sup>31</sup> Tali informazioni consistono in requisiti minimi per presupporre che si è in presenza di un titolare di ANF. La condizione di requisito minimo si sostanzia nella presenza di almeno una coordinata ANF con valori significativi a livello normativo.

<sup>32</sup> Uno degli errori ricorrenti nei casi di coordinate non corrette è l'indicazione di un numero di componenti familiari corrispondenti a valori ANF nulli. Vi è ragione di credere che, in questi casi, nel calcolo della dimensione del nucleo familiare sia stato escluso il dipendente titolare dell'assegno.

<sup>33</sup> Cfr. Cap.2, par.2.5, pag.41

<sup>11</sup> Nel caso di dipendenti (identificati tramite lo stesso codice fiscale) contemporaneamente presenti in due anni consecutivi, si sfrutta l'informazione sulle coordinate dei due anni per introdurre un elemento di variabilità tra coordinate del primo e secondo semestre dell'anno corrente. Nell'ipotesi in cui le coordinate dei due anni consecutivi non coincidano, per il primo semestre dell'anno in corso vengono attribuite le coordinate dell'anno precedente, mentre per il secondo semestre, le coordinate valide sono quelle dell'anno corrente. Ne segue, che l'invarianza delle coordinate si assume per i mesi interni ai due sottoperiodi ma non necessariamente tra un semestre e l'altro.

<sup>35</sup> Un'analisi sulle variazioni tra classi di reddito, tabelle e numero di componenti familiari ha evidenziato che le transizioni più significative riguardano le classi di reddito. In generale, tra l'anno di riferimento 2001 e il 2002 vi è una tendenza allo slittamento verso classi di reddito più elevate.

<sup>36</sup> In generale, le aziende che compilano il modello Dm10 sono tenute anche alla presentazione del modello 770, non vale tuttavia l'opposto. Una eccezione a questo caso è rappresentata dai lavoratori dipendenti, le cui imprese devono compilare entrambi i modelli.

<sup>37</sup> Sulla base di valutazioni sui flussi delle dichiarazioni, l'ente incaricato della raccolta del dato 770 (Sogei), ritiene che l'archivio raggiunga un adeguato livello di maturità tra il mese di marzo e aprile di due anni successivi a quello della scadenza di presentazione della dichiarazione. Ad esempio, per l'anno di dichiarazione 2003, la scadenza di presentazione è giugno 2003 e la data ritenuta adeguata da Sogei per la fornitura dell'archivio ai fini statistici è marzo-aprile 2005. La versione Inps dei dati previdenziali e assistenziali è inoltre soggetta a scrupolosi controlli amministrativi che prevedono il recupero di dichiarazioni pervenute anche in date successive rispetto alla scadenza normativa.

Per valutare la misura dell'impatto dei tre problemi, le due fonti citate sono state confrontate con l'archivio dei modelli 770 rilasciato dalla Sogei ed in particolare con la sezione dei dati assistenziali e previdenziali<sup>39</sup>. Come evidenziato dal prospetto 4.1, la comparazione del sotto universo di matricole aziendali presenti solo nell'archivio Inps dei 770, con quelle di provenienza Sogei ha evidenziato un'integrazione quasi completa delle due fonti. Delle 19.085 matricole aziendali presenti soltanto nell'archivio Inps dei modelli 770, il 74 per cento è presente nell'archivio di fonte Sogei con coordinate Anf significativamente diverse da zero. Diversamente, le 71.441 matricole aziendali presenti solo nell'archivio dei Dm10 trovano riscontro nell'archivio Sogei in misura molto più contenuta, solo il 3 per cento presenta coordinate non nulle, mentre il 71 per cento, pur essendo presente nella sezione dei dati previdenziali e assistenziali, non presenta coordinate Anf significative.

**Prospetto 4.1 - Distribuzione per tipo di appartenenza Sogei delle matricole solo 770 INPS e solo Dm10**

	matricole aziendali 770 presenti in Sogei (sez. G)		matricole aziendali 770 non presenti in Sogei (sez. G)	Totale
	coord Anf not null	coord Anf null		
matricole aziendali solo 770 INPS(*)	14.203 -74,4	2.974 -15,6	1.908 -10	19.085 -100
matricole aziendali solo Dm10	2.094 -2,9	50.981 -71,4	18.366 -25,7	71.441 -100

Note: (\*) include sono le dichiarazioni rilevate in corso d'anno e non quelle recuperate dall'anno precedente.

Per le 18.366 matricole aziendali presenti solo nell'archivio Dm10, che non trovano riscontro nell'archivio dei modelli 770 rilasciato dalla Sogei è plausibile ipotizzare una perdita di unità per effetto di problemi di copertura. Diversamente, per le restanti 50.981 aziende, che pur essendo presenti nella sezione previdenziale dell'archivio rilasciato dalla Sogei presentano coordinate Anf nulle, non è possibile quantificare quante di esse si perdono per effetto della diversa data di rilascio del dato e quante siano invece imputabili alla perdita di percettori infra-annuali. In tal senso, l'archivio Dm10 consente di rilevare mese per mese e per ciascuna impresa, gli importi Anf anticipati e posti a conguaglio con l'Inps<sup>40</sup>.

Il prospetto 4.2, riporta la distribuzione delle matricole aziendali che richiedono il conguaglio per Anf distinte a seconda che la domanda sia presentata o meno anche al mese di dicembre.

**Prospetto 4.2: Distribuzione delle matricole presenti nel Dm10 e non nel 770 fonte Inps per mese 'dicembre' conguagliato**

Periodo retributivo di riferimento	matricole aziendali presenti in Dm10 € a 770 Sogei (*)	matricole aziendali presenti in Dm10€ a 770 Sogei
Inclusive mese dicembre	6.879 (13,0)	5.325 (29,0)
Senza mese dicembre	46.196 (87,0)	13.041 (71,0)
<b>Totale</b>	<b>53.075</b> <b>(100,0)</b>	<b>18.366</b> <b>(100,0)</b>

Note: (\*) include sono le dichiarazioni rilevate in corso d'anno e non quelle recuperate dall'anno precedente.

Nell'archivio Dm10, il 13 per cento delle matricole aziendali comuni all'archivio dei modelli 770 rilasciati dalla Sogei (sezione G) presenta richieste di conguaglio per Anf corrisposti anche nel mese di dicembre

<sup>38</sup> In base alle regole di compilazione del mod. 770, le coordinate ANF devono essere fornite solo se il lavoratore dipendente percepisce gli assegni anche al mese di dicembre. Ciò implica la perdita di tutti i percettori infra-annuali.

<sup>39</sup> Si fa notare che la versione Inps dell'archivio modd. 770 include la sola sezione dei dati previdenziali e assistenziali di cui è competente l'Inps, mentre la versione Sogei contempla tutte le sezioni del modello di dichiarazione.

<sup>40</sup> È necessario chiarire che i dati da Dm10 sono disponibili solo aggregati a livello di matricola aziendale, e non disaggregati per singolo codice fiscale del lavoratore dipendente.

dell'anno di imposta. Tale quota sale al 29 per cento nel caso delle matricole aziendali presenti solo nell'archivio Dm10. L'ultima evidenza fornisce pertanto indizi a favore sia della presenza di potenziali imprese erogatrici di Anf che sfuggono alle stime da 770 (quale che sia la fonte di rilascio), in quanto relative a percettori in corso d'anno ma non al mese di dicembre, sia a favore di motivi più generali di natura normativa<sup>41</sup> o legati ad altri eventi (es. ritardi nella presentazione dei modelli 770 per effetto di eventi eccezionali) che ne giustificano l'assenza. La distribuzione delle matricole aziendali per stato di attività/inattività (prospetto 4.3) mostra che il 33 per cento circa delle matricole presenti solo nell'archivio dei modelli Dm10, è stata sottoposta nel corso dell'anno a procedure amministrative di cessazione, sospensione e/o riattivazione. È possibile che tali eventi abbiano determinato ritardi nella presentazione dei modelli di dichiarazione 770.

**Prospetto 4.3 - Distribuzione delle matricole solo Dm10 per stato di attività/inattività**

	Stato attività/inattività dell'impresa					Totale
	Attive	Cessate	Sospese	Riattivate	Non codificate	
Valori assoluti	47.278	699	6.159	16.849	472	71.441
Percentuale	(66,2)	(1,0)	(8,6)	(23,6)	(0,7)	(100,0)

L'evidenza dei dati impone pertanto due riflessioni. Da un lato, il fatto che la frazione di matricole aziendali potenziali erogatrici di Anf e rilevate solo dall'archivio dei Dm10 sia non trascurabile, sottolinea l'importanza di non sottovalutare il fenomeno della eventuale perdita di percettori in corso d'anno. Dall'altro, la constatazione che la percentuale di matricole aziendali non presenti né nell'archivio 770 di provenienza Inps, né in quello rilasciato dalla Sogei e che richiedono conguagli per Anf anche al mese di dicembre, non è trascurabile potrebbe fornire una duplice chiave di interpretazione: 1) mentre è possibile che vi siano problemi di sottocopertura dell'universo delle matricole aziendali nell'archivio 770 sia di provenienza Inps che Sogei (comprese quelle i cui ritardi nella presentazione delle dichiarazioni possono essere associati a eventi eccezionali); 2) è anche probabile che l'universo delle matricole aziendali Dm10 che richiede conguagli per Anf non coincida, per motivi normativi, con quello dei 770. Quest'ultima interpretazione giustificherebbe il motivo per cui l'universo delle matricole aziendali 770 di provenienza Inps è molto più simile a quello di fonte Sogei rispetto a quanto lo sia l'universo delle matricole aziendali Dm10.

#### 4.4 Unit e item non response

Non disponendo di un universo di riferimento alternativo è molto difficile quantificare la dimensione della mancata risposta totale. Tuttavia, poiché la sezione contenente i dati sulle coordinate Anf raccoglie informazioni rilevanti ai fini previdenziali ed è sottoposta a rigorosi controlli di natura amministrativa da parte dell'ente previdenziale, ci si attende che il rischio di mancata risposta totale sia molto contenuto.

Più rilevante in questo contesto è invece il trattamento della mancata risposta parziale associata ad una non completa o scorretta compilazione dei campi relativi alle coordinate Anf. Come anticipato in precedenza, prerequisito affinché un record affetto da informazione incompleta sia sottoposto a imputazione è che vi sia informazione minima per poter presupporre il percepimento di assegni, ovvero, almeno una coordinata deve essere significativa<sup>42</sup>. In questa sezione verranno discussi gli aspetti teorici e le implicazioni statistiche connesse all'utilizzo di tecniche d'imputazione di tipo donatore.

Così come la maggior parte delle tecniche di imputazione tradizionale, le procedure basate su *donatore* soffrono del vincolo di "ignorabilità" del processo generatore del dato mancante. Nel caso in cui l'ipotesi di

<sup>41</sup> Ad esempio, la normativa potrebbe prevedere per talune imprese l'obbligo di presentazione del modello Dm10 ma non del modello di dichiarazione 770.

<sup>42</sup> Per i dettagli tecnici sulla procedura utilizzata per l'imputazione del dato incompleto cfr. Cap.2, par.2.3, pag. 19.

“ignorabilità” di un processo non fosse vera, l'analisi basata sull'impiego dell'archivio ridotto sarebbe soggetta a distorsione<sup>43</sup>.

Formalmente, date due variabili  $X$  e  $Y$ , le quali possono essere o meno osservate, la probabilità di risposta  $R$ :

- 1) può dipendere da  $X$  ma non da  $Y$
- 2) può essere indipendente da  $X$  e  $Y$
- 3) può dipendere da  $X$  e  $Y$ .

Nel primo caso si dice che i dati mancanti sono di tipo *missing at random* (*MAR*), nel senso che i valori di  $Y$  osservati non sono necessariamente un sottocampione casuale dei valori campionati, bensì un campione casuale dei valori campionati all'interno di una sottoclasse definita dai valori di  $X$ . Nel secondo caso, i dati mancanti sono *missing at random* (*MAR*), mentre i dati osservati sono *observed at random* (*OAR*), o più semplicemente tutto il meccanismo di dati mancanti è di tipo *missing completely at random* (*MCAR*). Nel terzo caso infine, i dati incompleti e quelli osservati non sono né *MAR* né *OAR* e il meccanismo di dati mancanti è di tipo *non missing at random* (*NMAR*).

Più in generale, data la funzione di risposta  $f(R|Y_{obs}, Y_{mis}, \xi)$ , il meccanismo generatore dei dati mancanti può risultare tale che:

- a) se non dipende dalla componente inosservata  $Y_{mis}$ , ovvero

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|Y_{obs}, \xi)$$

allora il meccanismo generatore dei dati mancanti è di tipo *missing at random* (*MAR*);

- b) se non dipende né dalla componente inosservata  $Y_{mis}$ , né dalla componente osservata  $Y_{obs}$ , ovvero

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|\xi)$$

allora il meccanismo sottostante la generazione del dato mancante è di tipo *missing completely at random* (*MCAR*);

- c) se dipende sia dalla componente inosservata  $Y_{mis}$ , che dalla componente osservata  $Y_{obs}$ , ovvero

$$P(R|Y_{obs}, Y_{mis}, \xi) = P(R|Y_{obs}, Y_{mis}, \xi)$$

allora il meccanismo sottostante la generazione del dato mancante è di tipo *non missing at random* (*NMAR*).

Un meccanismo generatore del dato incompleto è definito “ignorabile” se i valori mancanti sono *missing at random*, nel senso che le unità osservate rappresentano un sottocampione casuale delle unità campionate. Se la probabilità che  $y_i$  sia osservata dipende dal valore di  $y_i$ , allora il meccanismo di dati mancanti non è ignorabile e l'analisi sul campione ridotto che non tenga in considerazione di questo fatto è soggetta a distorsione. L'assunzione di ignorabilità di un processo, benché criticabile in alcune circostanze, porta spesso a risultati migliori rispetto all'adozione di procedure *ad hoc* (ad esempio quelle basate sull'utilizzo della sola componente osservabile), in quanto consente di rimuovere tutta la distorsione generata dalla parte di non risposta spiegabile attraverso la componente osservata  $Y_{obs}$ . È utile tenere a mente che l'assunzione cruciale su cui si basano i metodi che ignorano il processo di mancata risposta, è che la propensione alla non risposta possa essere spiegata dai dati osservati e non che essa sia totalmente incorrelata con i dati mancanti. È per questo che procedure come quella utilizzata in questo contesto usualmente producono buoni risultati anche quando l'ignorabilità è sospettata. In generale, la plausibilità dell'ipotesi di ignorabilità dipende dal contenuto dei dati mancanti e dalla complessità del modello dei dati  $P(Y|\theta)$ ; se i dati osservati contengono informazione sufficiente per prevedere i dati mancanti e se il modello è sufficientemente elaborato da far uso di questa informazione, ci si aspetterà che la dipendenza residua di  $R$  da  $Y_{mis}$ , dopo aver condizionato ad  $Y_{obs}$ , sarà relativamente minore. In termini pratici, l'assunzione *MAR* implica che per ricostruire l'informazione mancante sia sufficiente transitare solo tramite l'informazione dei rispondenti. Tale ipotesi, seppure non sempre accettabile, è difendibile in questo contesto d'analisi in quanto è ragionevole assumere che il dato mancante sulle coordinate non sia generato da un sorta di meccanismo di autoselezione, derivando piuttosto da errori non direttamente legati alle caratteristiche del rispondente (es. incomprendimento delle regole di compilazione del 770 da parte del commercialista). Per valutare l'eventuale effetto distorsivo sulla distribuzione del dato finale sono state analizzate la distribuzione del dato non imputato (dataset dei potenziali donatori) e la distribuzione dopo l'imputazione del dato affetto da incompletezza o non

<sup>43</sup> Cfr. Little and Rubin (1987) sulle implicazioni statistiche della violazione delle ipotesi *MAR* e *MCAR*.

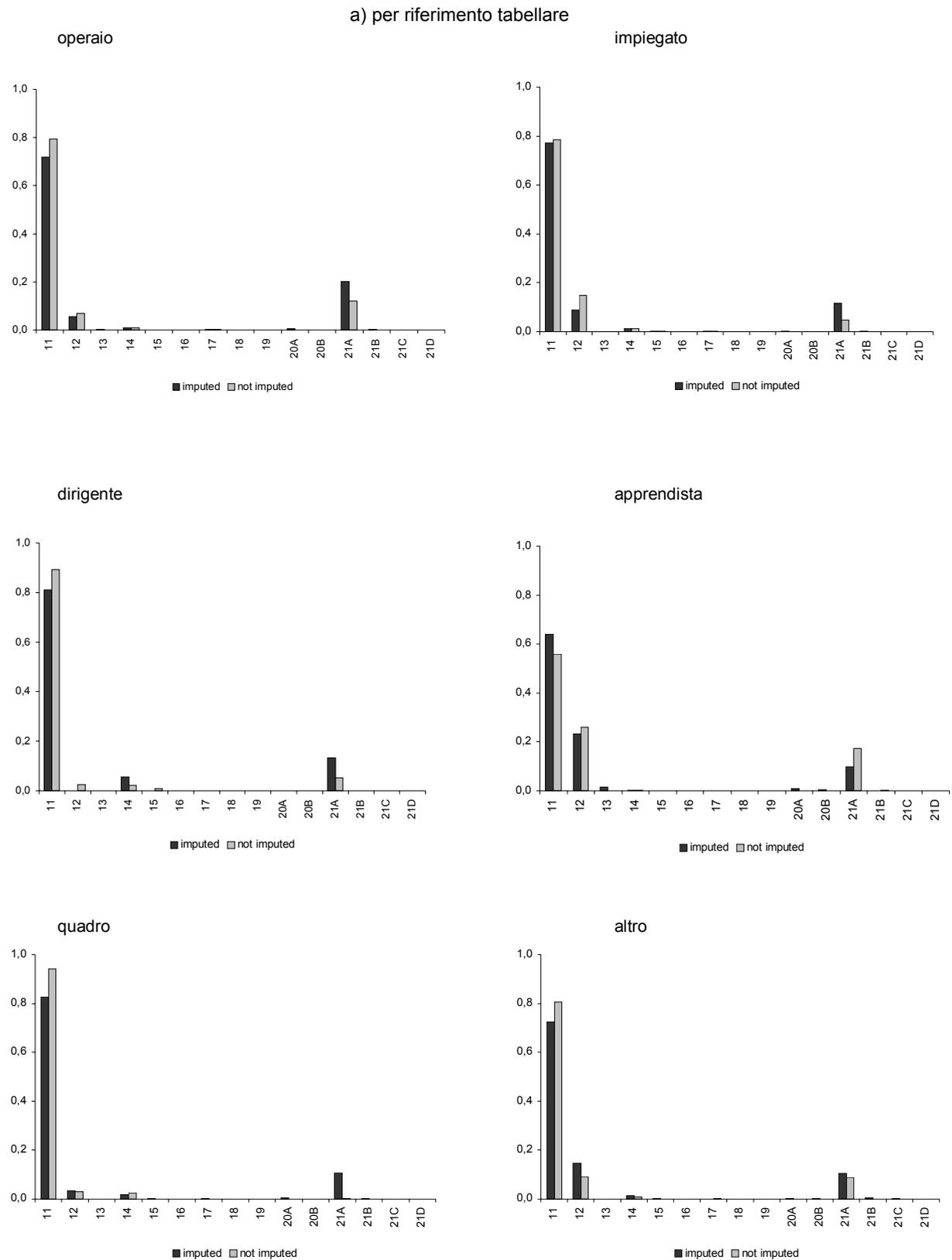
correttezza<sup>44</sup>. I record affetti da informazione mancante o scorretta sono risultati 94.648, pari a circa il 4 per cento dei record rilevanti ai fini della stima Anf. Di questi, solo il 26 per cento presenta informazione incompleta sulla tabella, mentre il 43 per cento sulla classe di reddito e il 98,5 per cento sul numero di componenti familiari. Benché la frazione di dato incompleto risulti statisticamente non significativa nel complesso, si è ritenuto opportuno investigare più approfonditamente gli effetti della procedura di imputazione, in particolare, per valutare l'eventuale distorsione sulla distribuzione finale. La figura 1 riporta per singola macro-qualifica professionale, le distribuzioni delle dichiarazioni complete e imputate rispetto alla tabella, al numero di componenti familiari e all'importo unitario Anf<sup>45</sup>.

---

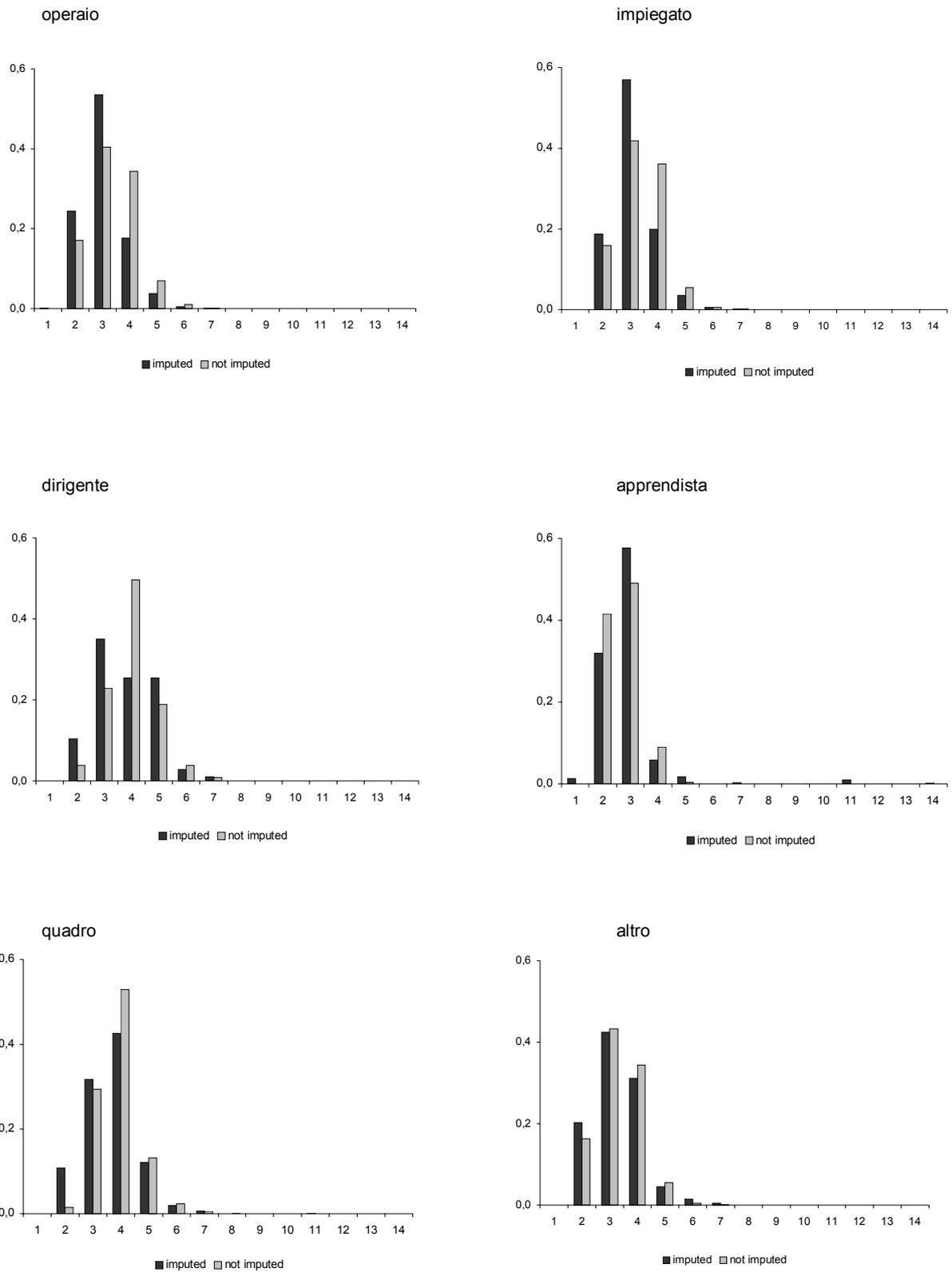
<sup>44</sup> L'analisi delle distribuzioni è stata condotta rispetto a tutte le variabili rilevanti ai fini dell'imputazione (qualifica, sesso, età). Nel testo vengono presentati i risultati solo rispetto alla qualifica lavorativa che è la variabile discriminativa nel primo *step* di selezione del *pool* dei donatori.

