

CONTRIBUTI ISTAT

n. /2004 L'indagine sperimentale sull'archivio fiscale modd.770
anno 1999: analisi della qualità del dato e stime campionarie

P. Consolini ()*

(*) ISTAT – Servizio Statistiche sulle istituzioni pubbliche e private

INDICE

1. Introduzione	
2. La struttura dei modelli 770-1999 del Ministero dell'Economia e delle Finanze	
3. Le procedure di sistemazione automatica dei record in un formato fisso leggibile in SAS	
4. Le procedure di controllo e correzione dei dati relativi al quadro SA	
5. Le procedure di controllo e correzione dei dati sul quadro SB	
6. Le stime campionarie sulle liquidazioni di fine rapporto e retribuzioni ridotte	
7. Conclusioni	
Riferimenti bibliografici	

Sommario

L'autore descrive alcune fasi dell'indagine sperimentale sulle dichiarazioni fiscali modd.770 - dati campionari di 22446 imprese, anno 1999 - del Ministero dell'Economia e delle Finanze. L'obiettivo prioritario di questa indagine sperimentale è stato di testare l'affidabilità dei dati economici sui trattamenti monetari non pensionistici (TMNP) presenti in archivio e di implementare le prime procedure di lettura, controllo e correzione dei dati, ai fini del loro impiego a regime nell'indagine TMNP (*codice Piano Statistico Nazionale: IST-00915*). Le soluzioni adottate costituiscono il primo tentativo di consolidamento dei dati economici sulle liquidazioni di fine rapporto e sulle retribuzioni ridotte. Si rammenta che, a causa della complessità e del numero delle informazioni contenute in archivio, si è voluto limitare l'analisi ad una sezione dei dati disponibili e pertanto le procedure utilizzate sono idonee a correggere solo alcune delle variabili di interesse. Successivi studi forniranno delle soluzioni che utilizzeranno organicamente il complesso delle informazioni presenti in archivio.

Abstract

The author describes some points of the experimental research on the "770-Form tax register" of the Ministry of Economy and Finance. This study covers a sample of 22.446 firms that are obliged to return information on "wages and social benefits", paid to their employees by means of the "770 Form, in the year 1999. The aims of this research were to test the data quality (reliability) of the variables of interest (non pension cash benefit benefits and recipients) recorded in 770-form register, and to implement the procedures for the detection and correction of errors. These procedures will be applied on the register of 770-Form as whole, in order to provide statistics regarding social benefits provided directly by employers to their employees. The criteria adopted for the error's detection and correction take into account only a specific group of variables of interest and provide a partial solution to the consolidation of the database on non pension cash

benefit. Therefore, this experimental study should be considered a starting point for the implementation of a consolidate procedure working on the whole set of variables and units of the 770-Form Register.

1. Introduzione¹

Le elaborazioni effettuate sull'archivio fiscale modd.770 - dati campionari² anno 1999 - hanno avuto come obiettivo prioritario la valutazione dell'affidabilità dei dati economici sui trattamenti monetari non pensionistici (TMNP), in esso presenti, ai fini del loro impiego a regime nell'indagine TMNP (*codice Piano Statistico Nazionale: IST-00915*). In particolare, l'archivio fiscale modd.770 consente di rilevare direttamente, a livello di singolo beneficiario, il dato finanziario su: 1) liquidazioni di fine rapporto del settore privato³ (comprendenti il TFR, gli incentivi all'esodo anticipato e le indennità sostitutive di preavviso); 2) trattamenti a carico esclusivo del datore di lavoro privato per malattia e per maternità (periodi non indennizzati dall'INPS ed integrazioni economiche rispetto all'indennità di base) previsti dai CCNL, denominati "retribuzioni ridotte". Inoltre, l'archivio fiscale modd.770 contiene informazioni che consentono di ripartire la spesa per le prestazioni non pensionistiche anticipate dai datori privati per conto dell'INPS, attribuendo a ciascun lavoratore il corrispondente flusso monetario. Al contrario, l'archivio DM10 dell'INPS rileva il dato economico sulle prestazioni anticipate a livello di impresa e non di singolo beneficiario.