<sup>45</sup> Le qualifiche professionali corrispondono al campo 1 della sezione 'dati previdenziali e assistenziali' di competenza Inps. Poiché alcune categorie professionali sono scarsamente rilevanti in termini numerici, si è provveduto ad una riclassificazione per macro-qualifiche: i codici 1, 2, 3 e Q sono stati tenuti distinti, i codici 4 e 5 sono stati aggregati nella categoria degli 'apprendisti', i restanti codici sono stati aggregati nella voce 'altri'. È utile sottolineare che fatta eccezione per i codici 1 e 2, i restanti codici coprono nell'aggregato poco più dell'1% del totale dei record.

**Figura 4.1 - Distribuzione dei dati completi e imputati per coordinate Anf e qualifica professionale**

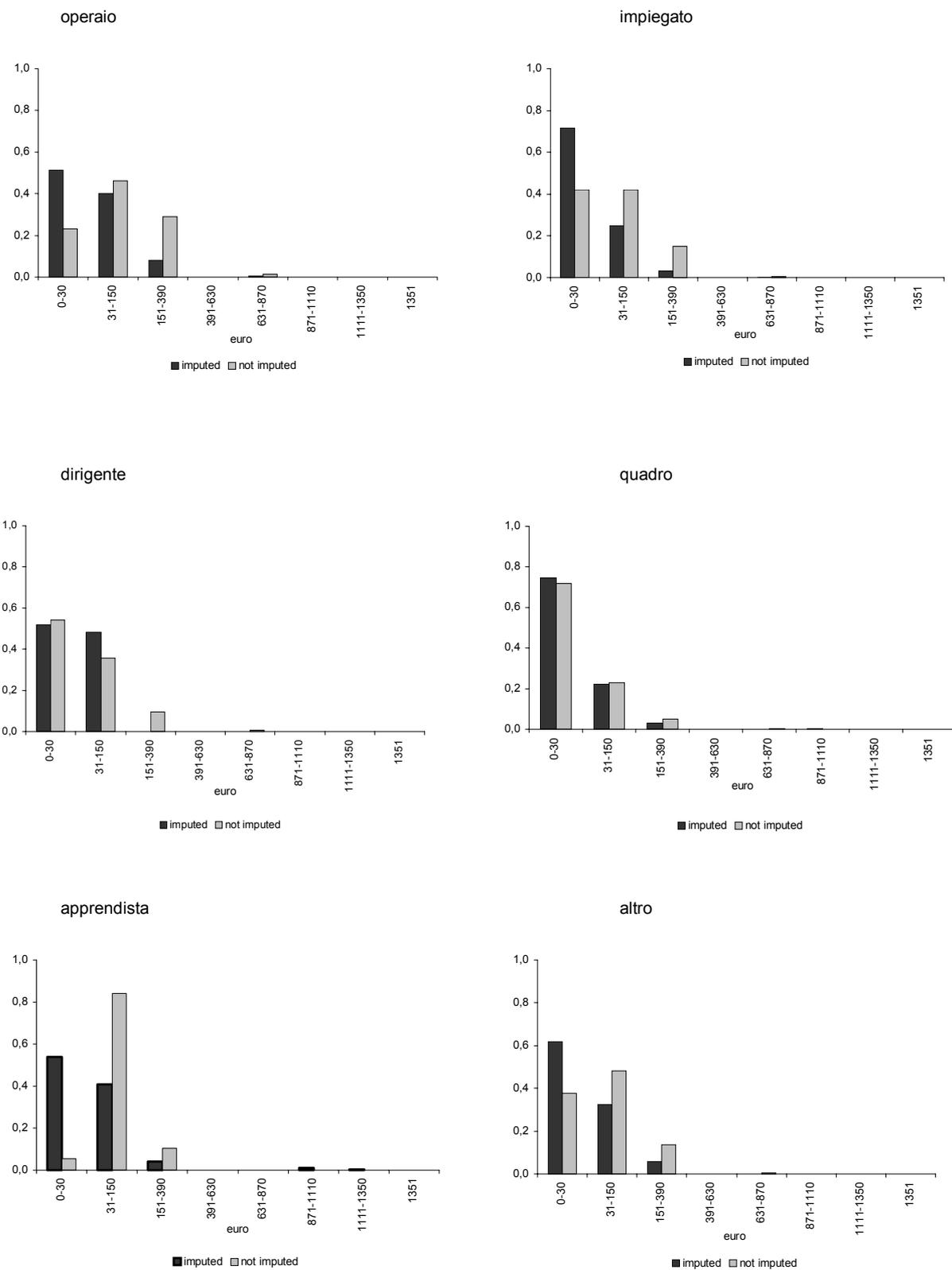


**Figura 4.1 (segue) : Distribuzione dei dati completi e imputati per coordinate Anf e qualifica professionale b) per numero di componenti del nucleo familiare beneficiario**



**Figura 4.1 (segue): Distribuzione dei dati completi e imputati per coordinate Anf e qualifica professionale**

c) per importo unitario Anf



Come evidenziato dalle figure 4.1a, 4.1b e 4.1c, la tecnica di imputazione tende in generale a riprodurre distribuzioni prossime o quasi prossime a quelle del dato completo. Relativamente alla prima coordinata, la similarità tra le due distribuzioni è il risultato della particolare conformazione del dato completo, il quale mostra una tendenza da parte dei potenziali donatori a localizzarsi solo su alcune tabelle (11, 12 e 21A). Scostamenti più apprezzabili, ma non rilevanti, si evidenziano invece per le distribuzioni rispetto alla coordinata 'numero di componenti familiari' e all'importo unitario Anf erogato. In entrambi i casi, la procedura di imputazione riproduce distribuzioni non completamente conformi a quella del dato completo (i dati tendono a concentrarsi verso i valori più bassi) ma senza che si riscontrino deviazioni fondamentali. Dall'analisi emerge pertanto che la metodologia di recupero dell'informazione incompleta, basata sulle caratteristiche socio-demografiche del donatore, è in grado di preservare la distribuzione del dato iniziale. Benché la procedura possa aver indotto distribuzioni del dato imputato lievemente distorte verso il basso rispetto a quelle del dato completo (con conseguente sottostima del fenomeno), la dimensione del dato da sottoporre a correzione (meno del 5 per cento) è tale da non generare ripercussioni statisticamente significative sulla distribuzione finale.

#### 4.5 Osservazioni conclusive

Questo capitolo chiarisce gli aspetti metodologici connessi al trattamento dei dati amministrativi dei modelli di dichiarazione fiscale 770, finalizzato alla produzione di informazioni statistiche sugli assegni al nucleo familiare (Anf). L'attenzione è stata posta sulle problematiche connesse all'individuazione e trattamento delle fonti di errore che potrebbero influire sull'accuratezza statistica delle stime.

Trattandosi di un'analisi sull'intera popolazione, non è stato necessario affrontare il problema della distorsione e perdita di efficienza delle stime derivanti da errori campionari, mentre l'attenzione è stata posta sui problemi di stima generati da errori non campionari. Sono state analizzate nel dettaglio le implicazioni connesse alla possibile incompleta copertura dell'archivio, alla mancata risposta parziale e agli errori di misurazione.

Per quanto riguarda i problemi di copertura dell'archivio, occorre sottolineare che lo studio è risultato limitato dall'impossibilità di disporre di fonti informative caratterizzate dallo stesso livello di disaggregazione del 770, con cui utilmente comparare la popolazione d'interesse. A tal fine, l'analisi è stata condotta a livello di unità d'analisi "azienda" e l'universo delle matricole aziendali 770 di fonte Inps è stato confrontato con quello di fonte Sogei e quello di provenienza Dm10. Una complessa procedura di integrazione dei tre archivi ha evidenziato due criticità. Primo, l'analisi ha messo in luce l'importanza di non sottovalutare il problema delle erogazioni infrannuali di Anf. Secondo, una modesta entità di matricole aziendali potenzialmente coinvolte in erogazioni di Anf ma non rilevate dal 770 è risultata caratterizzata da eventi "eccezionali" (sospensione, cessazione e riattivazione).

Mentre rispetto al primo problema, è stata predisposta un'opportuna procedura di recupero dei percettori infrannuali tramite l'integrazione longitudinale di archivi di anni successivi, il secondo problema rimane una questione aperta. Sarebbe infatti opportuno che l'ente incaricato del rilascio del dato ai fini statistici studiasse il livello di popolamento degli archivi a diverse scadenze e fornisse informazione utile per poter valutare il *trade off* tra costi e benefici connessi ad un eventuale ritardo nell'acquisizione degli archivi rispetto ad un maggior popolamento degli stessi.

Per quanto riguarda il trattamento della componente di errore non campionario connessa all'incompletezza delle coordinate Anf, è utile precisare, anzitutto, che pur risultando la dimensione del dato mancante statisticamente trascurabile (meno del 5 per cento), si è ritenuto opportuno predisporre già dall'immediato una procedura di imputazione che fosse valida anche per gli anni successivi. L'impianto che ne è scaturito, basato su un approccio *ad hoc* di natura deterministico-stocastico con imputazione da donatore, garantisce risultati statisticamente robusti. Analisi sul dato prima e dopo l'imputazione hanno evidenziato infatti, come essa tende a riprodurre distribuzioni prossime o quasi prossime a quelle del dato completo e quindi a non distorcere la distribuzione del dato finale.



## Appendice A

### *La scelta dell'archivio di riferimento per la stima degli assegni al nucleo familiare: studio di compatibilità tra archivi 770 di provenienza Inps e Sogei*

Per la produzione di statistiche sugli Anf si è deciso di utilizzare, tra le due versioni dell'archivio 770 disponibili, quella di fonte Inps. Questa scelta deriva da valutazioni emerse da uno studio di compatibilità in cui l'archivio di fonte Inps è stato posto a confronto con quello rilasciato direttamente dalla Sogei. Ai fini del progetto Tmnp, tale studio ha evidenziato la superiorità della fonte Inps, sia in termini di possibilità di sfruttamento di un archivio più popolato, che di utilizzo di un dato già sottoposto a controlli di qualità.

La sezione relativa ai *Dati previdenziali e assistenziali Inps* è sottoposta a scrupolosa verifica amministrativa volta a stabilire una compatibilità di contenuto tra le voci rilevanti della dichiarazione ed a garantire una completa copertura dell'universo dei soggetti tenuti alla presentazione dei modelli 770. Ne deriva pertanto che l'archivio rilasciato da parte dell'Inps potrebbe non coincidere completamente con quello Sogei, sia a livello di universo di matricole aziendali che di qualità del dato.

La figura A1 sintetizza gli *step* logici seguiti per effettuare la comparazione dei due archivi. L'integrazione delle due fonti è complicata dal fatto che uno stesso codice fiscale dipendente può apparire in più matricole aziendali, in quanto il dipendente ad esso associato potrebbe aver prestato attività lavorativa in più aziende, contemporaneamente o in diversi periodi dell'anno. Per tener conto di questa complicazione, la procedura prevede che, in linea generale (primo e secondo step) l'integrazione avvenga tramite la chiave di aggancio "matricola aziendale-codice fiscale dipendente" (o chiave di primo step), tuttavia, in casi residuali (terzo step) si prescinde dalla matricola aziendale e l'integrazione è effettuata tramite il solo codice fiscale dipendente. Il presupposto della procedura di integrazione si basa sul fatto che, in generale, ciascun record deve essere identificato in maniera univoca dalla chiave "matricola aziendale-codice fiscale dipendente", ovvero, ogni dipendente deve essere individuato rispetto a ciascuna matricola aziendale con cui ha intrapreso un rapporto di lavoro dipendente.

Una ulteriore complicazione deriva dal fatto che, per diverse motivazioni, non vi è perfetta coincidenza tra gli universi delle matricole aziendali dei due archivi. Primo, l'archivio Sogei è rilasciato nel suo formato 'grezzo', pertanto il campo matricola aziendale non sempre è compilato, o lo è in maniera non corretta. Invece l'Inps, per la rilevanza del dato ai fini contributivi, sottopone l'archivio a tutte le correzioni volte a ripristinare la completezza e correttezza dei record, risolvendo le situazioni in cui la matricola aziendale è assente o scorretta. In questi casi, ovvero quando i record non trovano riscontro mediante la chiave di primo step, si tenta una integrazione tramite il solo codice fiscale dipendente.

Secondo, è molto probabile che la fonte Inps sia caratterizzata da un più elevato livello di popolamento rispetto alla fonte Sogei, in quanto l'archivio viene continuamente alimentato anche da dichiarazioni pervenute oltre i termini di presentazione. Ciò implica che è possibile che risultino associazioni matricole aziendali-codici fiscali dipendenti non presenti nell'archivio Sogei. La procedura implementata consente di stimare esattamente l'incidenza di ciascun sottogruppo di record integrati e non integrati e pertanto di formulare valutazioni sulla qualità dei due archivi.

Il primo passo della procedura richiede la predisposizione di archivi omogenei dal punto di vista dell'universo di riferimento. Una serie di filtri preliminari sulla sezione dati previdenziali e assistenziali dell'archivio di fonte Sogei si è resa opportuna per isolare l'universo delle associazioni "matricola aziendale-codice fiscale dipendente" di competenza Inps e, all'interno di queste ultime, quelle rilevanti ai fini Anf. In un primo step di analisi si è ritenuto opportuno mantenere tutte le dichiarazioni di competenza Inps che avessero almeno una coordinata Anf compilata, indipendentemente dal suo contenuto. Relativamente all'archivio dei percettori Anf di fonte Inps, sono stati isolati i record relativi alle sole dichiarazioni correnti, ovvero sono state escluse tutte le dichiarazioni relative ai percettori recuperati grazie all'integrazione longitudinale degli archivi di anni successivi.

I filtri descritti hanno restituito, rispettivamente, 439.072 matricole aziendali Sogei corrispondenti a 1.812.041 codici fiscali dipendente, contro le 503.567 matricole Inps a cui sono associati 2.050.770 codici fiscali dipendente. Come previsto, ne è emerso un maggior popolamento dell'archivio di fonte Inps rispetto a quello di provenienza Sogei.

L'integrazione dei due archivi per matricola aziendale ha restituito nel complesso 436.304 matricole aziendali comuni, rappresentative rispettivamente dell'87 per cento delle matricole aziendali Inps e del 97

per cento di quelle Sogei. Il primo step dell'integrazione lascia fuori circa 67.000 matricole aziendali Inps e 2.600 Sogei, per una frazione di percettori Anf non del tutto trascurabile, 237.870 per la prima fonte e 5.874 per la seconda. Benché sia possibile che la mancata corrispondenza tra le associazioni matricola-codice fiscale possa essere originata da una mal codifica delle matricole Sogei, la forte sproporzione tra le due fonti potrebbe essere indizio di un problema diverso, legato al differente livello di popolamento dei due archivi.

Tra le matricole aziendali comuni ai due archivi, il secondo step della procedura genera un'integrazione quasi perfetta dei codici fiscali dipendente, 1.776.679 associazioni matricola aziendale-codice fiscale dipendente sono comuni ai due archivi, contro i circa 29.000 codici fiscali presenti solo nell'archivio Sogei e i circa 35.000 presenti solo nell'archivio Inps. La proporzionalità tra i due gruppi di codici fiscali dipendente fa sospettare che il problema della mancata integrazione dei due archivi non sia da attribuire a matricole aziendali errate, quanto piuttosto a codici fiscali dipendente errati, che non trovano pertanto corrispondenza nelle due fonti<sup>46</sup>. Per i record Inps e Sogei residuali rispetto all'integrazione per chiave "matricola-codice fiscale" è stata implementata una ulteriore procedura di recupero, basata sulla ricerca della corrispondenza del solo codice fiscale dipendente. Dei circa 270.000 record residuali Inps e 35.000 residuali Sogei, ne sono stati recuperati solo 600.

In conclusione, anche nel caso vi fossero stati errori di trascrizione nei codici fiscali dipendente dell'archivio Sogei, tali da inibire l'integrazione completa dei due archivi, la ridotta dimensione dei record residuali Sogei rispetto a quelli Inps non esclude (anzi rafforza) l'idea che vi siano problemi di diversa copertura degli universi.

In particolare, sarebbe l'archivio di fonte Sogei ad essere maggiormente affetto da problemi di copertura. Inoltre, l'incidenza del numero dei record residuali rispetto a quelli rilevanti ai fini della stima Anf è diversa nei due archivi, meno del 2 per cento per l'archivio Sogei e più del 15 per cento per l'archivio Inps.

Tali valutazioni fanno pertanto propendere per una superiorità dell'archivio Inps sull'archivio Sogei, almeno da un punto di vista del livello di popolamento dei due archivi e della minimizzazione della perdita di potenziali percettori.

Per valutare l'entità delle divergenze tra i due archivi, si sono posti a confronto i dati relativi ai due sotto-archivi costituiti dalle 1.776.679 dichiarazioni comuni derivanti dal primo e secondo step di integrazione. Il prospetto A1 mostra le distribuzioni delle dichiarazioni per valori, coincidenti e non, delle variabili rilevanti ai fini della stima degli Anf.

Le distribuzioni delle dichiarazioni per coordinate Anf e qualifiche sono sostanzialmente identiche nei due archivi. Da ciò si potrebbe dedurre che i dati, già in origine, risultassero coerenti e che le correzioni apportate dall'Inps siano state comunque di modesta entità. In particolare, rispetto alle coordinate Anf, lo scostamento sul numero di componenti familiari è quello che incide in misura lievemente superiore rispetto a tutti gli altri. Ciò potrebbe essere la conseguenza del tipo di procedura adottato per la correzione dei dati anomali, il quale prevede che per i valori superiori ad una soglia ritenuta ragionevolmente ammissibile (15 componenti), venga effettuato un riallineamento secondo regole deterministiche. È pertanto probabile che tra i dati non coincidenti, i valori fuori range abbiano un peso determinante.

#### **Prospetto A1 - Distribuzione per valori coincidenti e non delle variabili**

	Coincidenti	Non coincidenti
<i>tabella ANF</i>	1.787.942 (98.75)	22.319 (1.25)
<i>classe ANF</i>	1.750.914 (97.93)	37.028 (2.07)
<i>numero fam ANF</i>	1.746.976 (97.71)	40.966 (2.29)
<i>qualifica V1</i>	1.785.703 (99.87)	2.239 (0.13)
<i>qualifica V2</i>	1.783.125 (99.73)	4.817 (0.27)
<i>qualifica V3</i>	1.786.255 (99.91)	1.687 (0.09)
<i>mesi retribuiti</i>	1.408.944 (78.80)	378.998 (21.20)

<sup>46</sup> Un'analisi sulla lunghezza dei codici fiscali ha rilevato un numero molto esiguo di campi errati, pertanto eventuali errori dovrebbero riguardare la struttura interna del codice.

Divergenze più rilevanti sono invece osservabili per il campo ‘mesi retribuiti’, in cui più del 20 per cento di record restituisce informazione discordante. Un’analisi sulle restati variabili di computo del periodo lavorativo, ha evidenziato che solo il 3,4 per cento dei record, per le settimane retribuite e il 3,6 per cento, per i giorni retribuiti, è caratterizzato da un vettore di mesi non coincidente. Ciò fa sospettare che nella procedura di correzione delle variabili temporali, l’Inps attribuisca maggior affidabilità alla correttezza delle variabili disaggregate giorni e settimane, effettuando in funzione delle stesse l’adeguamento del vettore dei mesi.



## Capitolo 5 – Validazione e controllo di qualità delle stime degli Anf

### 5.1 Valutazione del livello di precisione delle stime

In generale, l'accuratezza statistica delle stime necessaria per giudicare la fattibilità di queste ultime a livello di diffusione dovrebbe essere valutata facendo ricorso a dati di *benchmark*. Cholette and Dagum (1994) definiscono il concetto di *benchmark* come segue "... si è in presenza di situazioni di *benchmark* quando sono disponibili due o più fonti di dati sulla stessa variabile d'interesse, con frequenze diverse, ad esempio, mensile contro annuale o mensile contro trimestrale. Generalmente, le due fonti di dati forniscono risultati diversi, ad esempio la somma delle misurazioni mensili potrebbe non coincidere con la misurazione annuale...*La misura più attendibile deve essere considerata il benchmark...*tipicamente, un processo di convergenza verso il *benchmark* consiste di una combinazione di due fonti di dati volta a raggiungere un miglioramento delle stime delle serie investigate. Sotto questa prospettiva, i *benchmark* rappresentano *osservazioni ausiliarie*"<sup>47</sup>.

Allo stato attuale non esistono tuttavia, fonti informative caratterizzate dallo stesso livello di disaggregazione del dato da 770, che consentono di valutare l'accuratezza statistica delle stime Anf generate tramite tale archivio.

Premesso che non può trattarsi di un *benchmark* nel senso sopra descritto, in quanto diversi sono i presupposti di rilevazione e i criteri di contabilizzazione con cui vengono rilevati gli eventi associati all'erogazione di Anf, l'archivio fiscale dei modelli Dm10 consente di effettuare, a livello di entità "matricola aziendale", utili analisi comparative. L'archivio dei modelli Dm10, con cui le aziende richiedono la restituzione delle somme anticipate per conto dell'Ente Previdenziale, contiene informazioni sugli Anf che possono essere disaggregate a diversi livelli di classificazione Ateco. Come anticipato sopra, le stime da Dm10 non possono considerarsi tuttavia un *benchmark* per le stime da 770 nel senso rigoroso del termine. Diversi sono infatti i presupposti di rilevazione, fiscale per il primo e di pura comunicazione per il secondo e i criteri di rilevazione, basato su un principio di cassa nel caso del Dm10 e di competenza nel 770. Rispetto a quest'ultima diversità, in particolare, potrebbero originarsi delle differenze anche sostanziali nelle distribuzioni degli importi stimati, nel caso in cui i conguagli relativi ad anni precedenti a quello di competenza fossero rilevanti. Un fenomeno di questo genere, in particolare è atteso negli aggregati più numerosi in cui confluiscono imprese anche con caratteristiche molto diverse.

### 5.2 Popolazione e parametri di interesse

Per valutare la precisione delle stime e quindi analizzare la fattibilità del metodo utilizzato si è fatto ricorso alla teoria tradizionale dei test. Nella validazione delle stime Anf tramite gli archivi amministrativi dei modelli 770 e Dm10, la comparazione riguarda parametri non di campioni ma di due rappresentazioni (simili ma non coincidenti) di una stessa popolazione. Ci si trova pertanto in presenza di una situazione sperimentale particolarmente favorevole, in cui la conoscenza dell'intero universo delle matricole aziendali equivale alla conoscenza della forma distributiva dell'intera popolazione, nonché alla validità del teorema del limite centrale<sup>48</sup>.

Prima di procedere alla descrizione delle varie fasi di validazione del dato è utile fornire alcuni chiarimenti tecnici sull'universo di riferimento e sulla variabile di aggregazione utilizzata per effettuare i confronti tra 770 e Dm10. L'unità di rilevazione del modello 770 è rappresentata dal dipendente (titolare di assegno nel caso Anf) mentre quella del Dm10 dall'azienda. La chiave di aggancio dei due archivi è rappresentata dalla

<sup>47</sup> Cfr. Draper D. and Bewater R., "Model Quality Report in Business Statistics", volume I, ch. 9, in Pam Davies and Paul Smith.

<sup>48</sup> Il *teorema centrale limite*, che si riferisce a "qualunque" popolazione da cui si estraggano dei campioni, afferma: "Si traggano campioni casuali di  $n$  osservazioni da una popolazione con media  $\mu$  e scarto quadratico medio finito (o deviazione standard)  $\sigma$ . Allora, quando  $n$  è grande, la media campionaria sarà approssimativamente distribuita normalmente, con media uguale a  $\mu$  e scarto quadratico medio  $\sigma/n^{1/2}$ . L'approssimazione diventerà sempre più accurata via via che  $n$  cresce".

matricola aziendale Inps. Per l'anno di imposta 2002, l'archivio mod. 770 ha rilevato la presenza di 532.379 matricole aziendali distinte. Per effetto della diversa data di rilascio, i due archivi non sono risultati completamente integrabili; 19.085 aziende non hanno trovato corrispondenza nell'archivio fiscale dei Dm10. Benché le imprese non integrate risultino in proporzione non trascurabile (circa il 4 per cento del totale delle matricole presenti nell'archivio 770), opportune analisi sulla distribuzione del dato in loro presenza o assenza hanno evidenziato che esse non possono considerarsi un campione autoselettivo della popolazione e pertanto la loro inclusione non produce effetti distorsivi sulle stime finali<sup>49</sup>.

La validazione del dato è stata pertanto condotta sull'universo delle matricole aziendali comuni all'archivio amministrativo 770 e fiscale Dm10 e è consistita in:

- studio della distribuzione empirica del dato sulle due popolazioni di riferimento e fattibilità dell'approssimazione alla distribuzione normale (analisi descrittiva delle principali misure di posizione e momenti della distribuzione, analisi teorica delle densità di probabilità);
- studio della significatività dei parametri delle distribuzioni, tramite test di localizzazione;
- studio del livello di accostamento delle stime Anf da mod. 770 a quelle da Dm10 tramite indici di accostamento, comparazione delle distribuzioni empiriche e test parametrici di similarità delle distribuzioni.

### 5.3 Ipotesi distributive e metodo di stima

L'assunzione cruciale su cui si basa la maggioranza dei test parametrici è la normalità della distribuzione sottostante la popolazione di interesse. Benché l'ipotesi di normalità sia spesso assunta in maniera implicita e conveniente, l'inferenza basata su modelli in cui essa fosse violata non sarebbe affidabile. Un modo semplice ed efficace per testare l'ipotesi di normalità consiste nello studiare la distribuzione empirica del dato e i suoi principali momenti e analizzarne l'eventuale significatività della deviazione rispetto alla distribuzione normale<sup>50</sup>.

I grafici costruiti sulle distribuzioni dei dati osservati, mostrano in generale, sia a livello aggregato sia a livello di sezione Ateco, una discreta approssimazione del modello teorico normale alla distribuzione empirica del dato stimato. Nel dettaglio, le figure 5.1 e 1a mettono in relazione funzioni empiriche cumulate dei dati stimati, rispettivamente tramite 770 e Dm10, con la densità cumulata normale. Ne emerge che per entrambe le fonti di dati amministrativi esiste un buon "match" tra dato osservato e modello teorico normale. Scostamenti dall'ipotesi di normalità, in misura non apprezzabile, si osservano in generale agli estremi delle distribuzioni.

La plausibilità dell'approssimazione alla distribuzione normale è ulteriormente convalidata dagli istogrammi di cui alle figure 5.2 e 5.2a, in cui le densità di probabilità assolute sono relazionate alla densità normale. Unitamente alle informazioni contenute nel prospetto 5.2, gli istogrammi consentono di argomentare ulteriormente sulle caratteristiche del dato osservato; il discreto livello di accostamento dei tre indicatori statistici di posizione (media, moda e mediana) e i valori degli indici di curtosi, generalmente prossimo allo zero, sono a sostegno dell'adeguatezza del modello teorico prescelto. La prevalenza del segno negativo per l'indice di simmetria, denota inoltre una tendenza del dato all'addensamento verso valori più elevati, anche se la modesta entità dell'indice non è in contrasto con l'ipotesi di normalità della distribuzione.

Rispetto alle misure di posizione e variabilità, il Box-Whisker *plot* di cui alla figura 5.3 e il prospetto 5.2, mostrano come vi sia una discreta aderenza tra i dati dei due archivi, anche se i risultati Anf da 770 mostrano una lieve sovrastima in media rispetto a quelli da Dm10 e questi ultimi una maggiore variabilità.

Poiché esiste sufficiente evidenza a favore dell'ipotesi di normalità delle due distribuzioni è possibile ricorrere alla tradizionale teoria dei test per studiare la conformità dei due sottouniversi. A scopo puramente descrittivo, i prospetti 5.3 e 5.4 riportano i risultati dei test di nullità dei parametri relativi agli importi medi e mediani Anf stimati tramite gli archivi 770 e Dm10, rispettivamente nelle versioni:

- parametrica, test *t* di Student, per la validazione dell'ipotesi di media non nulla;
- non parametrica, test dei segni e test di Wilcoxon (o test dei segni per ranghi), per la validazione dell'ipotesi di mediana non nulla.

<sup>49</sup> I motivi della non perfetta sovrapposizione dei due universi sono ampiamente discussi nell'Appendice A del cap. 4, pag. 56.

<sup>50</sup> Test sull'ipotesi di normalità non sono in realtà fondamentali in questo contesto di analisi. La dimensione dell'universo è tale che il teorema del limite centrale è pienamente valido.

I tre test concludono univocamente sulla non nullità dei parametri d'interesse, media e mediana, a livello di significatività  $\alpha = 0.05$ .

#### 5.4 Valutazione empirica della fattibilità delle stime

Data la possibilità di disporre di informazioni disaggregate a diversi livelli di classificazione Ateco, sono state effettuate una serie di analisi comparative tra 770 e Dm10, volte a studiare empiricamente il grado di similarità tra le due distribuzioni tramite il confronto dei parametri rilevanti. Nel condurre l'analisi ci si è trovati di fronte ad una situazione sperimentale piuttosto favorevole da un punto di vista statistico:

- si dispone dell'intero universo di matricole aziendali coinvolte nell'erogazione di Anf. Ciò equivale a supporre la validità del teorema del limite centrale, che per numerosità campionarie sufficientemente grandi, garantisce l'approssimazione alla distribuzione normale di certe statistiche utili al confronto tra universi;
- l'ipotesi di normalità è ulteriormente avvalorata dall'analisi empirica del dato, che mostra essere ben approssimabile dalla distribuzione teorica.

I prospetti 5.5 e 5.6 riportano i risultati dei test di differenze tra varianze e tra medie sulle distribuzioni degli importi stimati tramite archivio Dm10 e archivio 770. Le ipotesi da sottoporre ai test consistono nel verificare che non esistono differenze significative nella struttura di variabilità e negli importi medi (o mediani) stimati sui due archivi. Formalmente ciò equivale alla verifica dei seguenti sistemi d'ipotesi:

a) per gli importi medi stimati

$$\begin{aligned} H_0 : \mu_{DM} - \mu_{770} &= \delta && \text{medie non significativamente diverse} && (1) \\ H_1 : \mu_{DM} - \mu_{770} &\neq \delta && \text{medie significativamente diverse} \end{aligned}$$

con  $\delta = 0$ ;

b) per la struttura di variabilità degli importi stimati

$$\begin{aligned} H_0 : \sigma_{DM}^2 &= \sigma_{770}^2 && \text{universi omoschedastici} && (2) \\ H_1 : \sigma_{DM}^2 &> \sigma_{770}^2 && \text{universi eteroschedastici} \end{aligned}$$

La verifica dell'ipotesi di omoschedasticità degli universi di provenienza è pre-requisito ai fini della scelta della statistica più appropriata per valutare la significatività della differenza tra medie. Nel caso di omoschedasticità dei due universi di riferimento, ovvero di validità dell'ipotesi  $H_0$  relativa al sistema (2), la statistica più appropriata è quella basata sulla distribuzione normale:

$$\frac{(\bar{m}_{DM} - \bar{m}_{770}) - \delta}{\sqrt{\frac{\sigma_{DM}^2}{n_{DM}} + \frac{\sigma_{770}^2}{n_{770}}}} \sim N(0,1) \quad (3)$$

con  $\bar{m}$ , importo medio stimato sui due archivi,  $\sigma^2$ , rispettiva varianza,  $n$ , dimensione dei due archivi e  $\delta$ , scostamento ammesso (nel caso in esame pari a zero).

Nel caso di universi eteroschedastici, ovvero di violazione dell'ipotesi  $H_0$  relativa al sistema (2), il test appropriato è quello che utilizza una combinazione delle varianze dei due universi,  $s_p^2$ , anche detta varianza *pooled*<sup>51</sup> e è approssimabile alla distribuzione teorica *t*-Student:

<sup>51</sup> Pari a  $s_p^2 = \frac{(n_{DM}^2 - 1)s_{DM}^2 + (n_{770}^2 - 1)s_{770}^2}{n_{DM} + n_{770} - 2}$ .

$$\frac{(\bar{m}_{DM} - \bar{m}_{770}) - \delta}{\sqrt{s_p^2 \left[ \frac{1}{n_{DM}} + \frac{1}{n_{770}} \right]}} \sim t_{n_{DM} + n_{770} - 2} \quad (4)$$

È utile rilevare che, quando l'universo di provenienza è di tipo normale e la numerosità delle osservazioni è sufficientemente ampia, anche la statistica (4) tende ad approssimarsi ad una distribuzione normale standard. Il prospetto 5.5 riporta i risultati dell'applicazione del sistema d'ipotesi (2) alle matricole aziendali comuni all'archivio dei modelli 770 e Dm10. Il sistema di ipotesi (2) è stato testato utilizzando due versioni 'robuste' del test di omoschedasticità: il test di Levene sulle deviazioni quadratiche dalla media e il test di Brown-Forsythe sulle deviazioni assolute dalla mediana. Nella formulazione generale, i test basati sulle deviazioni sono approssimabili alla distribuzione  $F$  di Snedecor e possono essere formalizzati come segue:

$$W_0 = \frac{(N-k) \cdot \frac{\sum_{i=1}^K n_i (\bar{Z}_i - \bar{\bar{Z}})^2}{k-1}}{\sum_{i=1}^k \frac{\sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2}{n_i}} > F_{(\alpha, k-1, N-k)} \quad (5)$$

dove  $Z_{ij}$  rappresenta lo scarto in valore assoluto di ogni valore  $X_{ij}$  dalla media di gruppo  $\bar{X}_i$ ,  $\bar{Z}_i$  la media di gruppo di  $Z_{ij}$  e  $\bar{\bar{Z}}$  la media generale delle osservazioni  $Z_{ij}$ <sup>52</sup>.

Il test di Levene, nella versione basata sulla media aritmetica, si presta particolarmente bene quando la distribuzione empirica del dato è approssimabile ad una normale, la versione basata sulle mediane risulta invece più robusta quando si sospetta che la distribuzione del dato non sia perfettamente simmetrica. Benché esista sufficiente evidenza sulla fattibilità dell'approssimazione delle distribuzioni empiriche alla normale, si è ritenuto opportuno supportare i test di omoschedasticità basati sulle differenze rispetto alla media con un test più robusto rispetto a scostamenti dall'ipotesi di normalità. Gli indici di *skewness* hanno infatti denotato una lieve tendenza del dato ad addensarsi verso valori più alti (asimmetria negativa), si è ritenuto pertanto utile eseguire anche il test sulle differenze rispetto alla mediana, più robusto in presenza di distribuzioni asimmetriche.

I risultati dei due test per la validazione dell'ipotesi di omoschedasticità delle popolazioni (Prospetto 5.8) non sempre coincidono. Mentre il test basato sulle deviazioni quadratiche non rigetta mai l'ipotesi di uguaglianza tra varianze, né a livello generale né di sezione Ateco, il test di Brown Forsythe non permette di concludere in assoluto sull'omogeneità della struttura di variabilità dei due universi posti a confronto. In particolare, il test non supporta l'ipotesi nulla sia per l'intero universo delle matricole aziendali coinvolte in eventi di erogazione di Anf e che per le due sezioni Ateco F e G.

Il prospetto 5.6 riporta le statistiche test sulle differenze tra medie calcolate rispettivamente nel caso di validità dell'ipotesi di omoschedasticità (metodo *Pooled*) e nel caso di violazione (metodo *Satterthwaite*). I test  $t$  di Student condotti sull'universo delle matricole aziendali comuni ai due archivi non sempre concludono a favore dell'ipotesi di uguaglianza degli importi medi stimati tramite i due archivi. Ad eccezione delle sezioni B e M, il cui peso sul totale delle matricole aziendali è comunque contenuto (poco più dell'1%), i restanti casi in cui il test segnala una violazione dell'ipotesi nulla, riguardano le sezioni più rilevanti sia in termini di numerosità delle matricole che di importi Anf stimati. Nel complesso, tali sezioni (D, F e G) si riferiscono infatti al 70 per cento circa delle matricole aziendali, le quali erogano Anf per un ammontare complessivo pari al 68 per cento del totale. Un'analisi sulle dimensioni aziendali ha evidenziato che si tratta per lo più di imprese di piccole dimensioni, la cui eterogeneità giustificherebbe in parte i risultati.

Malgrado i test statistici non sempre forniscano sostegno empirico alla confrontabilità dei due universi, i risultati vanno interpretati con cautela e certamente non come chiave risolutiva per decidere sulla fattibilità delle stime tramite 770. Infatti, anche se l'archivio Dm10 è sufficientemente affidabile, per la sua natura di dato fiscale, non si può prescindere dal fatto che il criterio contabile di rilevazione delle informazioni sugli Anf è diverso da quello del 770. In un'ottica di lungo periodo, focalizzare l'attenzione su questa fonte di

<sup>52</sup> Ovviamente il test sulle medie quadratiche anziché sui valori assoluti delle differenze si basa sulle differenze quadratiche e il test di Brown Forsythe calcola le differenze rispetto alla mediana e non alla media.

errore potrebbe tuttavia non essere necessario. Quando gli archivi *Emens* mensilizzati saranno disponibili, non solo sarà possibile chiarire le cause delle discrepanze sopra evidenziate ma saranno superate tutte le limitazioni connesse alla verifica ex-post della qualità del dato stimato.

## 5.5 Osservazioni conclusive

L'attenzione, in questo capitolo, è stata posta sulle metodologie per la validazione e controllo di qualità dei risultati di stima degli assegni al nucleo familiare tramite i modelli 770. A tal fine, sono stati infine condotti una serie di test statistici finalizzati alla verifica delle ipotesi di uguaglianza dei parametri fondamentali della distribuzione degli importi Anf stimati tramite gli archivi 770 rispetto a quella derivata dagli archivi dei modelli di dichiarazione fiscale Dm10/2.

La disponibilità di un dato disaggregato a vari livelli di classificazione Ateco, ha consentito di effettuare utili analisi comparative tra archivio 770 e Dm10. Malgrado l'analisi esplorativa delle due distribuzioni non abbia evidenziato scostamenti fondamentali tra le due distribuzioni, sia in generale che a livello di sezione Ateco, i test parametrici non sempre sono risultati a sostegno della confrontabilità delle due distribuzioni stimate. Mentre l'ipotesi di uniformità tra le strutture di variabilità delle due distribuzioni è rifiutata solo in alcuni casi, il test condotto sulle misure di posizione rigetta l'ipotesi in un discreto numero di sezioni. Tali ipotesi sono violate, in particolare, nelle sezioni più popolate sia in termini di matricole aziendali che di importi stimati. Un'analisi sulle dimensioni aziendali ha evidenziato che si tratta per lo più di imprese di piccole dimensioni, la cui eterogeneità giustificerebbe in parte i risultati.

Benché i test statistici non sempre forniscano sostegno empirico alla confrontabilità dei due universi, i risultati vanno interpretati con cautela e certamente non come chiave risolutiva sulla fattibilità delle stime. A tal proposito sono necessarie alcune precisazioni. Primo, allo stato attuale l'unico dato di *benchmark* a cui riferire le stime da 770 è quello fornito in forma aggregata dalla contabilità nazionale e pubblicato a livello ufficiale. Rispetto a quest'ultimo, le stime 770 sono molto più aderenti di quanto non lo siano quelle da Dm10. Come spiegato in precedenza infatti, i modelli 770 e Dm10 rilevano l'informazione amministrativa secondo un diverso criterio contabile, il primo per competenza, il secondo per cassa. Nel valutare la misura dello scostamento tra le due stime occorre pertanto tener conto di questa differenza di contabilizzazione che, vista l'incidenza delle matricole aziendali contemporaneamente assenti nel 770 di fonte Inps e Sogei ma non nel Dm10, potrebbe essere non del tutto irrilevante. In un'ottica di lungo periodo, inoltre, le fonti informative che si renderanno disponibili (archivi *Emens*) inducono a privilegiare il criterio della competenza e se si vuole seguire un approccio uniforme, le stime da archivio 770 sono quelle che allo stato attuale garantiscono una continuità temporale. In relazione a quest'ultimo aspetto, c'è inoltre da rilevare che l'*Emens* nasce come un modello in cui, in ogni esercizio, il consolidamento viene realizzato tenendo conto delle quadrature effettuate con il modello Dm10. Quando gli archivi mensilizzati saranno disponibili quindi, non solo sarà possibile chiarire le cause delle discrepanze sopra evidenziate ma saranno superate tutte le limitazioni connesse alla verifica ex-post della qualità del dato stimato.

Figura 5.1 - Distribuzione degli importi Anf stimati tramite 770 e Dm10 (probability plot)

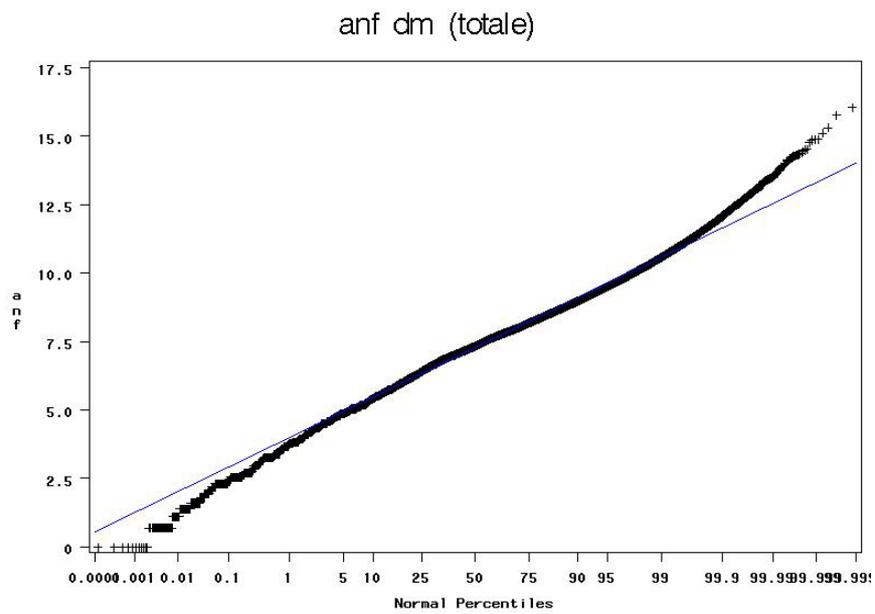
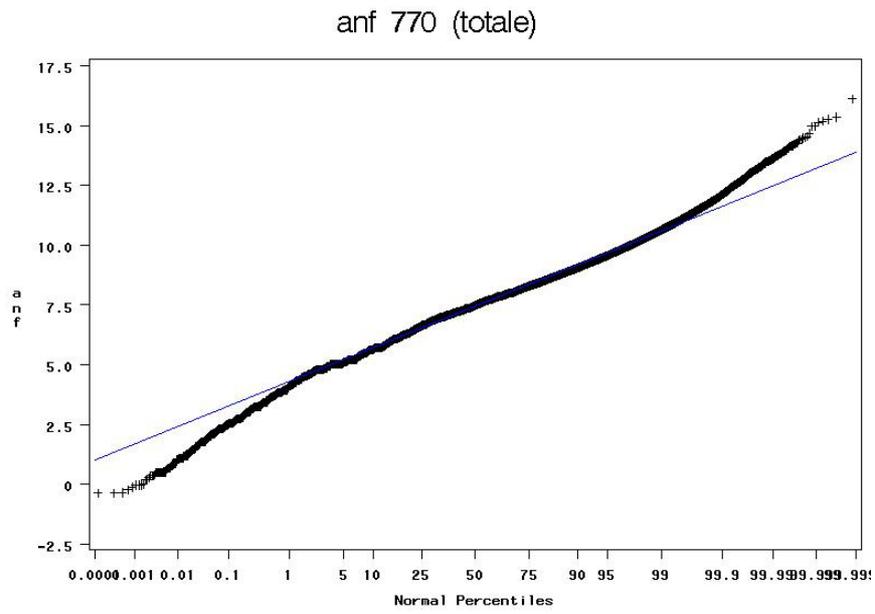


Figura 5.1a - Distribuzione degli importi Anf stimati tramite 770 per sezione Ateco (probability plot)

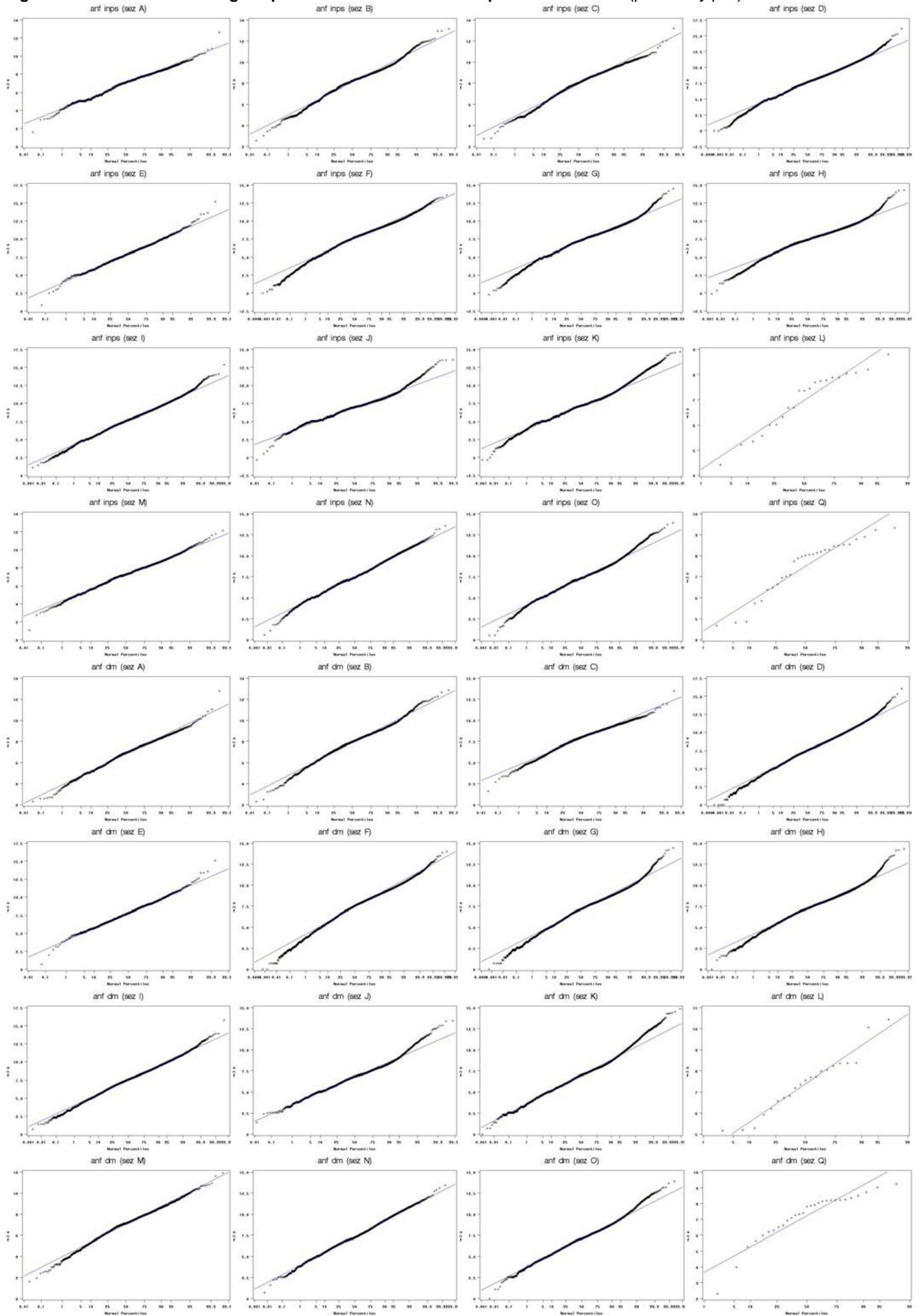


Figura 5.2 - Distribuzione empirica e approssimazione alla normale degli importi Anf stimati tramite 770 e Dm10