Le *unità di rilevazione* dell'archivio modd.770, utili ai fini dell'indagine TMNP, sono costituite dai datori di lavoro che erogano spesa non pensionistica direttamente o per conto dell'INPS. Esse sono rappresentate dai soggetti obbligati a presentare la dichiarazione del quadro SA, relativamente ai dati previdenziali ed assistenziali (ex modelli O1M dell'Inps) e/o dai soggetti tenuti alla compilazione del quadro SB. Questo è riservato ai sostituti d'imposta che nel corso dell'anno hanno corrisposto trattamenti di fine rapporto di lavoro dipendente, nonché altre indennità equipollenti di cui all'art.16, comma 1, lett.a, del D.P.R. 917/86 (indennità di mancato preavviso, incentivi all'esodo, liquidazioni in capitale di forme pensionistiche complementare).

Le *unità di analisi* sono costituite dalle singole dichiarazioni SA e/o SB del modd.770 associate a ciascun percipiente, per il quale il datore di lavoro ha corrisposto, nel corso dell'anno di riferimento, somme o valori soggetti a ritenuta alla fonte e/o a contributi previdenziali e assistenziali dovuti all'Inps.

L'indagine sperimentale sulle dichiarazioni fiscali modd.770 ha richiesto lo sviluppo delle seguenti fasi: *a)* studio della fonte del dato modd.770; *b)* predisposizione del piano di campionamento, con successiva estrazione delle unità campionarie da rilevare (codici fiscali/partite IVA delle imprese); *c)* acquisizione del campione di imprese dall'archivio modd.770 fornito dalla Sogei; *d)* lettura dei dati contenuti in archivio; *e)* migrazione dell'archivio in un formato leggibile in SAS; *f)* analisi della qualità del dato; *g)* confronti con altre basi di dati (*benchmark*); *h)* calcolo dei coefficienti di riporto all'universo; *i)* stima degli importi aggregati di spesa non pensionistica; *l)* analisi dei dati e commento dei risultati.

In questa sede ci limiteremo ad illustrare le fasi *f)*, *g)* e *h)*.

In sintesi, dall'indagine sperimentale sui modd.770 sono emersi i seguenti risultati:

¹ Particolari ringraziamenti vanno al dott. P.D. Falorsi per i validi suggerimenti recepiti in fase di elaborazione delle stime del paragrafo 6 ed al dott. V. Spinelli per aver fornito informazioni utili alla stesura del paragrafo 3. Naturalmente eventuali errori e omissioni sono imputabili esclusivamente all'Autore.

² Trattasi di un campione di 22.446 imprese, da cui si ricavano informazioni sugli importi di prestazioni sociali erogate direttamente dai datori di lavoro per ciascun lavoratore dipendente.

³ Le liquidazioni di fine rapporto sono classificate in ambito nazionale ed internazionale tra le prestazioni sociali, anche se nel caso del TFR si tratta di un'erogazione avente un ruolo di carattere retributivo con funzione previdenziale, costituendo esso quella parte della retribuzione la cui corresponsione viene differita al momento della cessazione del rapporto di lavoro. Appare tuttavia non pienamente giustificata la classificazione del TFR tra le prestazioni sociali quando viene erogato in seguito a dimissioni volontarie del lavoratore o per quiescenza. In tali circostanze esso non rappresenta una risposta ad un rischio o bisogno sociale. Per un approfondimento degli aspetti di classificazione si rinvia a P. Consolini (1998).