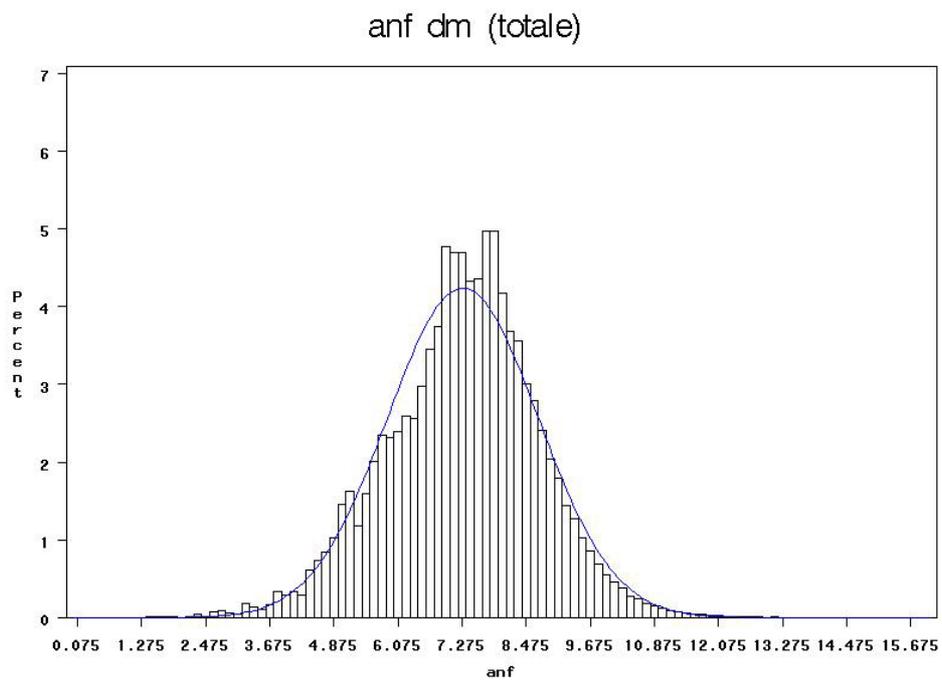
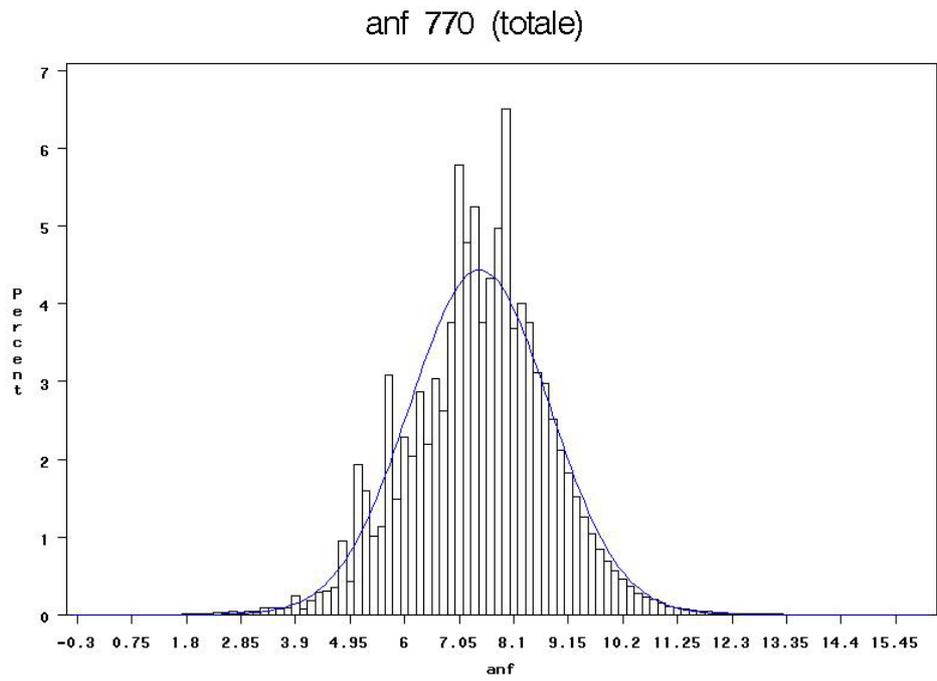


Figura 5.2a - Distribuzione empirica degli importi Anf stimati tramite 770 e Dm10 per sezione Ateco

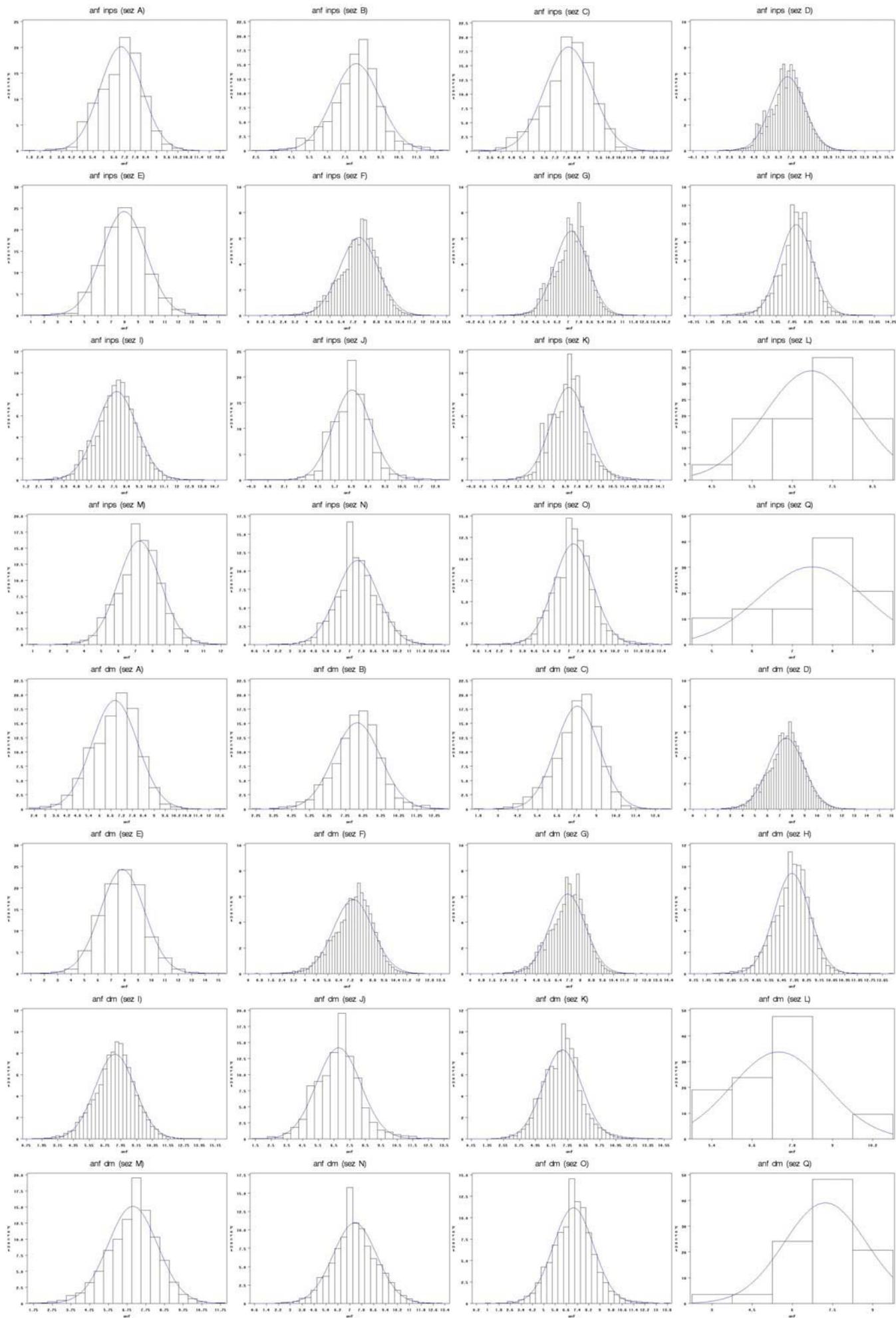
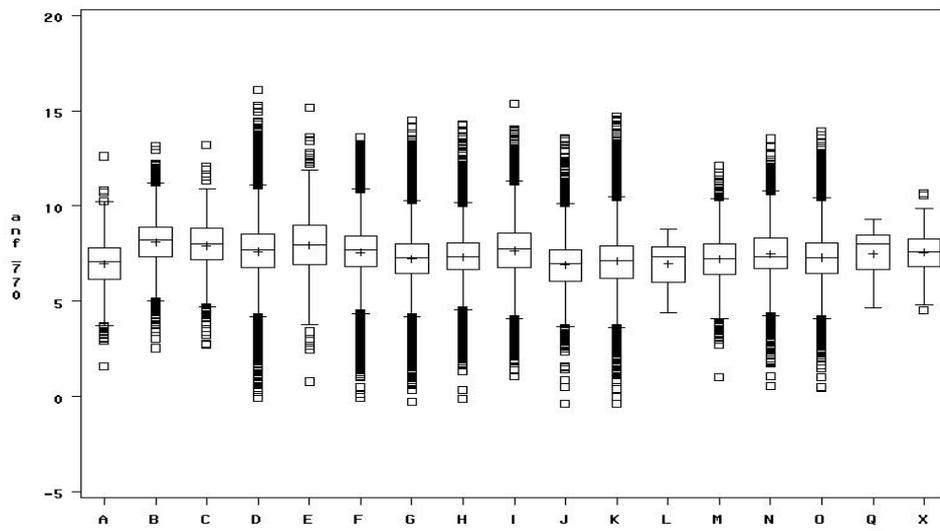
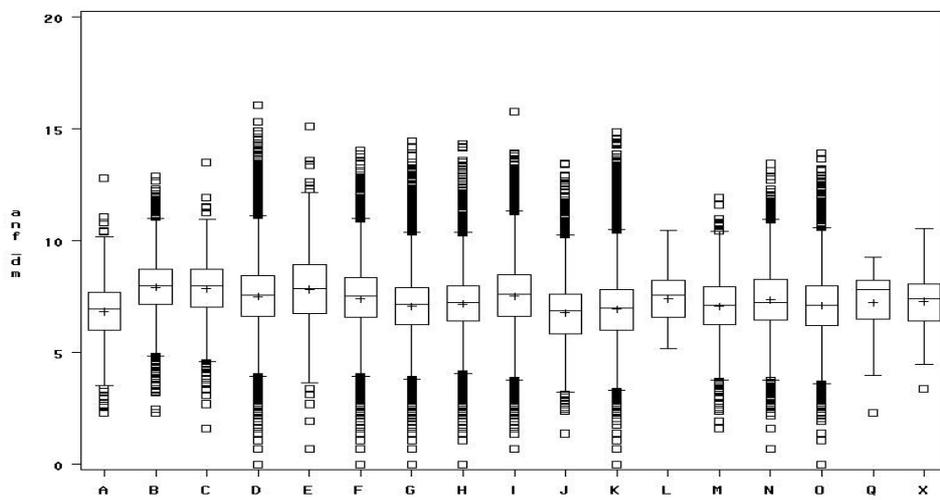


Figura 5.3 - Box and Whisker comparative plot degli importi Anf stimati per sezioni Ateco

a) da archivio 770



a) da archivio Dm10/2



**Prospetto 5.1 - Analisi descrittiva degli importi Anf stimati – livello di classificazione: sezione Ateco.**

ARCHIVIO modd. 770									
	A	B	C	D	E	F	G	H	I
<b>Misure statistiche di posizione</b>									
Media	6,983	8,105	7,921	7,614	7,950	7,571	7,220	7,326	7,664
Mediana	7,087	8,221	8,025	7,691	7,978	7,721	7,313	7,358	7,735
Moda	5,043	8,008	5,736	5,736	5,736	7,358	5,736	7,358	5,736
<b>Misure statistiche di variabilità</b>									
Deviazione Std	1,190	1,317	1,311	1,393	1,647	1,319	1,219	1,212	1,451
Varianza	1,416	1,733	1,719	1,940	2,714	1,740	1,485	1,470	2,106
Intervallo	11,019	10,604	10,475	16,138	14,380	13,632	14,728	14,389	14,279
Intervallo interquartile	1,634	1,544	1,644	1,727	2,077	1,631	1,530	1,402	1,801
<b>Momenti</b>									
Skewness	-0,198	-0,209	-0,482	-0,125	0,081	-0,428	-0,316	-0,386	-0,083
Kurtosis	0,326	0,888	0,325	0,431	0,769	0,595	0,613	1,406	0,468
Coeff di var	17,043	16,243	16,552	18,297	20,720	17,423	16,877	16,547	18,934
SS corretta	2.549,35	4.527,04	3.977,58	300.853,51	2.803,26	175.629,34	155.888,08	46.315,47	60.200,42
Errore std media	0,028	0,026	0,027	0,004	0,051	0,004	0,004	0,007	0,009
N	1.801	2.613	2.315	155.041	1.034	100.948	104.991	31.516	28.588
	J	K	L	M	N	O	Q	Non rip	Totale
<b>Misure statistiche di posizione</b>									
Media	6,927	7,133	6,976	7,217	7,496	7,307	7,493	7,569	7,444
Mediana	6,980	7,150	7,358	7,231	7,358	7,315	8,008	7,599	7,499
Moda	7,087	7,358	7,358	7,087	7,087	7,087	-	8,008	7,358
<b>Misure statistiche di variabilità</b>									
Deviazione Std	1,369	1,386	1,176	1,238	1,396	1,360	1,322	1,201	1,350
Varianza	1,873	1,920	1,384	1,533	1,950	1,848	1,746	1,443	1,822
Intervallo	13,908	15,032	4,380	11,082	12,963	13,431	4,637	6,159	16,475
Intervallo interquartile	1,634	1,713	1,863	1,599	1,636	1,589	1,800	1,452	1,655
<b>Momenti</b>									
Skewness	0,333	0,280	-0,602	-0,137	0,062	0,105	-0,830	-0,015	-0,128
Kurtosis	1,917	1,178	-0,579	0,564	0,513	1,147	-0,247	0,070	0,622
Coeff di var	19,760	19,428	16,865	17,158	18,628	18,607	17,636	15,867	18,131
SS corretta	8.590,76	100.074,25	27,68	4.477,17	20.131,70	26.472,33	48,90	183,20	934.962,27
Errore std media	0,020	0,006	0,257	0,023	0,014	0,011	0,245	0,106	0,002
N	4.587	52.112	21	2.921	10.325	14.324	29	128	513.294
ARCHIVIO modd. Dm10									
	A	B	C	D	E	F	G	H	I
<b>Misure statistiche di posizione</b>									
Media	6,830	7,909	7,832	7,498	7,822	7,399	7,053	7,172	7,510
Mediana	6,955	7,976	7,984	7,570	7,833	7,542	7,167	7,253	7,588
Moda	5,050	7,178	5,743	5,743	5,050	7,007	5,743	7,090	7,737
<b>Misure statistiche di variabilità</b>									
Deviazione Std	1,260	1,323	1,329	1,445	1,643	1,384	1,290	1,280	1,515
Varianza	1,587	1,750	1,768	2,088	2,699	1,916	1,665	1,638	2,296
Intervallo	10,488	10,574	11,872	16,047	14,417	14,013	14,459	14,329	15,067
Intervallo interquartile	1,719	1,547	1,678	1,801	2,190	1,760	1,648	1,585	1,882
<b>Momenti</b>									
Skewness	-0,268	-0,207	-0,495	-0,169	0,078	-0,427	-0,343	-0,280	-0,159
Kurtosis	0,400	0,961	0,577	0,407	0,792	0,472	0,506	0,960	0,447
Coeff di var	18,444	16,727	16,974	19,270	21,004	18,709	18,294	17,845	20,177
SS corretta	2.856,39	4.570,88	4.090,04	323.689,76	2.788,39	193.425,72	174.785,19	51.612,42	65.647,78
Errore std media	0,030	0,026	0,028	0,004	0,051	0,004	0,004	0,007	0,009
N	1.801	2.613	2.315	155.041	1.034	100.948	104.991	31.516	28.588
	J	K	L	M	N	O	Q	Non rip	Totale
<b>Misure statistiche di posizione</b>									
Media	6,785	6,951	7,394	7,055	7,339	7,108	7,229	7,286	7,291
Mediana	6,854	7,003	7,557	7,124	7,231	7,102	7,824	7,394	7,359
Moda	5,050	5,050	-	7,090	7,090	7,090	-	3,401	5,743
<b>Misure statistiche di variabilità</b>									
Deviazione Std	1,410	1,443	1,418	1,322	1,458	1,432	1,531	1,287	1,412
Varianza	1,988	2,083	2,011	1,747	2,126	2,050	2,345	1,657	1,994
Intervallo	12,074	14,865	5,282	10,305	12,738	13,899	6,942	7,144	16,047
Intervallo interquartile	1,768	1,805	1,647	1,667	1,789	1,750	1,714	1,652	1,777
<b>Momenti</b>									
Skewness	0,319	0,257	0,274	-0,337	0,062	0,114	-1,510	-0,397	-0,156
Kurtosis	1,241	0,985	0,100	0,518	0,307	0,834	2,825	0,429	0,503
Coeff di var	20,779	20,762	19,179	18,736	19,869	20,144	21,183	17,670	19,368
SS corretta	9115,802	108547,410	40.225	5101,971	21948,812	29366,326	65.652	210,482	1023680,460
Errore std media	0,021	0,006	0,309	0,024	0,014	0,012	0,284	0,114	0,002
N	4.587	52.112	21	2.921	10.325	14.324	29	128	513.294

Note: statistiche calcolate sui valori espressi in termini logaritmici.

**Prospetto 5.2 - Test di uguaglianza delle medie per gli importi Anf stimati da modello 770 e Dm/10 (statistiche)**

Sez Ateco	Var def	Mean			Std Dev			Std Err	N
		lower CL	upper CL	current	lower CL	upper CL	current		
A	Totale ANF Dm10	1.652,90	2.484,40	2.068,70	8.711,60	9.299,90	8.996,00	212,00	1801
	Totale ANF INPS	1.857,70	2.569,10	2.213,40	7.453,50	7.956,90	7.696,90	181,40	1801
	Diff (Dm10-INPS)	-691,70	402,30	-144,70	8.182,70	8.569,70	8.371,70	279,00	
B	Totale ANF Dm10	6.064,30	7.420,10	6.742,20	17.206,00	18.165,00	17.672,00	345,70	2613
	Totale ANF INPS	7.312,70	9.041,90	8.177,30	21.944,00	23.167,00	22.539,00	440,90	2613
	Diff (Dm10-INPS)	-2.534,00	-336,70	-1.435,00	19.871,00	20.648,00	20.252,00	560,30	
C	Totale ANF Dm10	4.859,10	6.248,90	5.554,00	16.572,00	17.555,00	17.050,00	354,40	2315
	Totale ANF INPS	5.244,10	6.395,50	5.819,80	13.730,00	14.545,00	14.126,00	293,60	2315
	Diff (Dm10-INPS)	-1.168,00	636,40	-265,80	15.344,00	15.982,00	15.656,00	460,20	
D	Totale ANF Dm10	5.079,70	5.423,50	5.251,60	34.409,00	34.652,00	34.530,00	87,70	155000
	Totale ANF INPS	5.413,90	5.781,90	5.597,90	36.830,00	37.090,00	36.959,00	93,90	155000
	Diff (Dm10-INPS)	-598,10	-94,50	-346,30	35.676,00	35.854,00	35.765,00	128,50	
E	Totale ANF Dm10	6.596,50	21.374,00	13.985,00	116.074,00	126.534,00	121.077,00	3.765,30	1034
	Totale ANF INPS	7.737,50	23.703,00	15.720,00	125.409,00	136.710,00	130.815,00	4.068,10	1034
	Diff (Dm10-INPS)	-12.606,00	9.135,80	-1.735,00	122.312,00	130.004,00	126.040,00	5.543,20	
F	Totale ANF Dm10	3.767,00	3.895,50	3.831,20	10.367,00	10.458,00	10.412,00	32,80	101000
	Totale ANF INPS	4.178,70	4.297,10	4.237,90	9.548,40	9.632,10	9.590,10	30,20	101000
	Diff (Dm10-INPS)	-494,00	-319,30	-406,70	9.978,70	10.040,00	10.010,00	44,60	
G	Totale ANF Dm10	2.528,00	2.675,90	2.602,00	12.174,00	12.278,00	12.226,00	37,70	105000
	Totale ANF INPS	2.778,50	2.924,20	2.851,30	11.996,00	12.099,00	12.047,00	37,20	105000
	Diff (Dm10-INPS)	-353,20	-145,60	-249,40	12.100,00	12.174,00	12.137,00	53,00	
H	Totale ANF Dm10	2.970,90	3.402,60	3.186,80	19.398,00	19.703,00	19.549,00	110,10	31516
	Totale ANF INPS	3.147,50	3.562,30	3.354,90	18.639,00	18.933,00	18.785,00	105,80	31516
	Diff (Dm10-INPS)	-467,50	131,20	-168,20	19.066,00	19.277,00	19.171,00	152,70	
I	Totale ANF Dm10	5.523,70	6.610,70	6.067,20	46.502,00	47.270,00	46.883,00	277,30	28588
	Totale ANF INPS	6.232,50	7.132,30	6.682,40	38.494,00	39.130,00	38.810,00	229,50	28588
	Diff (Dm10-INPS)	-1.321,00	90,40	-615,10	42.788,00	43.287,00	43.036,00	360,00	
J	Totale ANF Dm10	2.971,10	4.188,50	3.579,80	20.607,00	21.468,00	21.029,00	310,50	4587
	Totale ANF INPS	3.407,20	5.026,80	4.217,00	27.415,00	28.561,00	27.976,00	413,10	4587
	Diff (Dm10-INPS)	-1.650,00	375,80	-637,20	24.395,00	25.111,00	24.748,00	516,80	
K	Totale ANF Dm10	3.979,90	4.496,40	4.238,20	29.892,00	30.257,00	30.073,00	131,70	52112
	Totale ANF INPS	4.358,40	4.869,70	4.614,00	29.593,00	29.955,00	29.773,00	130,40	52112
	Diff (Dm10-INPS)	-739,20	-12,60	-375,90	29.796,00	30.053,00	29.924,00	185,40	
L	Totale ANF Dm10	623,80	8.288,70	4.456,20	6.441,30	12.158,00	8.419,30	1.837,20	21
	Totale ANF INPS	1.049,30	2.498,60	1.773,90	1.218,00	2.299,00	1.592,00	347,40	21
	Diff (Dm10-INPS)	-1.097,00	6.461,30	2.682,30	4.974,40	7.752,30	6.058,80	1.869,80	
M	Totale ANF Dm10	2.405,80	2.801,90	2.603,80	5.322,40	5.602,60	5.458,90	101,00	2921
	Totale ANF INPS	2.722,30	3.219,90	2.971,10	6.686,50	7.038,50	6.858,00	126,90	2921
	Diff (Dm10-INPS)	-685,20	-49,30	-367,30	6.087,60	6.312,50	6.198,00	162,20	
N	Totale ANF Dm10	4.430,80	4.983,10	4.706,90	14.122,00	14.513,00	14.315,00	140,90	10325
	Totale ANF INPS	4.831,00	5.435,30	5.133,10	15.451,00	15.879,00	15.662,00	154,10	10325
	Diff (Dm10-INPS)	-835,50	-16,90	-426,20	14.860,00	15.150,00	15.004,00	208,80	
O	Totale ANF Dm10	3.931,10	4.604,30	4.267,70	20.316,00	20.792,00	20.551,00	171,70	14324
	Totale ANF INPS	4.329,60	5.033,60	4.681,60	21.246,00	21.744,00	21.492,00	179,60	14324
	Diff (Dm10-INPS)	-900,90	73,10	-413,90	20.856,00	21.201,00	21.027,00	248,50	
Q	Totale ANF Dm10	1.717,90	3.613,30	2.665,60	1.977,20	3.369,60	2.491,50	462,70	29
	Totale ANF INPS	2.117,90	4.314,10	3.216,00	2.290,90	3.904,30	2.886,80	536,10	29
	Diff (Dm10-INPS)	-1.969,00	868,10	-550,40	2.276,40	3.307,80	2.696,40	708,10	
Non rip	Totale ANF Dm10	2.206,50	3.798,80	3.002,70	4.054,50	5.190,00	4.552,10	402,40	128
	Totale ANF INPS	2.850,50	5.024,90	3.937,70	5.536,40	7.087,00	6.215,90	549,40	128
	Diff (Dm10-INPS)	-2.276,00	406,00	-935,10	5.012,50	5.966,70	5.447,90	681,00	
Totale	Totale ANF Dm10	4.119,70	4.264,80	4.192,30	26.479,00	26.582,00	26.530,00	37,03	513000
	Totale ANF INPS	4.447,70	4.624,60	4.551,20	26.796,00	26.900,00	26.847,00	37,47	513000
	Diff (Dm10-INPS)	-462,20	-255,60	-358,90	26.653,00	26.726,00	26.689,00	52,68	

Note: statistiche calcolate sui valori non trasformati.