- a) rispetto al campione di 22.446 imprese selezionate dall'Archivio Statistico delle Imprese Attive (ASIA), la Sogei ha rilasciato all'Istat informazioni su 21.724 imprese che hanno compilato almeno uno dei seguenti quadri: prospetto, quadro SA e quadro SB. Il tasso di caduta è stato pari al 3,2%; valore che cresce al 3,6%, se si escludono le 75 imprese campione che non producono informazioni sui quadri analitici SA e SB. In particolare il numero di imprese che hanno compilato il quadro SA è stato pari a 21.644, mentre il numero di quelle che hanno compilato il quadro SB è stato pari a 11.191. Ad eccezione di sei casi, tutte le imprese che hanno trasmesso informazioni per il quadro SB effettuano contestualmente dichiarazioni relative al quadro SA (cfr. figura 1);
- b) ambedue i quadri SA e SB sono inficiati da errori che hanno richiesto l'utilizzo di procedure di correzione automatica. Le tre tipologie di errore riscontrate sono così classificabili: *i*) duplicazione sui dati dei percipienti; *ii*) valori anomali sulle variabili quantitative di interesse (retribuzioni e prestazioni) *iii*) valori mancanti (variabili qualitative e quantitative). In particolare il quadro SA ha presentato 1.810 "record duplicati" (record con dati esattamente ripetuti su tutte le variabili selezionate) e 1.791 "record parzialmente duplicati", cioè contenenti dati non tutti ripetuti sulle variabili esaminate riferite allo stesso percipiente (trattasi di dichiarazioni integrative/sostitutive). Queste ultime potrebbero generare errori di doppio conteggio sulle variabili quantitative di interesse, qualora fossero riportate informazioni ripetute per uno stesso individuo (identificato attraverso il corrispettivo codice fiscale). A sua volta, il quadro SB ha presentato 1.179 record duplicati e 6.792 record parzialmente duplicati rispetto allo stesso percipiente. Riguardo al punto *ii*) il quadro SA ha presentato 108 record con valori anomali (*outlier*) sulla variabile "importi retribuzioni ridotte"⁴ su un totale di 72.062 osservazioni valorizzate (0,15%). Sebbene i record coinvolti risultino inferiori al 2% essi determinano un innalzamento della spesa per retribuzioni ridotte di 93,4 miliardi di lire, pari al 52,9% del totale della spesa calcolata sulle retribuzioni ridotte non corrette. A ciò si aggiungono 932 casi in cui le "retribuzioni ridotte" (settimanali) assumono valori inconsistenti, cioè superano il limite massimo definito dalle retribuzioni complessive⁵ (settimanali). Anche in quest'ultima circostanza si produce un rigonfiamento della spesa per retribuzioni ridotte (circa 2 miliardi di lire). In relazione al punto *ii*) il quadro SB ha evidenziato 43 record contenenti valori anomali sugli importi delle liquidazioni rispetto a un totale di 75.431 osservazioni (0,06%). Questi ultimi determinano una crescita spropositata della spesa in questione, pari a 312,8 miliardi di lire su un importo complessivo di 769,2 miliardi calcolati sulla variabile originaria (41%);
- c) il confronto tra la fonte fiscale modd.770 e la fonte di benchmark ASIA'99 ha prodotto risultati molto soddisfacenti. Vi è una buona concordanza tra l'archivio fiscale modd.770 e la fonte ASIA (fonte ufficiale ISTAT) rispetto al numero medio annuo di lavoratori per impresa (calcolabile sulle 21.618 imprese campione che hanno fornito informazioni su lavoratori dipendenti all'interno del quadro SA). Infatti, prendendo a riferimento il corpo della distribuzione (comprendente il 90% dei casi) delle differenze assolute e relative tra il numero dei dipendenti medi rilevati nell'archivio modd.770 e quelli forniti dal registro statistico delle imprese (ASIA) per diverse classi di dipendenti, si notano scostamenti poco rilevanti. Nella classe 1-9 di dipendenti (di cui fanno parte 16.915 imprese campione) i valori del range "5° percentile-95° percentile" sono pari a (-0,22; +0,58) e a (-9,6%; +34,5%) rispettivamente per la distribuzione degli scostamenti assoluti e relativi. Nella classe 10-19 di dipendenti (1.691

⁴ La "retribuzione ridotta" consiste nei versamenti in denaro corrisposti al dipendente non come contropartita di una prestazione lavorativa, ma come trasferimenti senza contropartita (prestazioni sociali) durante il periodo di assenza dal lavoro del dipendente per malattia (generica o speciale), infortunio sul lavoro e maternità, in carenza di tutela previdenziale o in miglioramento del livello delle prestazioni di base erogate dagli enti di previdenza, così come disciplinato dai CCNL.

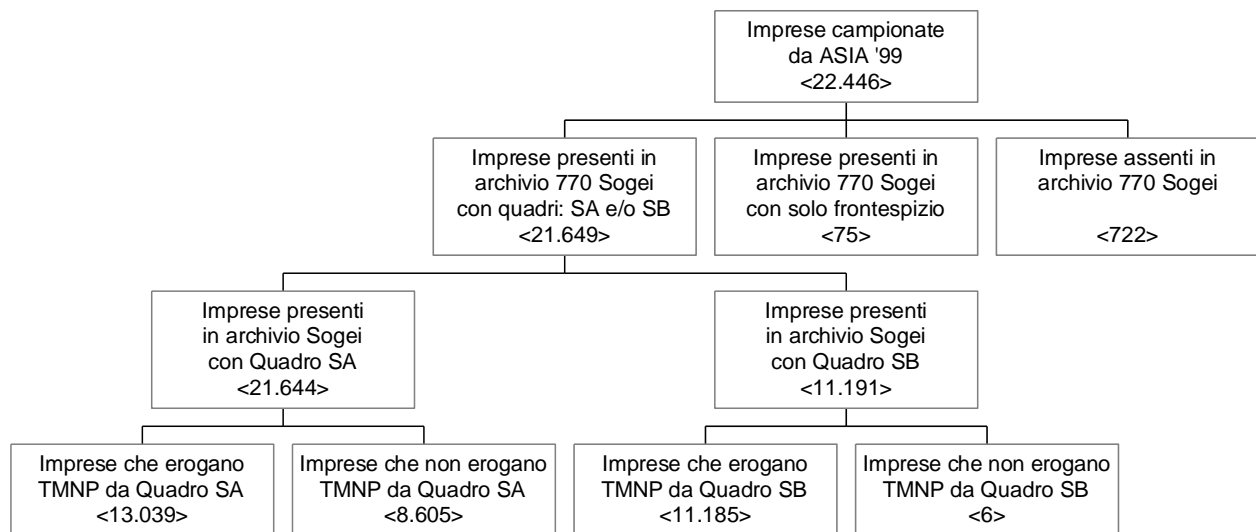
⁵ Per retribuzione complessiva si intende l'importo complessivo delle retribuzioni mensili dovute nell'anno solare, sia intere che ridotte. La retribuzione complessiva viene definita nel modello 770 con il termine competenze correnti ed include le seguenti componenti: stipendio base, contingenza, competenze accessorie.

imprese campione) si registrano nei due casi valori del *range* rispettivamente pari a (-0,89;+1,76) e a (-5,9%;+12,5%). Nella classe 20-49 dipendenti (839 imprese campione) l'intervallo tra i due percentili è pari a (-2,25;+5,00) e a (-6,8%;+15,8%). Nella classe 50-99 dipendenti (212 imprese campione) gli intervalli di cui sopra sono pari a (-3,33;+15,25) e a (-5,0%;+22,3%). Infine nella classe 100 e oltre dipendenti (182 imprese campione) si rilevano intervalli pari a (-51,4;+32,2) nelle differenze assolute e *range* pari a (-18,2%; +14,5%) nelle differenze relative. Il numero complessivo di dipendenti presenti in media nell'anno è pari rispettivamente a 168.788 unità secondo la fonte ASIA e a 170.533 unità per la fonte fiscale modd.770 (+1,03%);