**Prospetto 5.3 - Test di locazione:  $\mu=0$  e med=0 (modd. 770)**

sezioni	t-statistic	Pr >  t	dei segni	Pr >  M	Wilcoxon	Pr >  S
A	249,003	<,0001	900,5	<,0001	>10.000	<,0001
B	314,709	<,0001	1.306,5	<,0001	>10.000	<,0001
C	290,684	<,0001	1.157,5	<,0001	>10.000	<,0001
D	2.152,056	<,0001	>10.000	<,0001	>10.000	<,0001
E	155,190	<,0001	517	<,0001	>10.000	<,0001
F	1.823,571	<,0001	50.473	<,0001	>10.000	<,0001
G	1.919,885	<,0001	52.494,5	<,0001	>10.000	<,0001
H	1.072,853	<,0001	15.757	<,0001	>10.000	<,0001
I	755,261	<,0001	9.551	<,0001	>10.000	<,0001
J	342,754	<,0001	2.292,5	<,0001	>10.000	<,0001
K	1.175,013	<,0001	26.053	<,0001	>10.000	<,0001
L	27,172	<,0001	10,5	<,0001	115,5	<,0001
M	314,997	<,0001	1.460,5	<,0001	>10.000	<,0001
N	545,485	<,0001	5.162,5	<,0001	>10.000	<,0001
O	643,230	<,0001	7.162	<,0001	>10.000	<,0001
Q	30,534	<,0001	14,5	0,0002	217,5	0,0002
Non rip	71,304	<,0001	64	<,0001	4.128	<,0001
<b>Totale</b>	<b>3.951,59</b>	<b>&lt;,0001</b>	<b>&gt;10.000</b>	<b>&lt;,0001</b>	<b>&gt;10.000</b>	<b>&lt;,0001</b>

Note: statistiche calcolate sui valori espressi in termini logaritmici.

**Prospetto 5.4 - Test di locazione:  $\mu=0$  e med=0 (Dm10)**

sezioni	t-statistic	Pr >  t	dei segni	Pr >  M	Wilcoxon	Pr >  S
A	230,087	<,0001	900,5	<,0001	>10.000	<,0001
B	305,599	<,0001	1.306,5	<,0001	>10.000	<,0001
C	283,460	<,0001	1.157,5	<,0001	>10.000	<,0001
D	2.043,326	<,0001	>10.000	<,0001	>10.000	<,0001
E	153,091	<,0001	517	<,0001	>10.000	<,0001
F	1.698,242	<,0001	49.039	<,0001	>10.000	<,0001
G	1.771,166	<,0001	>10.000	<,0001	>10.000	<,0001
H	994,851	<,0001	>10.000	<,0001	>10.000	<,0001
I	837,978	<,0001	>10.000	<,0001	>10.000	<,0001
J	325,934	<,0001	2.293,5	<,0001	>10.000	<,0001
K	1.099,485	<,0001	>10.000	<,0001	>10.000	<,0001
L	23,893	<,0001	10,5	<,0001	115,5	<,0001
M	288,456	<,0001	1.460,5	<,0001	>10.000	<,0001
N	511,413	<,0001	5.162,5	<,0001	>10.000	<,0001
O	594,122	<,0001	7.161,5	<,0001	>10.000	<,0001
Q	25,422	<,0001	14,5	<,0001	217,5	<,0001
Non rip	64,027	<,0001	64	<,0001	4.128	<,0001
<b>Totale</b>	<b>3.699,13</b>	<b>&lt;,0001</b>	<b>&gt;10.000</b>	<b>&lt;,0001</b>	<b>&gt;10.000</b>	<b>&lt;,0001</b>

Note: statistiche calcolate sui valori espressi in termini logaritmici.

**Prospetto 5.5 - Test di uguaglianza delle medie per gli importi Anf stimati da modello 770 e Dm10 (test di omoschedasticità)**

Sez Ateco	DF (explanatory)	DF (error)	F- stat (1)	Pr > F	F- stat (2)	Pr > F
A	1	3.600	0,06	0,80	0,05	0,83
B	1	5.224	1,45	0,23	4,02	0,05
C	1	4.628	0,13	0,72	0,11	0,74
D	1	310.080	0,04	0,85	3,36	0,07
E	1	2.066	0,02	0,90	0,08	0,78
F	1	201.894	0,51	0,48	29,12	<,0001
G	1	209.980	0,00	0,95	6,06	0,01
H	1	63.030	0,03	0,87	0,03	0,86
I	1	57.174	0,14	0,71	1,69	0,19
J	1	9.172	1,36	0,24	1,19	0,28
K	1	104.222	0,00	0,95	2,04	0,15
L	1	40	2,15	0,15	2,08	0,16
M	1	5.840	1,12	0,29	2,99	0,08
N	1	20.648	0,22	0,64	1,96	0,16
O	1	28.646	0,06	0,80	1,14	0,29
Q	1	56	0,36	0,55	0,33	0,57
Non rip	1	254	0,89	0,35	1,24	0,27
<b>Totale</b>	<b>1</b>	<b>1.026.586</b>	<b>0,00</b>	<b>0,95</b>	<b>19,57</b>	<b>&lt;0,001</b>

Note: statistiche calcolate sui valori non trasformati

(1) Levene's test of Squared Deviations

(2) Brown and Forsythe's Test of Absolute Deviations from medians

**Prospetto 5.6 - Test di uguaglianza delle medie per gli importi degli Anf stimati da modello 770 e Dm10 (zero mean differences)**

Sez Ateco	Method	DF	t - statistic	Pr >  t
A	Pooled	3.600	-0,52	0,60
B	Pooled	5.224	-2,56	0,01
C	Pooled	4.628	-0,58	0,56
D	Pooled	310.000	-2,7	0,01
E	Pooled	2.066	-0,31	0,75
F	Satterthwaite	200.000	-9,13	<,001
G	Satterthwaite	210.000	-4,71	<,001
H	Pooled	63.000	-1,1	0,27
I	Pooled	57.000	-1,71	0,09
J	Pooled	9.172	-1,23	0,22
K	Pooled	100.000	-2,03	0,05
L	Pooled	40	1,43	0,16
M	Pooled	5.840	-2,26	0,02
N	Pooled	21.000	-2,04	0,05
O	Pooled	29.000	-1,67	0,10
Q	Pooled	56	-0,78	0,44
Non rip	Pooled	254	-1,37	0,17
<b>Totale</b>	<b>Satterthwaite</b>	<b>-</b>	<b>-6,81</b>	<b>&lt;.0001</b>

Note: statistiche calcolate sui valori non trasformati.

## Capitolo 6 - Le prospettive future del progetto Tmnp: l'introduzione del modello Emens

### 6.1 L'introduzione del modello *Emens*

Come già accennato nel paragrafo 1.1.2 del capitolo 1, le possibilità di produrre statistiche disaggregate sulle diverse tipologie di Tmnp subiscono un notevole incremento con l'introduzione del modello *Emens*. Secondo quanto stabilito dal DL 269/2003, art. 44, comma 9, convertito in legge n.326/2003 «[...] a partire dalle retribuzioni corrisposte con riferimento al mese di gennaio 2005, i sostituti di imposta tenuti al rilascio della certificazione di cui all'articolo 7-bis del decreto del Presidente della Repubblica 29 settembre 1973, n. 600, trasmettono mensilmente in via telematica [...] all'Istituto Nazionale della Previdenza Sociale (Inps) i dati retributivi e le informazioni necessarie per il calcolo dei contributi, per l'implementazione delle posizioni assicurative individuali e per l'erogazione delle prestazioni, entro l'ultimo giorno del mese successivo a quello di riferimento. Tale disposizione si applica anche nei confronti dell'Istituto Nazionale di Previdenza per i Dipendenti dell'Amministrazione Pubblica (Inpdap) [...]». A tale riguardo la Circolare Inps n.152 del 22 novembre 2004 stabilisce che a partire dal 1° gennaio 2005 i datori di lavoro «già tenuti alla compilazione della parte C, dati previdenziali ed assistenziali Inps del modello 770 semplificato», sono tenuti ad «inviare mensilmente i dati retributivi e le informazioni necessarie per il calcolo dei contributi, per le implementazioni delle posizioni assicurative individuali e per l'erogazione delle prestazioni».

Nei paragrafi successivi, oltre ad illustrare il contenuto informativo del modello *Emens*, e a confrontarlo con quello del modello 770, si cercherà di formulare ipotesi sui possibili effetti che l'introduzione della mensilizzazione delle denunce retributive, con la conseguente semplificazione della parte C del modello 770, potrà avere nella produzione di informazioni statistiche sui Tmnp.

Sin da subito è però possibile sottolineare che l'introduzione del modello *Emens* comporterà modifiche decisive su alcune delle sezioni del modello 770 utilizzate nell'ambito del progetto Tmnp. Questo renderà l'impiego del modello *Emens* non una scelta alternativa rispetto a quella del modello 770, bensì l'unica scelta percorribile. Coerentemente con questo, nel Psn 2007-2009 è stato introdotto «[...] un nuovo studio progettuale *Utilizzo a fini statistici della mensilizzazione delle denunce retributive dei lavoratori dipendenti a titolarità Istat in collaborazione con l'Inps* [...]». L'obiettivo dello studio è di verificare l'utilizzabilità a fini statistici degli archivi delle dichiarazioni mensili. »

Comunque l'analisi dei contenuti dei due modelli finora svolta, consente già di ipotizzare che l'*Emens* diventerà la fonte informativa principale per la produzione di statistiche su tutte le tipologie di Tmnp erogati ai lavoratori dipendenti del settore privato, con la sola esclusione della prestazione Tfr, per la cui rilevazione sarà ancora necessario l'utilizzo del 770.

Dal momento che non è stato ancora avviato uno studio sugli archivi informatici dei modelli *Emens*, ad oggi non è possibile definire con ragionevole accuratezza le risorse umane e tecnologiche necessarie per utilizzare al meglio gli archivi in questione. Tuttavia, l'esperienza già acquisita con i modelli 770, unita alla circostanza che si passerà da una fonte annuale ad una fonte mensile, permette di prevedere, sin da adesso, la necessità di un notevole sforzo tecnologico e conoscitivo da parte dell'istituto, a cui andrà unita la capacità di instaurare rapporti stabili e fruttuosi con gli Enti che forniranno i dati (Inps).

È altresì fondamentale sottolineare che malgrado il notevole incremento informativo garantito dall'introduzione del modello *Emens*, non sembra che questo possa risolvere un problema assai rilevante già emerso con il modello 770: gli unici trattamenti monetari non pensionistici rilevabili dal modello *Emens* sono quelli a cui si associa una riduzione della retribuzione, in corrispondenza della quale ai lavoratori vengono accreditati contributi figurativi. Quindi, come già evidenziato per il modello 770, eventi come la malattia inferiore ai 7 giorni, i congedi matrimoniali ed altri ai quali non corrisponde una riduzione della retribuzione percepita, non potranno essere rilevati tramite il modello *Emens*, così come non potevano esserlo tramite il modello 770.

Ulteriore aspetto meritevole d'attenzione è il rapporto tra modello *Emens* e modello Dm10.

Relativamente alle prestazioni monetarie non pensionistiche, il modello Dm10 contiene dati aggregati a livello aziendale (matricola aziendale) che nell'ambito del progetto d'indagine sui Tmnp vengono utilizzati per valutare la coerenza di stime che l'Istituto realizza, sul medesimo fenomeno, a partire da fonti

amministrative diverse dal modello 770<sup>53</sup>.

Anche da questo punto di vista l'introduzione dell'*Emens* rappresenta, almeno nelle intenzioni di chi l'ha ideato, un passo in avanti. Infatti a differenza di quanto avveniva per il modello 770, per il quale la quadratura con il modello Dm10 non era un 'prerequisito' logico né un principio istituzionalmente sancito, il modello *Emens*, nasce sin dall'origine come un modello del quale, in ogni esercizio, il consolidamento viene realizzato tenendo conto delle quadrature effettuate con il modello Dm10. Quindi non più una verifica ex-post, come avveniva tra modello 770 e modello Dm10, ma un parallelismo insito nei fondamenti stessi del modello *Emens*. Questo principio è esplicitamente formulato in documenti ufficiali dell'Inps. Dal punto di vista generale viene innanzitutto sottolineato come, con l'adozione dell'*Emens*, «...si enfatizza l'allineamento tra i cedolini paga ed i flussi trasmissivi dei modelli Dm10 e quelli delle nuove denunce mensili»<sup>54</sup>. Scendendo nel dettaglio, nella spiegazione della gestione delle denunce di rettifica viene specificato che «...prima di provvedere all'eliminazione della prima denuncia ed all'inserimento della seconda, l'Inps provvederà ad espletare i controlli di congruità con i flussi del Dm10 [...]»<sup>55</sup>. Così pure, nella descrizione della sezione *Mese Precedente*, viene specificato che tale sezione non potrà essere utilizzata per modificare i dati sulla qualifica, il profilo contributivo e l'imponibile del lavoratore «...in quanto elementi che garantiscono la quadratura con il modello Dm10»<sup>56</sup>.

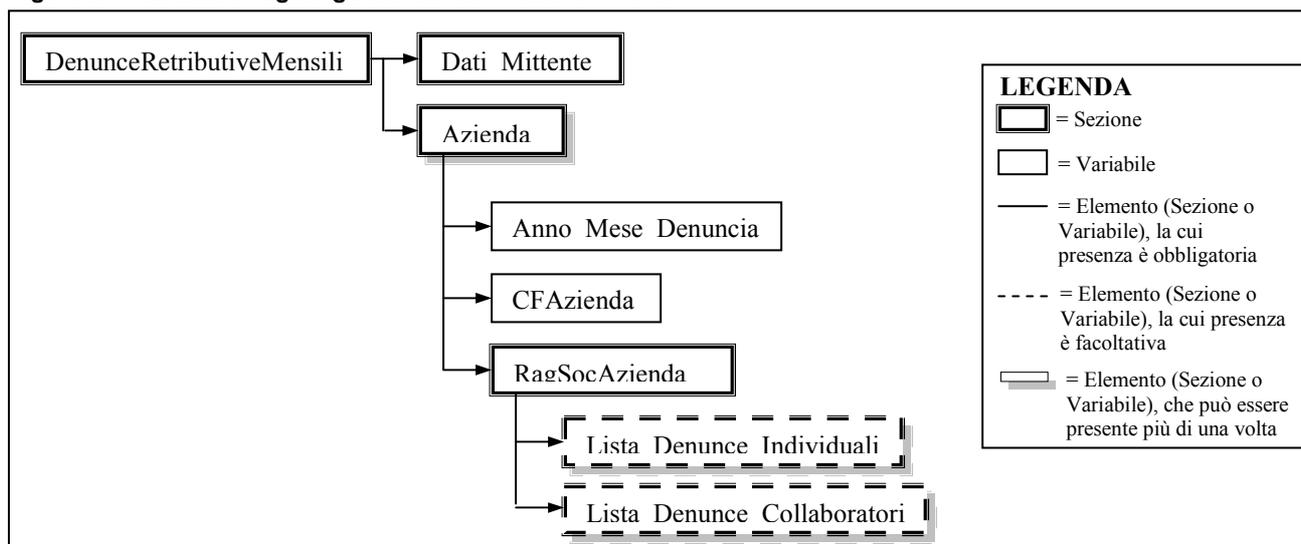
## 6.2 Il modello *Emens*: struttura generale

Dal punto di vista gerarchico-logico l'informazione contenuta nel modello *Emens* è organizzata come descritto nel figura 6.1. L'elemento 'Denunce\_Retributive\_Mensili' è l'elemento<sup>57</sup> radice, cioè l'elemento che, dal punto di vista gerarchico, domina l'intero flusso informativo e che permette di identificare ed instradare il flusso in questione come flusso informativo *Emens*.

Al suo interno questo elemento è articolato in due sottoelementi, entrambi obbligatoriamente presenti:

- 'Dati\_Mittente', destinato a raccogliere alcune informazioni sul soggetto che, essendone abilitato, ha trasmesso i dati;
- 'Azienda', contenente informazioni sul datore di lavoro e sul lavoratore, sia dipendente che collaboratore, cioè sulle due unità di analisi sulla base delle quali si intende indagare il fenomeno Tmpn. È quindi questa la sezione fondamentale ai fini del progett in questione.

Figura 6.1 - Struttura logico-gerarchica del modello *Emens*



<sup>53</sup> Cfr. cap.3, par. 3.2; pag. 37.

<sup>54</sup> Circ. Inps n.152/2004.

<sup>55</sup> Cfr *Documento tecnico per la compilazione dei flussi delle denunce retributive mensili*, Inps - Direzione Centrale Sistemi Informativi e Telecomunicazione, 2006, pag 8.

<sup>56</sup> Cfr *Documento tecnico per la compilazione dei flussi delle denunce retributive mensili*, Inps - Direzione Centrale Sistemi Informativi e Telecomunicazione, 2006, pag 54.

<sup>57</sup> I termini *elemento* e *sottoelemento*, utilizzati nei documenti ufficiali dell'Inps, verranno qui usati con lo stesso significato dei termini *sezione* e *sottosezione*, dai quali a volte verranno sostituiti.

### *Criteria di univocità delle dichiarazioni*

Le varie sezioni e sottosezioni del modello sono individuate da diversi criteri di univocità (chiavi primarie) elencati nel figura 6.2, in relazione gerarchica tra loro. Questo permette, in caso sia necessario rettificare dati già inviati, di intervenire su una specifica sottosezione del modello, senza che sia necessario inviarlo nuovamente nella sua interezza. Questo aspetto verrà approfondito nel successivo paragrafo 3.3.

**Figura 6.2 - Criteri di Univocità (Chiavi primarie) del modello *EMens***

<b>Sezione</b>	<b>Variabili – Criteri di Univocità</b>
Azienda	Anno Mese Denuncia, CFAzienda
Lista Denunce Individuali	Matricola Aziendale
Denuncia Individuale	CFLavoratore, Qualifica1, Qualifica2, Qualifica3, Tipo di contribuzione
Lista collaboratori	Univoca nell'ambito dell'azienda, quindi identificata da CFAzienda
Collaboratore	CFCollaboratore, Tipo Rapporto, Aliquota

### *Invio del modello*

L'invio del modello *Emens* deve avvenire, esclusivamente per via telematica, entro l'ultimo giorno del mese successivo a quello di competenza o, qualora questo sia un giorno festivo, entro il primo giorno lavorativo del mese successivo. L'invio può essere concentrato in un unico flusso o frazionato, anche nel caso di singola azienda o singola posizione contributiva.

L'elemento base della dichiarazione *Emens* rimane quindi la denuncia individuale: l'aggregazione o segmentazione di tali denunce è libera<sup>58</sup>.

Quanto invece alle dichiarazioni rettificative o di regolarizzazione, l'Inps non ha previsto modalità di invio specifiche. Questo aspetto verrà approfondito nel paragrafo 6.3.

## **6.3 Il contenuto informativo del modello *Emens***

Per meglio comprendere le innovazioni introdotte dal modello *Emens*, è opportuno analizzare in maniera più approfondita il contenuto informativo dello stesso.

### *La sezione Azienda*

La sezione Azienda contiene le variabili 'Anno\_Mese\_Denuncia', 'CFAzienda' e 'RagioneSocialeAzienda' e le sottosezioni 'Lista\_Denunce\_Individuali', contenente le denunce individuali dei singoli lavoratori dipendenti e 'Lista\_Collaboratori', contenente le denunce individuali dei lavoratori parasubordinati. L'anno ed il mese della denuncia, il codice fiscale e la ragione sociale dell'azienda sono variabili obbligatorie (essendo, tra l'altro, criteri di univocità della sezione). Quanto invece alla 'Lista\_Denunce\_Individuali' e 'Lista\_Collaboratori', nella denuncia almeno una delle due deve essere presente.

### *La sezione Azienda/Lista Denunce Individuali*

Prima di tutto occorre specificare che si parla di 'Lista' di denunce perché in questa sottosezione del modello *Emens* sono raggruppate le denunce individuali dei lavoratori dipendenti, quelle cioè che riportano dati contributivi e retributivi riferiti al singolo lavoratore.

Le denunce individuali sono suddivise in base alla Matricola Aziendale Inps, un codice di dieci caratteri numerici che deve essere obbligatoriamente presente e che sintetizza il profilo contributivo che caratterizza il rapporto del datore di lavoro con l'Inps. Ad ogni azienda, a cui corrisponde un unico codice fiscale aziendale, possono corrispondere una o più matricole aziendali, ad esempio nel caso di plurilocalizzazione degli impianti. Quando questo avviene, nella denuncia *Emens* saranno presenti più sezioni 'Lista\_Denunce\_Individuali', ognuna individuata da una specifica Matricola Aziendale.

<sup>58</sup> Sul punto cfr. Messaggio Inps n.16329 del 22 aprile 2005: «relativamente alla struttura del flusso si ribadisce che lo stesso può essere composto con la massima flessibilità: possono essere separati i flussi tra lavoratori dipendenti e collaboratori; possono essere separati i flussi relativi a matricole diverse della stessa azienda; possono essere separati i flussi di taluni lavoratori anche nell'ambito della stessa matricola aziendale».

La sezione ‘Caratteristiche contributive’ è facoltativa e contiene due codici, entrambi assenti nel modello 770: il Csc (Codice Statistico Contributivo) ed il Ca (Codice Autorizzazione), che consentono un’ulteriore analisi del ruolo che il datore di lavoro ha nei confronti dell’Inps<sup>59</sup>.

*La sezione Azienda/Lista Denunce Individuali/Denuncia Individuale*

Venendo al contenuto di una singola denuncia individuale (Figura 6.3), è in questa sezione che si concentrano le informazioni fondamentali per ricostruire le caratteristiche del lavoratore, percettore di Tmnp, e per calcolare il valore della prestazione ricevuta.

Ai fini del progetto d’indagine sui Tmnp di particolare rilievo sono, oltre ai dati anagrafici, quelli contenuti nella sezione *Dati Retributivi* e nelle sezioni su *Anf, Tfr e Cig*.

*La sezione Azienda/Lista Denunce Individuali/Denuncia Individuale/Dati retributivi*

Nella sezione destinata alla raccolta dei dati retributivi sono riportate informazioni sul trattamento economico ricevuto dal lavoratore, con un dettaglio nettamente maggiore di quello presente nel modello 770, soprattutto con riferimento alla possibilità di individuare con elevata specificità la tipologia di trattamento monetario non pensionistico ricevuto.

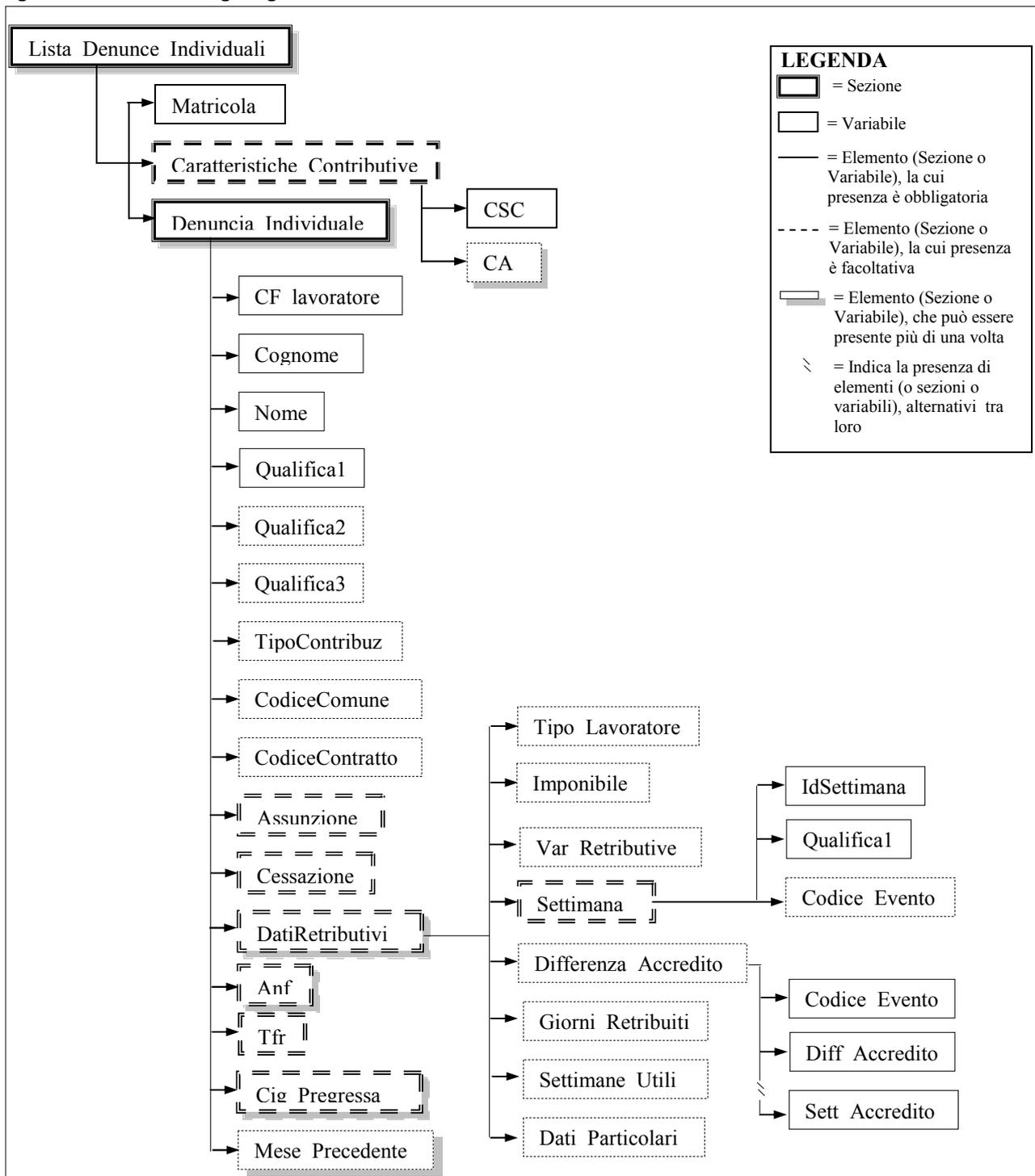
Questo maggiore contenuto informativo risulta evidente esplodendo la sezione Settimana. In questa sezione deve essere indicato il numero progressivo della settimana di calendario (da 1 a 54), il tipo di copertura retributiva goduta dal lavoratore (totale, parziale, nulla o assente...) ed il codice dell’evento associato a settimane di copertura parziale o nulla, che consente di distinguere le diverse tipologie di prestazioni. Come evidenziato nella figura 6.4, nel modello 770 l’informazione relativa ad alcune prestazioni veniva resa in maniera aggregata (ad esempio non è possibile distinguere tra le prestazioni per malattia e quelle per infortuni, come pure tra quelle per maternità e quelle per congedi parentali), sebbene le singole prestazioni dessero luogo a trattamento retributivi molto diversi fra loro, rendendo così impossibile una stima degli importi singolarmente corrisposti. Nel modello *Emens* è invece presente un maggior dettaglio classificatorio delle diverse prestazioni, al quale si aggiunge il portato informativo della variabile *IdSettimana* che consente di individuare con massima precisione la settimana in cui l’evento Tmnp si è verificato.