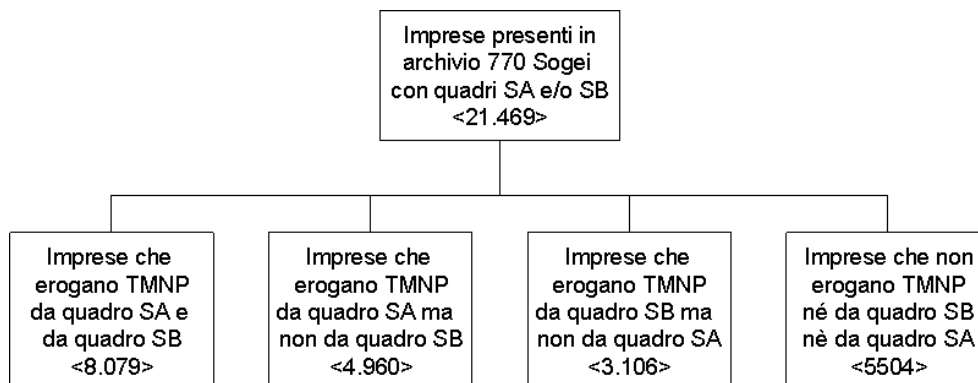
- d) delle 21.644 imprese che hanno compilato il quadro SA, ben 21.618 hanno erogato retribuzioni. Esse hanno corrisposto redditi a favore di 256.567 lavoratori (codici fiscali distinti) e disposto di un numero di dipendenti medi nell'anno pari a 168.826 unità. Le imprese che hanno erogato retribuzioni ridotte sono in numero pari a 13.039 e trasferiscono una spesa corretta di 33,2 milioni di euro in favore di 62.553 *beneficiari*;
- e) l'insieme delle imprese che ha compilato il quadro SB è pari a 11.191, di esse 11.185 erogano liquidazioni di fine rapporto in favore di 68.459 beneficiari, per un importo complessivo pari a 237,8 milioni di euro.

Figura 1. - Struttura di acquisizione dei dati 770 - Sogei

A) Imprese campionate ed imprese rilevate per quadro di riferimento



B) Classificazione delle imprese per erogazione di trattamenti monetari non pensionistici (TMNP) e quadro di riferimento



2. La struttura dei modelli 770-1999 del Ministero dell'Economia e delle Finanze

L'analisi delle informazioni sui trattamenti non pensionistici raccolte dal modello ha inizio con le dichiarazioni relative al 1998. A partire da quell'anno è stata infatti modificata la procedura di registrazione dei dati previdenziali ed è stato introdotto nel *mod.770* un apposito riquadro, specifico per lavoratore dipendente, dove devono essere riportati tutti gli elementi necessari all'INPS per l'accredito reale e figurativo dei contributi dovuti. L'innovazione ha avuto come conseguenza che, per la parte previdenziale, la dichiarazione del sostituto d'imposta effettuata col *mod.770* ha sostituito di fatto quella precedentemente trasmessa all'INPS attraverso i *mod.O1/M* e *mod.O3/M*; i dati retributivi e contributivi che il lavoratore precedentemente riceveva con una copia del *mod.O1/M* sono invece ora trasmessi con il *mod.CUD*⁶.

La dichiarazione *mod.770* si compone di un modello base con i dati identificativi del dichiarante (datore di lavoro)⁷ e di 21 quadri staccati.

Tra questi, quelli più interessanti ai fini dell'acquisizione dei dati utili per l'indagine sulle prestazioni monetarie non pensionistiche sono i primi due, denominati rispettivamente quadro SA e quadro SB. Nel primo sono contenute le informazioni relative ai redditi da lavoro dipendente e ai redditi assimilati, mentre nel secondo sono raccolti i dati relativi al trattamento di fine rapporto di lavoro dipendente e alle indennità equipollenti, alle altre indennità erogate in occasione della cessazione del rapporto, nonché alle prestazioni in forma di capitale corrisposte dai regimi pensionistici complementari.

In quanto segue verrà data una breve descrizione dei dati attinenti esclusivamente questi due quadri, rimandando alle istruzioni per la compilazione del *modello 770/2000*, fascicolo redatto dalla stessa Agenzia delle Entrate del Ministero dell'Economia e delle Finanze, per una completa esposizione di tutte le informazioni contenute nel modello.

Il **quadro SA** viene compilato dal datore di lavoro per ciascun lavoratore o assicurato e ricalca in parte il contenuto informativo del precedente *mod.O1/M*, da cui tuttavia si differenzia per la natura di alcune informazioni di carattere demografico rilevate sull'assicurato e per una maggiore quantità di informazioni in esso contenute. Il quadro è suddiviso a sua volta in cinque riquadri denominati: dati identificativi; dati fiscali; dati previdenziali ed assistenziali INPS e INPDAI; dati previdenziali e assistenziali INPDAP; dati assicurativi relativi all'INAIL.

Nel riquadro dati identificativi (ex-quadro A del *mod.O1/M*) sono riportate le informazioni anagrafiche (cognome e nome, codice fiscale, sesso, luogo e data di nascita, domicilio) e la qualifica professionale dell'assicurato. Quest'ultima informazione, collocata in un punto differente rispetto al precedente *mod.O1/M*⁸, permette di rilevare la qualifica assicurativa, le condizioni contrattuali, la categoria particolare e l'ulteriore categoria rilevata ai soli fini INAIL sul lavoratore assicurato. Pur se in linea di massima la tipologia dei dati anagrafici registrati nel *mod.770* è analoga a quella riportata nel *mod.O1/M*, sussistono alcune differenze: in luogo delle informazioni sulla residenza del *mod.O1/M* vengono ora registrate le informazioni sul domicilio⁹ e al posto della

⁶ Con l'entrata in vigore dei D.Lgs. n. 314/97 e n.241/97 in materia di certificazioni e di dichiarazioni unificate a fini fiscali e previdenziali, a decorrere dal 1998 i modelli O1M e O3M sono di fatto decaduti e sostituiti dal quadro SA e SS del *mod.770* (dichiarazione del sostituto d'imposta) per i dati da inviare all'INPS e il *mod.CUD* (Certificazione Unica del Dipendente) per i dati da inviare al lavoratore. Per maggiori dettagli sull'argomento si rimanda a P.Consolini e R. De Carli (2002).

⁷ I dati identificativi del dichiarante sono riportati nel frontespizio, mentre sono collocati nella seconda facciata i riquadri relativi al tipo di dichiarazione, al dichiarante, al rappresentante o altro soggetto che sottoscrive la dichiarazione, alla scelta dell'utilizzo delle eccedenze di versamenti, alla sottoscrizione della dichiarazione, alle modalità di presentazione e all'apposizione del visto di conformità.

⁸ Nell'intestazione del *mod.O1/M* viene riportato il solo codice individuale INPS e la matricola INPS dell'azienda, mentre il codice relativo alla qualifica viene riportato nel quadro B dello stesso modello.

⁹ Secondo l'art.43 del codice civile, per domicilio si intende il luogo dove la persona ha stabilito la sede principale dei suoi affari e interessi, mentre per residenza si intende il luogo ove essa vive abitualmente.

nazionalità viene riportata la cittadinanza¹⁰ del lavoratore assicurato. Se per quest'ultima informazione si può assumere una adeguata rispondenza ai principi di continuità storica dei dati raccolti per gli anni successivi al 1998 rispetto al vecchio *mod.OI/M*, non altrettanto si può dire per il domicilio del lavoratore che può essere diverso rispetto alla residenza.

Il riquadro dati fiscali contiene le informazioni relative alla tassazione dei redditi corrisposti nell'anno di imposta e alle correlate detrazioni fiscali¹¹. Le informazioni contenute in questo quadro, integrative rispetto al precedente *mod.OI/M*, pur se di notevole interesse dal punto di vista di una analisi descrittiva degli effetti redistributivi della politica fiscale, non risultano tuttavia strettamente utilizzabili ai fini del sistema di rilevazione in oggetto, non rientrando le detrazioni per carichi familiari nel campo delle prestazioni previdenziali e assistenziali così come da definizione ESSPROS¹².

Il riquadro dati previdenziali ed assistenziali INPS e INPDAI è a sua volta suddiviso in tre sezioni, che ricalcano rispettivamente i quadri B, C e D del *mod.OI/M*, in cui vengono registrati i dati relativi alla contribuzione previdenziale e assistenziale dovuta all'INPS e/o all'INPDAI. Come nel quadro B del *mod.OI/M*, ma con una quantità e una qualità dell'informazione maggiore¹³, nella prima sezione vengono indicati i dati relativi all'anno di contribuzione e al rapporto di lavoro¹⁴, l'accantonamento sul fondo trattamento fine rapporto complessivamente spettante, e, di particolare importanza ai fini dell'indagine, l'eventuale erogazione di assegni al nucleo familiare con le relative coordinate.