Già questo consente di affermare con certezza che, per quanto concerne l’individuazione del tipo di prestazione non pensionistica goduta (il ‘quale?’) ed il periodo in cui questa sia stata erogata (il ‘quando?’), il passaggio dal modello 770 al modello *Emens* garantisce un progresso certo.

---

<sup>59</sup> Il codice Csc è stato utile nella procedura di stima degli Anf, in quanto ha consentito di ripulire l’archivio 770 dalle imprese pubbliche, estranee all’universo su cui la statistica deve essere prodotta. Questo è stato però possibile solo grazie all’utilizzo dell’archivio 770 Inps, in quanto nell’archivio originale Sogei, che non riporta variabili ulteriori rispetto a quelle del modello cartaceo, il Csc non è presente. Quindi, in sintesi, il codice Csc è un’informazione utile ai fini del progetto sui Tmnp, che grazie all’introduzione dell’*Emens* sarà immediatamente disponibile nel modello originario, a differenza di quanto avveniva nel modello 770.

Figura 6.3 - Struttura logico-gerarchica della sezione *Denunce Individuali*



Quanto invece alla determinazione degli importi (il ‘quanto?’), risultano di centrale rilievo le informazioni contenute nella sottosezione ‘Differenza Accredito’. Al riguardo va immediatamente sottolineata la differenza concettuale tra modello 770 e modello *Emens* rispetto agli importi relativi alle prestazioni Tmnp (figura 6.4). Nel modello 770 venivano indicati gli importi corrisposti a titolo di retribuzione ridotta, cioè importi riconosciuti in corrispondenza di eventi come malattia, infortuni, maternità, ecc., in coincidenza dei quali il lavoratore non percepisce l’intera retribuzione prevista dal contratto, ma solo una parte della stessa. Nel modello *Emens* la prospettiva è invece opposta o, meglio, complementare. Non si indica la retribuzione ridotta, ma la riduzione subita rispetto alla retribuzione spettante per una giornata lavorativa ‘standard’. Questa informazione, se correttamente compilata, incrociata con la variabile Codice\_Evento e con quella

IdSettimana, potrebbe permettere, se sottratta all'ammontare totale spettante settimanalmente da contratto collettivo, di determinare con estrema precisione gli importi delle prestazioni.

**Figura 6.4 - classificazione dei Tmpn nel modello Emens e nel modello 770**

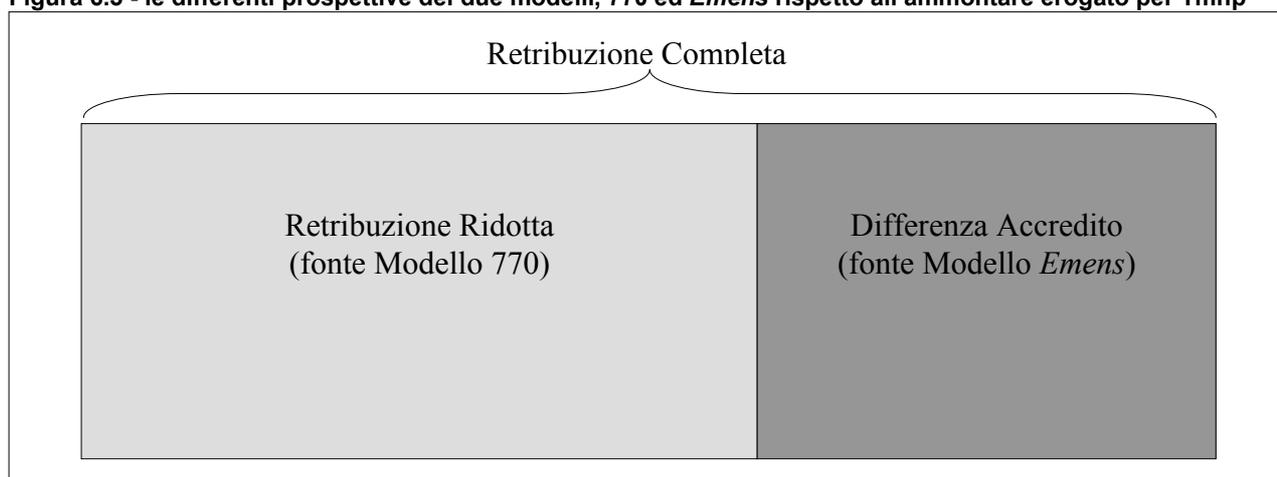
Dettaglio nel Modello Emens		Dettaglio nel Modello 770
Codice	Descrizione	Descrizione
MAL	Malattia per eventi di durata non inferiore ai sette giorni	Malattia e/o infortuni per eventi di durata non inferiore ai sette giorni
INF	Infortunio per eventi di durata non inferiore ai sette giorni	
MA1	Congedo di maternità e paternità ex artt. 16, 17, 20 e 28, D.Lgs.n.151/2001	Maternità e Congedi Parentali D.Lgs. 151/2001
MA2	Periodi di congedo parentale disciplinati dall'art.35, comma 1, D.Lgs. 151/2001 (6 mesi entro i tre anni di vita del bambino)	
MA3	Periodi di congedo per malattia del bambino di età inferiore a 3 anni, disciplinati dall'art.49, comma 1, D.Lgs. 151/2001	
MA4	Prolungamento del congedo parentale fino a 3 anni di vita del bambino con handicap - art. 33, comma 1, D.Lgs. 151/2001 (art. 33, comma 1, L.104/1992)	
MA5	Permessi mensili per figli con handicap gravi - art.42, commi 2 e 3, D.Lgs. 151/2001 (art. 33, comma 3, L.104/1992)	
MA6	Permessi mensili per lavoratore con handicap grave - art.33, comma 6, L. 104/1992	
MA7	Permessi mensili per assistere parenti ed affini entro il terzo grado, portatori di handicap grave - art.33, comma 3, L.104/1992	
MB1	Riposi giornalieri fino al primo anno di vita del bambino, ex permessi per allattamento	Congedi Parentali D.Lgs. 151/2001
MB2	Congedi parentali oltre i 6 mesi entro i 3 anni di vita del bambino ovvero fruti entro il 3° e l'8° anno	
MB3	Riposi giornalieri per figli con handicap gravi fino a 3 anni di vita del bambino	
MB4	Congedi per malattia del bambino di età compresa fra i 3 e gli 8 anni, fruibili alternativamente, nel limite di 5 giorni l'anno per ciascun genitore	
MB5	Riposi giornalieri per lavoratore portatore di handicap grave	
MC1	Congedi di cui all'art.42, comma 5, D.Lgs. 151/2001	
ACT	Legge n.88/1997	Legge n.88/1997
CGO	Cassa integrazione Guadagni Ordinaria	Cassa Integrazione Guadagni
CGS	Cassa integrazione Guadagni Straordinaria	
DON	Assenza per donazione di sangue	Assenza per donazione di sangue
SOL	Contratto di solidarietà	Non ancora previsti dalla normativa
DMO	Assenza per donazione di midollo osseo (dal 1/1/2006)	

Il problema sorge però proprio con riferimento alla determinazione della retribuzione giornaliera totale, informazione necessaria per ottenere, sottraendole la 'Differenza Accredito', l'ammontare del trattamento monetario non pensionistico erogato. L'individuazione di tale ammontare presuppone uno studio piuttosto approfondito dei contratti collettivi, in quanto la retribuzione giornaliera dipenderà non solo dal contratto collettivo di riferimento, ma anche dalla qualifica e, in alcuni casi, dall'anzianità di servizio.

Al riguardo la strada più agevolmente percorribile pare invece quella di prendere in considerazione il dato sul totale imponibile (vedi figura 6.3) in uno dei mesi in cui il lavoratore è stato retribuito pienamente, cioè mesi con codice del 'Tipo di copertura' uguale ad 'X' (copertura totale) per tutte le settimane. Tale valore andrebbe poi diviso per il numero di settimane, per ottenere la retribuzione settimanale piena 'tipo'. Sottraendo a questo valore l'ammontare della 'Differenza Accredito' presente nelle settimane con un codice Tipo di copertura 2 (a qualunque mese appartengano), si otterrebbe il valore della prestazione monetaria non pensionistica erogata<sup>60</sup>.

<sup>60</sup> Sul significato ed il calcolo del valore della *Differenza Accredito* cfr. Messaggio Inps n.16329 del 22 aprile 2005, pagg 14 - 17.

**Figura 6.5 - le differenti prospettive dei due modelli, 770 ed *Emens* rispetto all'ammontare erogato per Tmnp**



Quanto invece al contenuto della sottosezione 'Dati Particolari', ai fini dell'indagine sui Tmnp possono essere di interesse i dati relativi al *Bonus* riconosciuto ai lavoratori che posticipano il pensionamento ai sensi della L.243/2004, e quelli nella sottosezione *Preavviso*, che raccoglie dati sull'indennità sostitutiva del preavviso. Di questi elementi deve però essere preventivamente verificata la possibilità e l'opportunità di una ricomprensione nell'elenco di quelli che vengono classificati come Tmnp<sup>61</sup>. In altri termini, queste erogazioni al lavoratore dipendente non sono state inserite nella classificazione originaria dei Tmnp. La prima perché non ancora esistente, la seconda per una scelta di merito. Oggi sembra opportuno riconsiderare con attenzione la classificazione originaria, per confermarla o, eventualmente, aggiornarla e/o modificarla.

*La sezione Azienda/Lista Denunce Individuali/Denuncia Individuale/Anf*

Uscendo dalla sezione *Dati Retributivi*, ma rimanendo nell'ambito della sezione *Denuncia Individuale*, la sottosezione successiva è la *Anf*. In questa sezione sono contenute le stesse coordinate già presenti nel modello 770: il numero della tabella, il numero dei componenti del nucleo familiare, il codice della classe di reddito. La differenza fondamentale deriva dal fatto che nell'*Emens* questa informazione viene fornita con riferimento mensile. Questo comporta una notevolissima semplificazione delle procedure di stima degli importi degli Anf<sup>62</sup>. Avendo direttamente disponibile il dato sulle coordinate Anf, mese per mese, sarebbero automaticamente risolti i problemi connessi:

1. alla mancata rilevazione dei percettori Anf in corso d'anno, ma non nel mese di dicembre;
2. alla possibile non coincidenza delle coordinate Anf del primo semestre con quelle del secondo semestre;
3. alla mancata indicazione diretta del numero di mesi nei quali gli Anf sono stati percepiti.

Notevoli semplificazioni concettuali deriverebbero anche per la procedura di imputazione delle coordinate Anf scorrette od incomplete<sup>63</sup>, in quanto la circostanza che queste coordinate vengano ripetute, per uno stesso percettore, dodici mesi l'anno (o comunque in ogni mese in cui gli assegni siano stati erogati), agevola la ricostruzione del dato mancante attraverso informazioni relative allo stesso percettore, seppur attinte da dichiarazioni relative a mesi diversi rispetto a quelle in cui il dato va imputato. Quindi è prevedibile che in molti casi non sarà più necessario il ricorso a procedure da donatore, procedure che rimarrebbero comunque necessarie nei casi in cui l'informazione sulle coordinate fosse incompleta per l'intero periodo di riferimento.

*La sezione Azienda/Lista Denunce Individuali/Denuncia Individuale/Tfr*

La sezione destinata al Tfr va compilata esclusivamente nel mese di febbraio. Deve esservi indicato l'importo accantonato, maturato al 31/12 dell'anno precedente, limitatamente ai lavoratori in forza nel mese. Allo stato attuale pare quindi che per la produzione di statistiche sul Tfr il modello *Emens* non possa essere utilizzato, in quanto vi viene rilevata solo la quota Tfr maturata nell'anno e non la prestazione effettivamente

<sup>61</sup> Cfr. Paolo Consolini, Rita De Carli, *op.cit.*

<sup>62</sup> Sulla procedura di stima degli importi Anf, cfr. cap. 3.

<sup>63</sup> Sulla procedura di imputazione delle coordinate Anf, cfr. cap. 2, par. 4, pag. 34

erogata alla conclusione del rapporto lavorativo. Limitatamente a questa prestazione rimarrà quindi irrinunciabile l'utilizzo della parte B, sezione *Indennità di fine rapporto*, del modello 770.

*La sezione Azienda/Lista Denunce Individuali/Denuncia Individuale/Cig\_Pregressa*

Proseguendo nell'analisi del contenuto informativo della sezione *Denuncia Individuale* del modello *Emens* (Figura 6.3), ai fini dell'indagine sulle prestazioni monetarie non pensionistiche è altresì rilevante l'analisi della sottosezione *Cig Pregressa*, che viene esplosa nel figura 6.6.

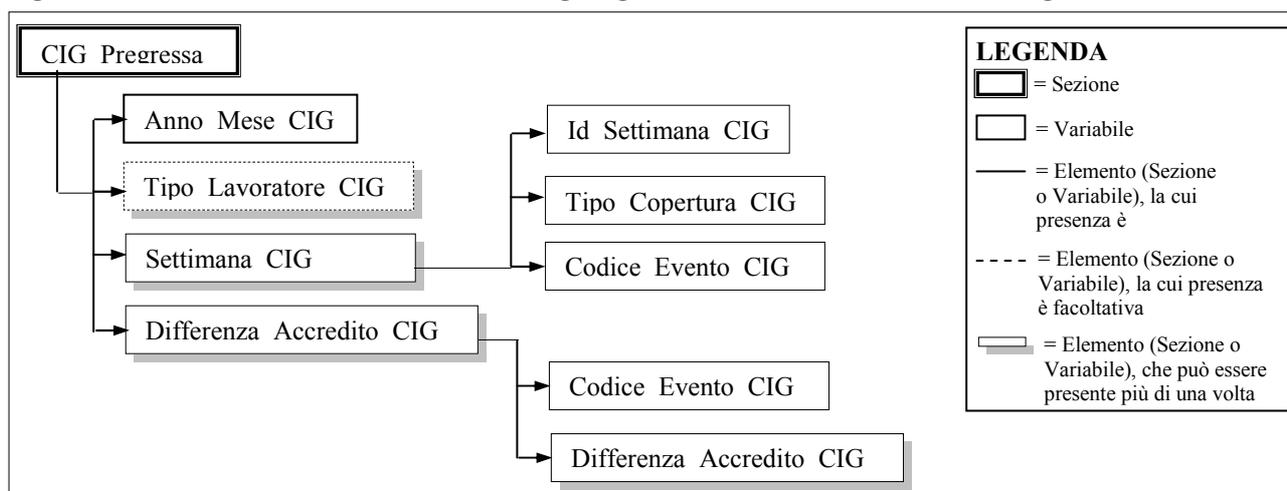
Anche con riferimento alla Cig, il contenuto informativo del modello *Emens* è assolutamente maggiore rispetto a quello del modello 770. In quest'ultimo, infatti, era esclusivamente indicato il numero delle settimane a retribuzione ridotta, o senza alcuna retribuzione, connesse all'evento cassa integrazione (che, tra l'altro, non era ulteriormente distinta in 'ordinaria' e 'straordinaria'). L'eventuale differenza retributiva da accreditare era poi riportata in un'unica voce (la 37)<sup>64</sup>, essa stessa non utile alla ripartizione dell'ammontare nelle diverse tipologie di cassa integrazione (cfr. cap.1, par.1.2.1).

Come è evidente dalla figura 6.6, nel modello *Emens* le informazioni raccolte sulla cassa integrazione sono maggiori. Prima di tutto da un riferimento annuale si passa ad un riferimento settimanale. Inoltre la prestazione è disaggregata nelle tre tipologie di Ordinaria, Straordinaria e di Solidarietà. Quanto all'ammontare percepito, la variabile 'Differenza\_Accredito\_Cig' è assolutamente analoga a quella già presente nel modello 770. In altri termini per questa prestazione non c'è stato, tra i due modelli, il mutamento di prospettiva descritto nella figura 6.5, in quanto il modello 770 già registrava la 'Differenza Accredito'. Le possibilità di riuscire a stimare l'ammontare della Cig sono comunque maggiori, stante le più numerose variabili *di supporto* (ad esempio tutte quella connesse alla rilevazione della retribuzione) presenti nell'*Emens*, variabili che risulteranno probabilmente utili nella procedura di stima che verrà implementata per la rilevazione della Cig.

Quindi, se correttamente compilato, l'*Emens* consentirà di individuare sia l'esatta tipologia di Cassa Integrazione erogata (il 'cosa'), sia la durata della prestazione stessa (il 'quando?'). Per quanto concerne invece l'ammontare (il 'quanto?'), una risposta definitiva si avrà quando sarà approntata una procedura di stima dedicata. Nel modello 770 era invece possibile solo determinare, ed in maniera approssimata, la durata del trattamento.

Per completezza di informazione va specificato che nel modello *Emens*, a differenza di quanto avviene per le altre informazioni, i dati sulla CIG non riguardano il mese di riferimento della dichiarazione (quello indicato nella variabile *Anno\_Mese\_Denuncia*, cfr. figura 6.1) ma mesi precedenti, da cui appunto il nome della sezione, 'CigPregressa'.

**Figura 6.6 - Contenuto informativo e struttura logico-gerarchica della sottosezione CIGPregressa**



<sup>64</sup> Cfr Cap.1, figura 1.3, pag.13.

## 6.4 Le dichiarazioni correttive e rettificative

All'incremento informativo descritto nelle pagine precedenti, che deriva essenzialmente dalla maggiore disaggregazione temporale e classificatoria con la quale sono riportate le informazioni sulle diverse tipologie di Tmnp, si aggiunge quello connesso alle modalità di rinvio di dichiarazioni correttive e rettificative.

In questo ambito è necessario distinguere due casistiche fondamentali:

- devono essere corretti i criteri di univocità della dichiarazione originariamente inviata;
- non devono essere corretti i criteri di univocità della dichiarazione originariamente inviata

In questo secondo caso è sufficiente inviare una nuova denuncia identificata dai criteri di univocità elencati al par. 6.2, coincidenti con quelli della dichiarazioni originaria, che verrà in tal modo eliminata e sostituita dalla denuncia rettificativa.

Qualora invece sia necessario rettificare una denuncia nei suoi criteri di univocità, sarà necessario eliminare la denuncia originale inviandone una seconda che riporti gli stessi dati identificativi di quella da eliminare e valorizzando 'S' nella variabile 'Elimina'. La nuova dichiarazione, con i criteri di univocità corretti, potrà essere inviata contestualmente o meno.

Quanto invece alla tempistica, le dichiarazioni rettificative possono essere prodotte nel medesimo esercizio della dichiarazione originaria o nell'esercizio successivo, a consolidamento annuale già avvenuto.

Nel primo caso la dichiarazione originaria viene immediatamente sostituita con la nuova.

Nel secondo caso l'Inps effettua controlli per verificare la congruità della nuova denuncia con i flussi del Dm10 e con il già effettuato utilizzo contabile/amministrativo della denuncia originaria.

Qualora invece, nell'ambito della stessa azienda, dello stesso lavoratore e nello stesso mese, cambino uno o più elementi identificativi (qualifica e/o tipo di contribuzione), andranno prodotte due distinte denunce individuali.

Nella figura 6.7 sono schematizzate le diverse modalità di rettifica delle dichiarazioni in base alle variabili per le quali si rende necessaria la correzione.

**Figura 6.7 - gestione e tempistica delle dichiarazioni rettificative**

Variazione	Rettifica	Porre l'attributo Elimina='S'	Tempistica e Notazioni
Dei criteri di univocità	Rinvio delle dichiarazione con identici dati identificativi (cd. Criteri di univocità);	Si	La dichiarazione inviata determinerà l'eliminazione dell'originaria denuncia corrispondente. Sarà così possibile l'invio della nuova dichiarazione, caratterizzata dai criteri di univocità corretti. Le variazioni posso essere inviate in qualsiasi momento ma: - la denuncia di variazione inviata nello stesso mese di invio di quella originale sarà considerata integralmente sostituiva di quest'ultima - la denuncia di variazione inviata successivamente opererà in sostituzione solo delle singole sezioni presenti: corrente, mese precedente, CIG pregressa.
	Invio della dichiarazione con i dati identificativi modificati	No	Le due dichiarazioni devono essere inviate nello stesso mese. Questo è quanto avviene quando i criteri di univocità variati nel mese riguardano il singolo lavoratore dipendente (qualifica e/o tipo di contribuzione)
Di dati di dettaglio	Rinvio delle dichiarazione con identici dati identificativi (cd. Criteri di univocità)	No	La tempistica segue le stesse regole esposte per il caso di rinvio di dichiarazioni con identici dati identificativi e Attributo Elimina='S'
	Utilizzo della sottosezione <i>Mese precedente</i> della Denuncia Individuale	No	I dati di dettaglio da rettificare riguardano esclusivamente la Denuncia Individuale, mentre i criteri di univocità rimangono invariati. La rettifica mediante la sezione <i>Mese Precedente</i> deve avvenire nella dichiarazione relativa al mese successivo a quello cui le correzioni di riferiscono.

Sempre allo scopo di rettificare dati inseriti in dichiarazioni precedentemente inviate, esiste poi una sezione del modello specificamente destinata ad accogliere informazioni utili a modificare dati inviati nella sezione *Denuncia Individuale* del mese precedente. In questo caso, quindi, non si tratta di inviare una dichiarazione rettificativa, ma di usare una sezione ad hoc presente in una dichiarazione *Emens* 'ordinaria'. La sezione in questione è la *Mese precedente*, esplosa nella figura 6.8.

Come è evidente le informazioni riportate in questa sezione consentono la correzione/integrazione dei dati contenuti nella sezione *Denuncia Individuale* inviata il mese precedente. L'attributo 'tipo consolidamento'

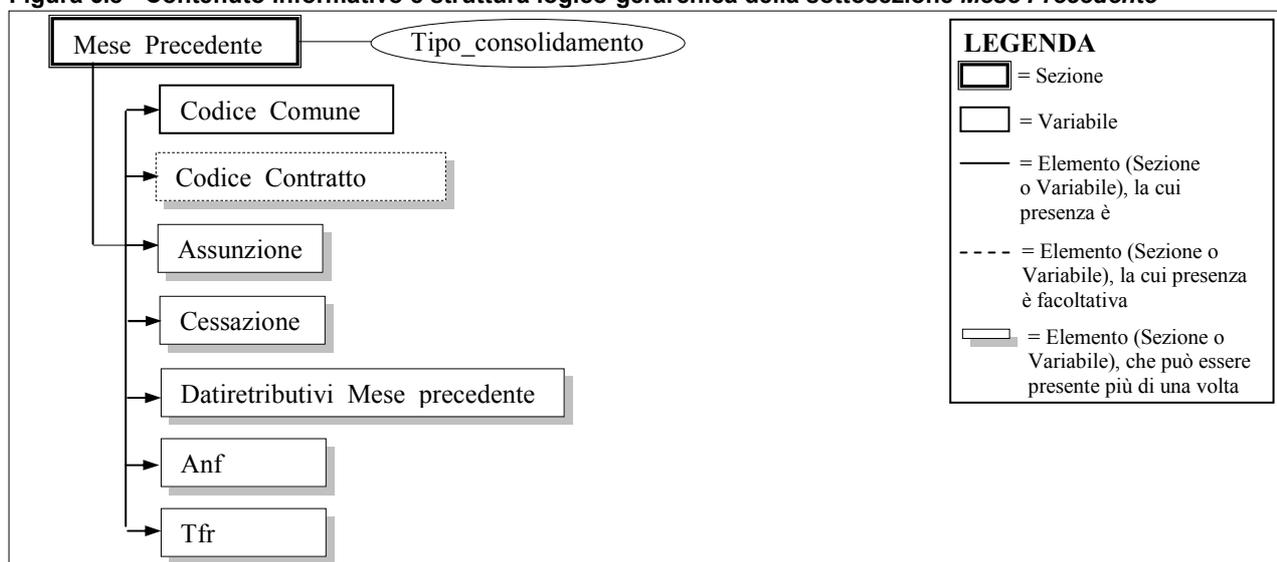
consente poi di specificare se i dati contenuti nella sezione debbano essere considerati come ‘sostituzione completa’ (codice ‘S’) o come integrazione/sostituzione dei soli dati variati.

Rispetto a quanto avveniva nel modello 770 questa procedura di rettifica comporterà dei vantaggi e degli svantaggi. I vantaggi derivano da procedure di rettifica molto tempestive, in quanto i dati vengono corretti entro il mese successivo a quello di riferimento. Lo svantaggio deriva dalla circostanza che la cadenza mensile potrebbe comportare un incremento delle dichiarazioni rettificative, con un conseguente incremento delle problematiche connesse al consolidamento dell’archivio.

Al riguardo vanno poi considerati altri due fattori.

Il primo è che la sezione *Mese Precedente* consente la rettifica di dati contenuti nella sezione *Denuncia Individuale* mentre qualora fosse necessaria la rettifica di dati a più alto livello gerarchico nella struttura informativa del modello *Emens*, ad esempio quelli corrispondenti ai criteri di univocità della denuncia, la procedura da seguire è quella illustrata nella prima parte di questo paragrafo<sup>65</sup>. Quindi, in sintesi, la rettifica può riguardare la denuncia *Emens* nel suo complesso o i soli dati relativi ad un singolo lavoratore dipendente, circostanza che va ulteriormente differenziata a seconda che si debbano correggere dati che rappresentano criteri di univocità (ed allora deve essere inviata una nuova *Denuncia Individuale*) o dati di dettaglio, nel qual caso basterà utilizzare la sezione *Mese Precedente*.

**Figura 6.8 - Contenuto informativo e struttura logico-gerarchica della sottosezione *Mese Precedente***



Da quanto descritto è evidente come la gestione dei dati rettificativi risulti estremamente più articolata. Altro elemento da prendere in considerazione è quello relativo al trasferimento dei dati da parte dell’Inps. Questi potranno essere richiesti ad uno stadio più o meno avanzato di raffinamento. Nel primo caso potrebbero essere già risolti i problemi di consolidamento delle dichiarazioni rettificative. Nel secondo sarebbe invece maggiore la tempestività nella consegna dei dati, dovendo poi essere risolti presso strutture interne dell’Istat i problemi di consolidamento, con i conseguenti investimenti in termini di risorse umane e tecnologiche.

## 6.5 L’introduzione del modello *Emens* e le modifiche apportate alla parte C del modello 770

Con l’introduzione del modello *Emens* la parte C (sezione Dati previdenziali ed assistenziali Inps) del modello 770 subisce rilevanti modifiche.

<sup>65</sup> Al riguardo cfr *Documento tecnico per la compilazione dei flussi delle denunce retributive mensili*, Inps - Direzione Centrale Sistemi Informativi e Telecomunicazione, 2006: «...con questo elemento [Mese Precedente] possono essere modificati i dati relativi al mese precedente la denuncia. Con la sola esclusione della *Qualifica1*, *Qualifica2*, *Qualifica3*, *Tipo di contribuzione* e *Imponibile*, in quanto elementi che garantiscono la quadratura con il Dm10.»

La versione successiva all'introduzione del modello *Emens* (figura 6.9) risulta infatti nettamente ridimensionata rispetto a quella antecedente (figura 6.10). In particolare viene meno:

- la sezione 1, destinata a raccogliere informazioni sulla qualifica ed il contratto di lavoro del dipendente, sui mesi, settimane e giorni retribuiti, sulla quota Tfr accantonata e sulle coordinate Anf;
- la sezione 2, destinata a raccogliere dati sulle retribuzione particolari;
- la sezione 3, destinata a raccogliere dati sui periodi caratterizzati da contribuzione figurativa.

Con l'introduzione del modello *Emens*, la parte C (sezione Dati previdenziali ed assistenziali Inps) risulta quindi svuotata dei contenuti informativi originariamente utili ai fini del progetto Tmnp e ridotta a poco più di un mero rimando al modello *Emens*. Le voci 10 ed 11 sono infatti destinate all'indicazione dei mesi per i quali è stata presentata la denuncia *Emens*.

Figura 6.9 - Parte C del modello 770: versione successiva all'introduzione del modello *Emens* (esercizio 2005)

<b>PARTE C</b> <b>DATI PREVIDENZIALI ED ASSISTENZIALI INPS</b> SEZIONE 1 LAVORATORI SUBORDINATI	Matricola azienda		Ente pensionistico		CONTRIBUTI DOVUTI								
	1	2	INPS	Altro	3	4	Imponibile previdenziale	5	Sono stati interamente versati	6	Sono stati parzialmente versati	7	Non sono stati versati
	Contributi a carico del lavoratore trattenuti		Bonus L. 243/2004		MESI PER I QUALI È STATA PRESENTATA LA DENUNCIA EMens								
8	9	Tutti		10	11	Tutti con l'esclusione di			T G F M A M G L A S O N D				

Figura 6.10 - Parte C del modello 770: versione antecedente l'introduzione del modello *Emens* (esercizio 2004)

Codice fiscale del percipiente		Progressivo Comunicazione																			
<b>PARTE C</b> <b>DATI PREVIDENZIALI ED ASSISTENZIALI INPS</b> SEZIONE 1	1	Qualifica	2	3	Matricola azienda	4	Prox. Lav.	5	IVS	6	DS	7	8	9	Competenze correnti	10	Altre competenze				
	11	Settimane retrib.	12	Settimane utili	13	Giorni retrib.	MESI RETRIBUITI NELL'ANNO			Tutti con esclusione di			14	15	CONTRATTO	16	Codice	17	Tipo	18	Livello inquadramento
	Data cessazione		19	giorno	20	retrib.	21	Trasf. rapporto	22	Accantonamento TFR, speltante	COORD. ASSEGNI FAMILIARI			23	Tabella	24	Nim. fam.	25	Classe		
<b>SEZIONE 2</b> <b>RETRIBUZIONI PARTICOLARI</b>	26	Tipo	27	giorno	retrib.	28	giorno	retrib.	29	Retribuzione	30	GG.retrib.	31	GG.utili.	32	GG. non retrib.	33	Retribuzione pensionabile			
	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50				
	Tot.anno retrib. a retrib. ridotta		Retribuzioni ridotte		Differenza retribuzioni da accreditare per C/C		MALATTIA E/O INFORTUNIO		MAT e CONG. D. Lgs. 151/01		CONG. D. Lgs. 151/01										
	L.88/87		GG		Congedi art. 42 c. 5 D.Lgs. 151/01		Donat. sangue L107/90														
<b>SEZIONE 3</b> <b>CONTRIBUZIONE FIGURATIVA</b>	44	45	46	47	48	49	50														
	51	52	53	54	55																
<b>Contributi dovuti</b>		Contributi a carico del lavoratore trattenuti		Sono stati interamente versati		Sono stati parzialmente versati		Non sono stati versati		Bonus L. 243/2004											



## BIBLIOGRAFIA

- Casella G., R.L. Berger (1990), *Statistical Inference*, Duxbury Press.
- Consolini P. (2000), “Le prestazioni monetarie non pensionistiche: aspetti istituzionali e classificazioni statistiche”, Documenti Istat, n. 2/2000.
- Consolini P., R. De Carli (2000), “Le prestazioni sociali monetarie non pensionistiche: unità d'analisi, fonti e rappresentazione statistica dei dati”, Documenti Istat, n. 1/2002.
- Davies P. and P. Smith, *Model Quality Report in Business Statistics*, in Theory and Methods for Quality Evaluation, volume I.
- Eurostat (1996), *Esspros Manual*, Population and social conditions, Methods, Luxembourg.
- Eurostat (2001). *Esspros Manual*, Living Conditions, Methods. Luxembourg.
- Grande E., O. Luzi (2003), *Metodologie per l'imputazione delle mancate risposte parziali: Analisi critica e soluzioni diponibili in Istat*, Contributi Istat n. 6/2003.
- Greene W.H. (2000), *Econometric Analysis*, Fourth edition, Prentice-Hall International Editions, Upper Saddle River, New Jersey.
- Little R.J.A. and D.B. Rubin (1987), *Statistical Analysis with Missing Data*. J. Wiley & Sons, New York.
- Maddala G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
- Schafer, J.L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London.
- Spinelli V., M. Tancioni (2004), “I Trattamenti Monetari Non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10”, Contributi Istat, n. 28/2004.
- AA.VV. (2004), *Memento pratico – Previdenza*, Ipsoa-Francis Lefebvre



## GLOSSARIO

**Assegni al Nucleo Familiare:** Prestazione previdenziale erogata con cadenza mensile su richiesta del lavoratore o del pensionato, finalizzata ad integrarne la retribuzione in presenza di determinati requisiti di reddito e di dimensione/composizione familiare.

**Ateco (Classificazione delle attività economiche):** Classificazione che distingue le unità di produzione secondo l'attività da esse svolta e finalizzata all'elaborazione di statistiche di tipo macroeconomico, aventi per oggetto i fenomeni relativi alla partecipazione di tali unità ai processi economici. La classificazione comprende 88 categorie, raggruppate in 514 classi, 224 gruppi, 62 divisioni, 17 sezioni, 16 sottosezioni.

**Benchmark:** Cholette & Dagum (1994) definiscono il *benchmark* come segue "... si è in presenza di situazioni di *benchmark* quando sono disponibili due o più fonti di dati sulla stessa variabile d'interesse, con frequenze diverse, ad esempio, mensile contro annuale o mensile contro quadrimestrale. Generalmente, le due fonti di dati forniscono risultati diversi, ad esempio la somma delle misurazioni mensili potrebbe non coincidere con la misurazione annuale [...]. La misura più attendibile deve essere considerata il benchmark [...]. Tipicamente, un processo di convergenza verso il *benchmark* consiste di una combinazione di due fonti di dati volta a raggiungere un miglioramento delle stime delle serie investigate. Sotto questa prospettiva, i *benchmark* rappresentano *osservazioni ausiliarie*."

**Beneficiario del trattamento Anf:** qualsiasi membro del nucleo familiare a cui viene erogato il trattamento Anf.

**Codice Statistico Contributivo (Csc):** come da denominazione, il Codice statistico contributivo ha funzioni che sono nel contempo statistiche e contributive: consente rilevazioni statistiche ma determina anche quale deve essere il comportamento contributivo dell'azienda. Il Csc è un codice numerico formato da cinque caratteri, di cui il primo rappresenta il ramo di attività, il secondo ed il terzo rappresentano la classe e gli ultimi due la categoria.

**Componente a carico:** Membro del nucleo familiare che non contribuisce (ad eccezione di altre prestazioni sociali eventualmente percepite) alla formazione del reddito familiare. .

**Coordinate Anf:** Caratteristiche economiche (reddito) e demografiche (numero di componenti e composizione) del nucleo familiare, sulla base delle quali viene stabilito l'importo dell'assegno familiare mensilmente spettante.

**Disabili:** Persone con una o più forme di handicap (fisico, psichico, sensoriale o plurimo).

**Giorni retribuiti:** Giornate per le quali è stata percepita retribuzione, intera o ridotta.

**Importo Anf:** Ammontare degli assegni al nucleo familiare percepiti dal nucleo familiare nell'anno solare.

**Importo medio mensile Anf:** Rapporto tra importo totale e numeri di mesi per i quali si ha diritto alla corresponsione dell'assegno al nucleo familiare.

**Matricola Aziendale:** La matricola aziendale è un codice numerico assegnato dall'Inps per individuare la posizione contributiva di un'impresa. Ogni impresa può avere una o più matricole aziendali, ad esempio nel caso in cui l'attività venga svolta in diverse unità locali. I primi tre caratteri del codice individuano la sede Inps provinciale alla quale l'impresa è assegnata. I successivi 5 sono un numero progressivo assegnato indipendentemente dal ramo di appartenenza dell'azienda. I 2 caratteri rimanenti rappresentano il controcodice, una chiave di controllo dell'esatta trascrizione dei caratteri numerici che lo precedono e costituenti il codice base. Il controcodice si determina come segue: 1) si sommano procedendo da sinistra verso destra separatamente prima le cifre del codice base situate in posizioni dispari, poi quelle situate in posizioni pari; 2) dei due totali che si ottengono, si prendono le unità e le si accoppiano nell'ordine e cioè prima l'unità del totale dei numeri dispari, poi quella del totale dei numeri pari. Si forma in tal modo il numero di due cifre che viene definito controcodice.

**Mesi (di percepimento) Anf:** Mesi per i quali si ha diritto alla corresponsione dell'assegno familiare.

**Mesi Retribuiti:** Mesi per i quali è stata percepita retribuzione, intera e/o ridotta, per almeno un giorno solare.

**Modello Dm10:** modello di dichiarazione compilato mensilmente dalle imprese. In estrema sintesi raccoglie le voci a credito (prestazioni previdenziali anticipate) e a debito (contributi previdenziali dovuti) che caratterizzano il rapporto economico-contributivo tra l'impresa-datore di lavoro e l'Inps.

**Modello 770:** modello di dichiarazione compilato annualmente dal datore di lavoro con riferimento a ciascun lavoratore occupato (dipendente, collaboratore, domestico, autonomo...) per finalità fiscali e contributive.

**Nucleo familiare (famiglia):** Insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi, coabitanti ed aventi dimora abituale nello stesso comune (anche se non sono ancora iscritte nell'anagrafe della popolazione residente del comune medesimo). Una famiglia può essere costituita anche da una sola persona. L'assente temporaneo non cessa di appartenere alla propria famiglia sia che si trovi presso altro alloggio (o convivenza) dello stesso comune, sia che si trovi in un altro comune. Con riferimento alla dimensione/composizione della nucleo familiare viene stabilito l'ammontare dell'assegno familiare spettante. A tal fine, oltre al richiedente, fanno parte del nucleo familiare (cfr. art.2, comma 6, L. 153/88):

- coniuge non separato sia legalmente che effettivamente;
- figli minorenni (compresi i cd equiparati)
- figli maggiorenni se si trovano nell'assoluta impossibilità di lavorare a causa di infermità, difetto fisico o mentale;
- fratelli, sorelle e minori minorenni;
- fratelli, sorelle e minori maggiorenni, qualora si trovino nell'impossibilità di lavorare a causa di infermità fisica o mentale e siano comunque orfani di entrambi i genitori, senza diritto al pensione superstiti.

**Numero Assegni:** Coincide con il numero di Titolari dell'assegno al nucleo familiare, che a sua volta coincide con il numero delle famiglie beneficiarie degli stessi. Quindi è un valore non confrontabile con il numero di assegni mensili percepiti in senso stretto nell'arco dell'anno, dipendenti dal numero di mesi in cui si è beneficiato del trattamento.

**Popolazione residente:** per ciascun comune è costituita dalle persone aventi dimora abituale nel comune, anche se alla data considerata sono assenti perché temporaneamente presenti in altro comune italiano o all'estero

**Prestazione lavorativa minima:** Prestazione lavorativa minima richiesta al lavoratore affinché abbia diritto alla corresponsione dell'intero assegno familiare individuato dalle *Coordinate Anf* di riferimento.

Secondo l'art.59 del DPR 797/55) tale limite è fissato nel modo seguente: per gli operai 24 ore per periodi di paga settimanali, 48 ore per periodi di paga quattordicinali, 52 ore per periodi di paga quindicinali, 104 ore per periodi di paga mensili. Per gli impiegati i limiti minimi di prestazione lavorativa sono invece, rispettivamente, di 30, 60, 65 e 130 ore.

Nel caso in cui i limiti minimi siano rispettati l'assegno familiare sarà computato con riferimento a tutte le giornate lavorative comprese nel periodo di paga. In caso contrario sarà ridotto in misura proporzionale.

**Qualifica (professionale):** Inquadramento della posizione nella professione dei lavoratori dipendenti, classificabile nelle seguenti voci: dirigenti, quadri, impiegati, operai, altri.

**Reddito familiare:** È una delle tre coordinate Anf necessarie alla determinazione dell'ammontare dell'assegno familiare spettante. È costituito dalla somma dei redditi del titolare del trattamento e degli altri soggetti componenti il nucleo familiare. I redditi da considerare sono quelli dichiarati per l'anno precedente quello della richiesta ed hanno valore per la corresponsione dell'assegno fino al 30 giugno dell'anno successivo.

**Reddito Familiare netto:** La somma dei redditi (salari, stipendi, proventi professionali o imprenditoriali, pensioni, indennità, rendite sussidi ecc.) di ciascun componente la famiglia (avente almeno 16 anni compiuti) al netto di imposte e contributi sociali.

**Retribuzione ridotta:** retribuzione inferiore a quella ordinaria spettante in base al contratto di lavoro di riferimento. La corresponsione della retribuzione ridotta è tipicamente associata ad eventi particolari come la malattia, la maternità, i congedi parentali, i periodi di integrazione salariale.

**Settimane retribuite:** Settimane per le quali è stata percepita retribuzione, intera e/o ridotta, per almeno un giorno solare.

**Spesa media mensile:** è calcolata dividendo la spesa totale relativa all'acquisto di beni (o servizi) per il numero complessivo delle famiglie italiane (senza distinguere, cioè, quelle che hanno effettuato la spesa specifica da quelle che non l'hanno effettuata)

**Titolare del trattamento Anf:** Membro del nucleo familiare che ha presentato l'apposito modulo di domanda al proprio datore di lavoro ed al quale viene conseguentemente intestato il trattamento. Tale soggetto viene rilevato nel modello 770 tramite i seguenti criteri:

- è titolare del trattamento qualunque persona fisica individuata da una dichiarazione per l'anno t, contenente un valore significativo per almeno una delle tre coordinate Anf;
- è titolare del trattamento qualunque persona fisica presente, pur senza coordinate Anf, nell'archivio dell'anno t per periodi comprensivi del mese di gennaio e risultante percettore, secondo il criterio sopra esposto, nell'archivio dell'anno t-1 per periodi retributivi comprensivi del mese di dicembre

**Trattamenti monetari non pensionistici:** erogazioni temporanee in denaro che danno luogo a trasferimenti di risorse monetarie dalle istituzioni o dalle imprese alle famiglie, per la copertura di rischi o bisogni che sono compresi nella protezione sociale.

**Unità locale:** Luogo fisico nel quale un'unità giuridico-economica (impresa, istituzione) esercita una o più attività economiche. L'unità locale corrisponde ad un'unità giuridico-economica o ad una sua parte, situata in una località topograficamente identificata da un indirizzo e da un numero civico. In tale località, o a partire da tale località, si esercitano delle attività economiche per le quali una o più persone lavorano (eventualmente a tempo parziale) per conto della stessa unità giuridico-economica. Costituiscono esempi di unità locale le seguenti tipologie: agenzia, albergo,

ambulatorio, bar, cava, deposito, domicilio, garage, laboratorio, magazzino, miniera, negozio, officina, ospedale, ristorante, scuola, stabilimento, studio professionale, ufficio, ecc.



## Contributi ISTAT(\*)

- 1/2002 - Francesca Biancani, Andrea Carone, Rita Pistacchio e Giuseppina Ruocco - *Analisi delle imprese individuali*
- 2/2002 - Massimiliano Borgese - *Proposte metodologiche per un progetto d'indagine sul trasporto aereo alla luce della recente normativa comunitaria sul settore*
- 3/2002 - Nadia Di Veroli e Roberta Rizzi - *Proposta di classificazione dei rapporti di lavoro subordinato e delle attività di lavoro autonomo: analisi del quadro normativo*
- 4/2002 - Roberto Gismondi - *Uno stimatore ottimale in presenza di non risposte*
- 5/2002 - Maria Anna Pennucci - *Le strategie europee per l'occupazione dal Libro bianco di Delors al Consiglio Europeo di Cardiff*
- 1/2003 - Giovanni Maria Merola - *Safety Rules in Statistical Disclosure Control for Tabular Data*
- 2/2003 - Fabio Bacchini, Pietro Gennari e Roberto Iannaccone - *A new index of production for the construction sector based on input data*
- 3/2003 - Fulvia Ceroni e Enrica Morganti - *La metodologia e il potenziale informativo dell'archivio sui gruppi di impresa: primi risultati*
- 4/2003 - Sara Mastrovita e Isabella Siciliani - *Effetti dei trasferimenti sociali sulla distribuzione del reddito nei Paesi dell'Unione europea: un'analisi dal Panel europeo sulle famiglie*
- 5/2003 - Patrizia Cella, Giuseppe Garofalo, Adriano Paggiaro, Nicola Torelli e Caterina Viviano - *Demografia d'impresa: l'utilizzo di tecniche di abbinamento per l'analisi della continuità*
- 6/2003 - Enrico Grande e Orietta Luzi - *Metodologie per l'imputazione delle mancate risposte parziali: analisi critica e soluzioni disponibili in Istat*
- 7/2003 - Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino - *Indagine sperimentale sui posti di lavoro vacanti*
- 8/2003 - Mario Adua - *L'agricoltura di montagna: le aziende delle donne, caratteristiche agricole e socio-rurali*
- 9/2003 - Franco Mostacci e Roberto Sabbatini - *L'euro ha creato inflazione? Changeover e arrotondamenti dei prezzi al consumo in Italia nel 2002*
- 10/2003 - Leonello Tronti - *Problemi e prospettive di riforma del sistema pensionistico*
- 11/2003 - Roberto Gismondi - *Tecniche di stima e condizioni di coerenza per indagini infraannuali ripetute nel tempo*
- 12/2003 - Antonio Frenda - *Analisi delle legislazioni e delle prassi contabili relative ai gruppi di imprese nei paesi dell'Unione Europea*
- 1/2004 - Marcello D'Orazio, Marco Di Zio e Mauro Scanu - *Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints*
- 2/2004 - Giovanna Brancato - *Metodologie e stime dell'errore di risposta. Una sperimentazione di reintervista telefonica*
- 3/2004 - Franco Mostacci, Giuseppina Natale e Elisabetta Pugliese - *Gli indici dei prezzi al consumo per sub popolazioni*
- 4/2004 - Leonello Tronti - *Una proposta di metodo: osservazioni e raccomandazioni sulla definizione e la classificazione di alcune variabili attinenti al mercato del lavoro*
- 5/2004 - Ugo Guarnera - *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi: il software Quis*
- 6/2004 - Patrizia Giaquinto, Marco Landriscina e Daniela Pagliuca - *La nuova funzione di analisi dei modelli implementata in Genesees v. 3.0*
- 7/2004 - Roberto Di Giuseppe, Patrizia Giaquinto e Daniela Pagliuca - *MAUSS (Multivariate Allocation of Units in Sampling Surveys): un software generalizzato per risolvere il problema dell'allocazione campionaria nelle indagini Istat*
- 8/2004 - Ennio Fortunato e Liana Verzicco - *Problemi di rilevazione e integrazione della condizione professionale nelle indagini sociali dell'Istat*
- 9/2004 - Claudio Pauselli e Claudia Rinaldelli - *La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*
- 10/2004 - Eugenio Arcidiacono, Marina Briolini, Paolo Giuberti, Marco Ricci, Giovanni Sacchini e Giorgia Telloli - *Procedimenti giudiziari, reati, indagati e vittime in Emilia-Romagna nel 2002: un'analisi territoriale sulla base dei procedimenti iscritti nel sistema informativo Re.Ge.*
- 11/2004 - Enrico Grande e Orietta Luzi - *Regression trees in the context of imputation of item non-response: an experimental application on business data*
- 12/2004 - Luisa Frova e Marilena Pappagallo - *Procedura di now-cast dei dati di mortalità per causa*
- 13/2004 - Giorgio DellaRocca, Marco Di Zio, Orietta Luzi, Emanuela Scavalli e Giorgia Simeoni - *IDEA (Indices for Data Editing Assessment): sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI*
- 14/2004 - Monica Pace, Silvia Bruzzone, Luisa Frova e Marilena Pappagallo - *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe*
- 15/2004 - Elisa Berntsen - *Modello Unico di Dichiarazione ambientale: una fonte amministrativa per l'Archivio delle Unità Locali di Asia*
- 16/2004 - Salvatore F. Allegra e Alessandro La Rocca - *Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la definizione di un indice composto*
- 17/2004 - Francesca R. Pogelli - *Un'applicazione del modello "Country Product Dummy" per un'analisi territoriale dei prezzi*
- 18/2004 - Antonia Manzari - *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*
- 19/2004 - Claudio Pauselli - *Intensità di povertà relativa: stima dell'errore di campionamento e sua valutazione temporale*
- 20/2004 - Maria Dimitri, Ersilia Di Pietro, Alessandra Nuccitelli e Evelina Paluzzi - *Sperimentazione di una metodologia per il controllo della qualità di dati anagrafici*
- 21/2004 - Tiziana Pichiorri, Anna M. Sgamba e Valerio Papale - *Un modello di ottimizzazione per l'imputazione delle mancate risposte statistiche nell'indagine sui trasporti marittimi dell'Istat*

- 22/2004 – Diego Bellisai, Piero D. Falorsi, Annalisa Lucarelli, Maria A. Pennucci e Leonello G. Tronti – *Indagine pilota sulle retribuzioni di fatto nel pubblico impiego*
- 23/2004 – Lidia Brondi – *La riorganizzazione del sistema idrico: quadro normativo, delimitazione degli ambiti territoriali ottimali e analisi statistica delle loro caratteristiche strutturali*
- 24/2004 – Roberto Gismondi e Laura De Sandro – *Provisional Estimation of the Italian Monthly Retail Trade Index*
- 25/2004 – Annamaria Urbano, Claudia Brunini e Alessandra Chessa – *I minori in stato di abbandono: analisi del fenomeno e studio di una nuova prospettiva d'indagine*
- 26/2004 – Paola Anzini e Anna Ciammola – *La destagionalizzazione degli indici della produzione industriale: un confronto tra approccio diretto e indiretto*
- 27/2004 – Alessandro La Rocca – *Analisi della struttura settoriale dell'occupazione regionale: 8° Censimento dell'industria e dei servizi 2001 7° Censimento dell'industria e dei servizi 1991*
- 28/2004 – Vincenzo Spinelli e Massimiliano Tancioni – *I Trattamenti Monetari non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10*
- 29/2004 – Paolo Consolini – *L'indagine sperimentale sull'archivio fiscale modd.770 anno 1999: analisi della qualità del dato e stime campionarie*
- 1/2005 – Fabrizio M. Arosio – *La stampa periodica e l'informazione on-line: risultati dell'indagine pilota sui quotidiani on-line*
- 2/2005 – Marco Di Zio, Ugo Guarnera e Orietta Luzi – *Improving the effectiveness of a probabilistic editing strategy for business data*
- 3/2005 – Diego Moretti e Claudia Rinaldelli – *EU-SILC complex indicators: the implementation of variance estimation*
- 4/2005 – Fabio Bacchini, Roberto Iannaccone e Edoardo Otranto – *L'imputazione delle mancate risposte in presenza di dati longitudinali: un'applicazione ai permessi di costruzione*
- 5/2005 – Marco Broccoli – *Analisi della criminalità a livello comunale: metodologie innovative*
- 6/2005 – Claudia De Vitiis, Loredana Di Consiglio e Stefano Falorsi – *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*
- 7/2005 – Edoardo Otranto e Roberto Iannaccone – *Continuous Time Models to Extract a Signal in Presence of Irregular Surveys*
- 8/2005 – Cosima Mero e Adriano Pareto – *Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie*
- 9/2005 – Filippo Oropallo – *Enterprise microsimulation models and data challenges*
- 10/2005 – Marcello D' Orazio, Marco Di Zio e Mauro Scanu – *A comparison among different estimators of regression parameters on statistically matched files through an extensive simulation study*
- 11/2005 – Stefania Macchia, Manuela Murgia, Loredana Mazza, Giorgia Simeoni, Francesca Di Patrizio, Valentino Parisi, Roberto Petrillo e Paola Ungaro – *Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI*
- 12/2005 – Piero D. Falorsi, Monica Scannapieco, Antonia Boggia e Antonio Pavone – *Principi Guida per il Miglioramento della Qualità dei Dati Toponomastici nella Pubblica Amministrazione*
- 13/2005 – Ciro Baldi, Francesca Ceccato, Silvia Pacini e Donatella Tuzi – *La stima anticipata OROS sull'occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento*
- 14/2005 – Stefano De Francisci, Giuseppe Sindoni e Leonardo Tininini – *Da Winci/MD: un sistema per data warehouse statistici sul Web*
- 15/2005 – Gerardo Gallo e Evelina Palazzi – *I cittadini italiani naturalizzati: l'analisi dei dati censuari del 2001, con un confronto tra immigrati di prima e seconda generazione*
- 16/2005 – Saverio Gazzelloni, Mario Albisinni, Lorenzo Bagatta, Claudio Ceccarelli, Luciana Quattrociochi, Rita Ranaldi e Antonio Toma – *La nuova rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*
- 17/2005 – Maria Carla Congia – *Il lavoro degli extracomunitari nelle imprese italiane e la regolarizzazione del 2002. Prime evidenze empiriche dai dati INPS*
- 18/2005 – Giovanni Bottazzi, Patrizia Cella, Giuseppe Garofalo, Paolo Misso, Mariano Porcu e Marianna Tosi – *Indagine pilota sulla nuova imprenditorialità nella Regione Sardegna. Relazione Conclusiva*
- 19/2005 – Fabrizio Martire e Donatella Zindato – *Le famiglie straniere: analisi dei dati censuari del 2001 sui cittadini stranieri residenti*
- 20/2005 – Ennio Fortunato – *Il Sistema di Indicatori Territoriali: percorso di progetto, prospettive di sviluppo e integrazione con i processi di produzione statistica*
- 21/2005 – Antonella Baldassarini e Danilo Birardi – *I conti economici trimestrali: un approccio alla stima dell'input di lavoro*
- 22/2005 – Francesco Rizzo, Dario Camol e Laura Vignola – *Uso di XML e WEB Services per l'integrazione di sistemi informativi statistici attraverso lo standard SDMX*
- 1/2006 – Ennio Fortunato – *L'analisi integrata delle esigenze informative dell'utenza Istat: Il contributo del Sistema di Indicatori Territoriali*
- 2/2006 – Francesco Altarocca – *I design pattern nella progettazione di software per il supporto alla statistica ufficiale*
- 3/2006 – Roberta Palmieri – *Le migranti straniere: una lettura di genere dei dati dell'osservatorio interistituzionale sull'immigrazione in provincia di Macerata*
- 4/2006 – Raffaella Amato, Silvia Bruzzone, Valentina Delmonte e Lidia Fagiolo – *Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage*
- 5/2006 – Alessandro La Rocca – *Fuzzy clustering: la logica, i metodi*
- 6/2006 – Raffaella Cascioli – *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province*
- 7/2006 – Gianluca Brogi, Salvatore Cusimano, Giuseppina del Vicario, Giuseppe Garofalo e Orietta Patacchia – *La realizzazione di Asia Agricoltura tramite l'utilizzo di dati amministrativi: il contenuto delle fonti e i risultati del processo di integrazione*
- 8/2006 – Simonetta Cozzi – *La distribuzione commerciale in Italia: caratteristiche strutturali e tendenze evolutive*
- 9/2006 – Giovanni Seri – *A graphical framework to evaluate risk assessment and information loss at individual level*

- 10/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Feasibility studies for the coverage of public institutions in sections N and O*
- 11/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Quarterly labour cost index in public education*
- 12/2006 – Silvia Montagna, Patrizia Collesi, Florinda Damiani, Danila Fulgenzio, Maria Francesca Loporcaro e Giorgia Simeoni – *Nuove esperienze di rilevazione della Customer Satisfaction*
- 13/2006 – Lucia Coppola e Giovanni Seri – *Confidentiality aspects of household panel surveys: the case study of Italian sample from EU-SILC*
- 14/2006 – Lidia Brondi – *L'utilizzazione delle surveys per la stima del valore monetario del danno ambientale: il metodo della valutazione contingente*
- 15/2006 – Carlo Boselli – *Le piccole imprese leggere esportatrici e non esportatrici: differenze di struttura e di comportamento*
- 16/2006 – Carlo De Gregorio – *Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione*
- 1/2007 – Paolo Roberti, Maria Grazia Calza, Filippo Oropallo e Stefania Rossetti – *Knowledge Databases to Support Policy Impact Analysis: the EuroKy-PIA Project*
- 2/2007 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, e Marina Sorrentino – *Production of job vacancy statistics: coverage*
- 3/2007 – Carlo Lucarelli e Giampiero Ricci – *Working times and working schedules: the framework emerging from the new Italian lfs in a gender perspective*
- 4/2007 – Monica Scannapieco, Diego Zardetto e Giulio Barcaroli – *La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESSEES/SAS*
- 5/2007 – Giulio Barcaroli e Tiziana Pellicciotti – *Strumenti per la documentazione e diffusione dei microdati d'indagine: il Microdata Management Toolkit*
- 6/2007 – AA.VV. – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 1<sup>a</sup> giornata*
- 7/2007 – Raffaella Cianchetta, Carlo De Gregorio, Giovanni Seri e Giulio Barcaroli – *Rilevazione sulle Pubblicazioni Scientifiche Istat*
- 8/2007 – Emilia Arcaleni e Barbara Baldazzi – *Vivere non insieme: approcci conoscitivi al Living Apart Together*
- 9/2007 – Corrado Peperoni e Francesca Tuzi – *Trattamenti monetari non pensionistici metodologia sperimentale per la stima degli assegni al nucleo familiare*



## Documenti ISTAT(\*)

- 1/2002 – Paolo Consolini e Rita De Carli - *Le prestazioni sociali monetarie non pensionistiche: unità di analisi, fonti e rappresentazione statistica dei dati*
- 2/2002 – Stefania Macchia - *Sperimentazione, implementazione e gestione dell'ambiente di codifica automatica della classificazione delle Attività economiche*
- 3/2002 – Maria De Lucia - *Applicabilità della disciplina in materia di festività nel pubblico impiego*
- 4/2002 – Roberto Gismondi, Massimo Marciani e Mauro Giorgetti - *The italian contribution towards the implementation of an european transport information system: main results of the MESUDEMO project*
- 5/2002 – Olimpio Cianfarani e Sauro Angeletti - *Misure di risultato e indicatori di processo: l'esperienza progettuale dell'Istat*
- 6/2002 – Riccardo Carbinì e Valerio De Santis – *Programma statistico nazionale: specifiche e note metodologiche per la compilazione delle schede identificative dei progetti*
- 7/2002 – Maria De Lucia – *Il CCNL del personale dirigente dell'area 1 e la valutazione delle prestazioni dei dirigenti*
- 8/2002 – Giuseppe Garofalo e Enrica Morganti – *Gruppo di lavoro per la progettazione di un archivio statistico sui gruppi d'impresa*
- 1/2003 – Francesca Ceccato, Massimiliano Tancioni e Donatella Tuzi – *MODSIM-P: Il nuovo modello dinamico di previsione della spesa pensionistica*
- 2/2003 – Anna Pia Mirto – *Definizioni e classificazioni delle strutture ricettive nelle rilevazioni statistiche ufficiali sull'offerta turistica*
- 3/2003 – Simona Spirito – *Le prestazioni assistenziali monetarie non pensionistiche*
- 4/2003 – Maria De Lucia – *Approfondimenti di alcune tematiche inerenti la gestione del personale*
- 5/2003 – Rosalia Coniglio, Marialuisa Cugno, Maria Filmeno e Alberto Vitalini – *Mappatura della criminalità nel distretto di Milano*
- 6/2003 – Maria Letizia D'Autilia – *I provvedimenti di riforma della pubblica amministrazione per l'identificazione delle "Amministrazioni pubbliche" secondo il Sec95: analisi istituzionale e organizzativa per l'anno 2000*
- 7/2003 – Francesca Gallo, Pierpaolo Massoli, Sara Mastrovita, Roberto Merluzzi, Claudio Pauselli, Isabella Siciliani e Alessandra Sorrentino – *La procedura di controllo e correzione dei dati Panel Europeo sulle famiglie*
- 8/2003 – Cinzia Castagnaro, Martina Lo Conte, Stefania Macchia e Manuela Murgia – *Una soluzione in-house per le indagini CATI: il caso della Indagine Campionaria sulle Nascite*
- 9/2003 – Anna Pia Maria Mirto e Norina Salamone – *La classificazione delle strutture ricettive turistiche nella normativa delle regioni italiane*
- 10/2003 – Roberto Gismondi e Anna Pia Maria Mirto – *Le fonti statistiche per l'analisi della congiuntura turistica: il mosaico italiano*
- 11/2003 – Loredana Di Consiglio e Stefano Falorsi – *Alcuni aspetti metodologici relativi al disegno dell'indagine di copertura del Censimento Generale della Popolazione 2001*
- 12/2003 – Roberto Gismondi e Anna Rita Giorgi – *Struttura e dinamica evolutiva del comparto commerciale al dettaglio: le tendenze recenti e gli effetti della riforma "Bersani"*
- 13/2003 – Donatella Cangialosi e Rosario Milazzo – *Fabbisogni formativi degli Uffici comunali di statistica: indagine rapida in Sicilia*
- 14/2003 – Agostino Buratti e Giovanni Salzano – *Il sistema automatizzato integrato per la gestione delle rilevazioni dei documenti di bilancio degli enti locali*
- 1/2004 – Giovanna Brancato e Giorgia Simeoni – *Tesori del Sistema Informativo di Documentazione delle Indagini (SIDI)*
- 2/2004 – Corrado Peperoni – *Indagine sui bilanci consuntivi degli Enti previdenziali: rilevazione, gestione e procedure di controllo dei dati*
- 3/2004 – Marzia Angelucci, Giovanna Brancato, Dario Camol, Alessio Cardacino, Sandra Maresca e Concetta Pellegrini – *Il sistema ASIMET per la gestione delle Note Metodologiche dell'Annuario Statistico Italiano*
- 4/2004 – Francesca Gallo, Sara Mastrovita, Isabella Siciliani e Giovanni Battista Arcieri – *Il processo di produzione dell'Indagine ECHP*
- 5/2004 – Natale Renato Fazio e Carmela Pascucci – *Gli operatori non identificati nelle statistiche del commercio con l'estero: metodologia di identificazione nelle spedizioni "groupage" e miglioramento nella qualità dei dati*
- 6/2004 – Diego Moretti e Claudia Rinaldelli – *Una valutazione dettagliata dell'errore campionario della spesa media mensile familiare*
- 7/2004 – Franco Mostacci – *Aspetti Teorico-pratici per la Costruzione di Indici dei Prezzi al Consumo*
- 8/2004 – Maria Frustaci – *Glossario economico-statistico multilingua*
- 9/2004 – Giovanni Seri e Maurizio Lucarelli – *"Il Laboratorio per l'analisi dei dati elementari (ADELE): monitoraggio dell'attività dal 1999 al 2004"*
- 10/2004 – Alessandra Nuccitelli, Francesco Bosio e Luciano Fioriti – *L'applicazione RECLINK per il record linkage: metodologia implementata e linee guida per la sua utilizzazione*
- 1/2005 – Francesco Cuccia, Simone De Angelis, Antonio Laureti Palma, Stefania Macchia, Simona Mastroluca e Domenico Perrone – *La codifica delle variabili testuali nel 14° Censimento Generale della Popolazione*
- 2/2005 – Marina Peci – *La statistica per i Comuni: sviluppo e prospettive del progetto Sisco.T (Servizio Informativo Statistico Comunale. Tavole)*
- 3/2005 – Massimiliano Renzetti e Annamaria Urbano – *Sistema Informativo sulla Giustizia: strumenti di gestione e manutenzione*
- 4/2005 – Marco Broccoli, Roberto Di Giuseppe e Daniela Pagliuca – *Progettazione di una procedura informatica generalizzata per la sperimentazione del metodo Microstrat di coordinamento della selezione delle imprese soggette a rilevazioni nella realtà Istat*
- 5/2005 – Mauro Albani e Francesca Pagliara – *La ristrutturazione della rilevazione Istat sulla criminalità minorile*
- 6/2005 – Francesco Altarocca e Gaetano Sberno – *Progettazione e sviluppo di un "Catalogo dei File Grezzi con meta-dati di base" (CFG) in tecnologia Web*

- 7/2005 – Salvatore F. Allegra e Barbara Baldazzi – *Data editing and quality of daily diaries in the Italian Time Use Survey*
- 8/2005 – Alessandra Capobianchi – *Alcune esperienze in ambito internazionale per l'accesso ai dati elementari*
- 9/2005 – Francesco Rizzo, Laura Vignola, Dario Camol e Mauro Bianchi – *Il progetto "banca dati della diffusione congiunturale"*
- 10/2005 – Ennio Fortunato e Nadia Mignolli – *I sistemi informativi Istat per la diffusione via web*
- 11/2005 – Ennio Fortunato e Nadia Mignolli – *Sistemi di indicatori per l'attività di governo: l'offerta informativa dell'Istat*
- 12/2005 – Carlo De Gregorio e Stefania Fatello – *L'indice dei prezzi al consumo dei testi scolastici nel 2004*
- 13/2005 – Francesco Rizzo e Laura Vignola – *RSS: uno standard per diffondere informazioni*
- 14/2005 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino – *Launching and implementing the job vacancy statistics*
- 15/2005 – Stefano De Francischi, Massimiliano Renzetti, Giuseppe Sindoni e Leonardo Tininini – *La modellazione dei processi nel Sistema Informativo Generalizzato di Diffusione dell'ISTAT*
- 16/2005 – Ennio Fortunato e Nadia Mignolli – *Verso il Sistema di Indicatori Territoriali: rilevazione e analisi della produzione Istat*
- 17/2005 – Raffaella Cianchetta e Daniela Pagliuca – *Soluzioni Open Source per il software generalizzato in Istat: il caso di PHPSurveyor*
- 18/2005 – Gianluca Giuliani e Barbara Boschetto – *Gli indicatori di qualità dell'Indagine continua sulle Forze di Lavoro dell'Istat*
- 19/2005 – Rossana Balestrino, Franco Garritano, Carlo Cipriano e Luciano Fanfoni – *Metodi e aspetti tecnologici di raccolta dei dati sulle imprese*
- 1/2006 – Roberta Roncati – [www.istat.it](http://www.istat.it) (versione 3.0) *Il nuovo piano di navigazione*
- 2/2006 – Maura Seri e Annamaria Urbano – *Sistema Informativo Territoriale sulla Giustizia: la sezione sui confronti internazionali*
- 3/2006 – Giovanna Brancato, Riccardo Carbini e Concetta Pellegrini – *SIQual: il sistema informativo sulla qualità per gli utenti esterni*
- 4/2006 – Concetta Pellegrini – *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza SIDI*
- 5/2006 – Maurizio Lucarelli – *Una valutazione critica dei modelli di accesso remoto nella comunicazione di informazione statistica*
- 6/2006 – Natale Renato Fazio – *La ricostruzione storica delle statistiche del commercio con l'estero per gli anni 1970-1990*
- 7/2006 – Emilia D'Acunto – *L'evoluzione delle statistiche ufficiali sugli indici dei prezzi al consumo*
- 8/2006 – Ugo Guarnera, Orietta Luzi e Stefano Salvi – *Indagine struttura e produzioni delle aziende agricole: la nuova procedura di controllo e correzione automatica per le variabili su superfici aziendali e consistenza degli allevamenti*
- 9/2006 – Maurizio Lucarelli – *La regionalizzazione del Laboratorio ADELE: un'ipotesi di sistema distribuito per l'accesso ai dati elementari*
- 10/2006 – Alessandra Bugio, Claudia De Vitiis, Stefano Falorsi, Lidia Gargiulo, Emilio Gianicolo e Alessandro Pallara – *La stima di indicatori per domini sub-regionali con i dati dell'indagine: condizioni di salute e ricorso ai servizi sanitari*
- 11/2006 – Sonia Vittozzi, Paola Giacchè, Achille Zuchegna, Piero Crivelli, Patrizia Collesi, Valerio Tiberi, Alexia Sasso, Maurizio Bonsignori, Giuseppe Stassi e Giovanni A. Barbieri – *Progetto di articolazione della produzione editoriale in collane e settori*
- 12/2006 – Alessandra Coli, Francesca Tartamella, Giuseppe Sacco, Ivan Faiella, Marcello D'Orazio, Marco Di Zio, Mauro Scanu, Isabella Siciliani, Sara Colombini e Alessandra Masi – *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'Indagine Banca d'Italia sui bilanci delle famiglie italiane*
- 13/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Intrastat*
- 14/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Extrastat*
- 15/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: comparazione tra rilevazione Intrastat ed Extrastat*
- 16/2006 – Fabio M. Rapiti – *Short term statistics quality Reporting: the LCI National Quality Report 2004*
- 17/2006 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Valutazione delle possibilità di uso di dati fiscali a supporto della rilevazione PMI*
- 18/2006 – Mauro Albani – *La nuova procedura per il trattamento dei dati dell'indagine Istat sulla criminalità*
- 19/2006 – Alessandra Capobianchi – *Review dei sistemi di accesso remoto: schematizzazione e analisi comparativa*
- 20/2006 – Francesco Altarocca – *Gli strumenti informatici nella raccolta dei dati di indagini statistiche: il caso della Rilevazione sperimentale delle tecnologie informatiche e della comunicazione nelle Pubbliche Amministrazioni locali*
- 1/2007 – Giuseppe Stassi – *La politica editoriale dell'Istat nel periodo 1996-2004: collane, settori, modalità di diffusione*
- 2/2007 – Daniela Ichim – *Microdata anonymisation of the Community Innovation Survey data: a density based clustering approach for risk assessment*
- 3/2007 – Ugo Guarnera, Orietta Luzi e Irene Tommasi – *La nuova procedura di controllo e correzione degli errori e delle mancate risposte parziali nell'indagine sui Risultati Economici delle Aziende Agricole (REA)*
- 4/2007 – Vincenzo Spinelli – *Processo di Acquisizione e Trattamento Informativo degli Archivi relativi al Modello di Dichiarazione 770*
- 5/2007 – Anna Di Carlo, Maria Picci, Laura Posta, Michaela Raffone, Giuseppe Stassi e Fiorella Tortora – *La progettazione dei Censimenti generali 2010-2011: 1 - Analisi, valutazione e proposte in merito ad atti di normazione e finanziamento*
- 6/2007 – Silvia Bruzzone, Antonia Manzari, Marilena Pappagallo e Alessandra Reale – *Indagine sulle Cause di Morte: Nuova procedura automatica per il controllo e la correzione delle variabili demo-sociali*
- 7/2007 – Maura Giacommo, Carlo Vaccari e Monica Scannapieco – *Indagine sulle Scelte Tecnologiche degli Istituti Nazionali di Statistica*
- 8/2007 – Lamberto Pizzicannella – *Sviluppo del processo di acquisizione e trattamento informatico degli archivi relativi al modello di dichiarazione 770. Anni 2004 – 2005*
- 9/2007 – Damiano Abbadini, Lorenzo Cassata, Fabrizio Martire, Alessandra Reale, Giuseppina Ruocco e Donatella Zindato – *La progettazione dei Censimenti generali 2010-2011 2 - Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani*

- 10/2007 – Marco Fortini, Gerardo Gallo, Evelina Paluzzi, Alessandra Reale e Angela Silvestrini – *La progettazione dei censimenti generali 2010–2011 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*
- 11/2007 – Domenico Adamo, Damiana Cardoni, Valeria Greco, Silvia Montecolle, Sante Orsini, Alessandro Ortensi e Miria Savioli – *Strategie di correzione del questionario sulla qualità della vita dell'infanzia e dell'adolescenza. Indagine multiscopo sulle famiglie. Aspetti della vita quotidiana 2005*
- 12/2007 – Carlo Nappi – *Manuale per la preparazione di originali "ready to print"*