La seconda e la terza sezione, di cui la prima è denominata "retribuzioni particolari", contengono lo stesso tipo di dati rilevate nei quadri C e D del *mod.OI/M*; ai fini dell'indagine forniscono, allo stesso livello di dettaglio¹⁵ del precedente modello, le informazioni sull'indennità

¹⁰ Per cittadinanza deve intendersi la condizione giuridica di chi appartiene ad uno Stato ed è titolare dei diritti politici che si esercitano secondo il suo ordinamento e degli obblighi relativi. La legge sulla cittadinanza è la n.555 del 1912, modificata dalla L. n.123/1983. I motivi di acquisizione della cittadinanza possono ricondursi a quattro ipotesi: acquisizione per nascita, per estensione del diritto (*iuris communicatio*), per beneficio di legge e per naturalizzazione. Per nazionalità si intende l'appartenenza ad una Nazione, concetto questo di origine settecentesca affermatosi nell'ottocento che si riferisce all'insieme di genti legate da comunanza di tradizioni storiche, lingua, costumi ed aventi coscienza di tali comuni vincoli. L'insieme dei cittadini appartenenti ad uno Stato non si identifica necessariamente con il concetto di Nazione: ne è la riprova la presenza di numerosi stati plurinazionali che si affiancano agli Stati uninazionali come il nostro.

¹¹ In particolare in esso viene indicato il numero dei giorni per i quali spetta il credito di imposta, il numero dei giorni preso a base per il calcolo delle detrazioni per redditi da lavoro dipendente e i relativi importi, gli importi dei redditi per cui non è possibile usufruire di dette detrazioni, il totale delle detrazioni di imposta attribuite, le detrazioni per carichi familiari, le ritenute effettuate in eccedenza, gli importi dell'addizionale regionale all'Irpef, dell'addizionale comunale all'Irpef, del contributo per l'Europa restituito e da restituire al percipiente (per l'anno di imposta 1999), il totale degli oneri detraibili e il totale degli oneri deducibili.

¹² È previsto tuttavia lo sviluppo di un modulo "prestazioni fiscali" che verrà in futuro affiancato al "core" del ESSPROS. Cfr. ESSPROS, Manual 1996, Luxemburg.

¹³ In particolare, se alcune variabili trovano soltanto una diversa collocazione - la matricola d'azienda viene indicata in questo riquadro, ove nel *mod.OI/M* veniva riportata nell'intestazione - altre sono aggiuntive o sostitutive rispetto al precedente modello. Tra le assicurazioni coperte, è prevista qui anche la casella con l'indicazione esplicita dell'iscrizione al Fondo garanzia trattamento fine rapporto; accanto al codice contrattuale va ora indicato anche il tipo di contratto - stipulato a livello regionale, provinciale, aziendale, o altrimenti - qualora non sia applicato un contratto nazionale (salvo il caso in cui il contratto del lavoratore è relativo agli enti pubblici); per gli assegni erogati al nucleo familiare, in luogo della variabile binaria (SI, NO) riportata nel *mod.OI/M*, vanno ora fornite tutte le coordinate relative alla corresponsione degli assegni da parte del datore di lavoro - il numero della tabella riferita alla tipologia del nucleo familiare utilizzata per la determinazione dell'assegno spettante, il numero dei componenti del nucleo e il numero progressivo che individua la fascia di reddito dello stesso nucleo.

¹⁴ In particolare: la matricola d'azienda; la provincia di lavoro; le assicurazioni coperte; le competenze correnti e le altre competenze; i mesi, le settimane e il numero di giornate lavorative retribuite nell'anno; il codice e la tipologia del contratto; il livello di inquadramento; l'eventuale data di cessazione del rapporto di lavoro; il tipo di rapporto, l'eventuale trasformazione del rapporto di lavoro; il numero di settimane riferite al rapporto a tempo parziale utili per la determinazione della misura delle prestazioni pensionistiche (settimane utili).

¹⁵ Nella sezione II vengono indicati distintamente: il codice della retribuzione particolare, il periodo cui questa si riferisce, l'importo della retribuzione corrisposta, il numero delle settimane e delle giornate lavorative retribuite, il numero

sostitutiva del preavviso e sull'importo e la durata della "retribuzione ridotta" a causa di malattia indennizzata e infortunio sul lavoro, malattia specifica L88/87, maternità obbligatoria e facoltativa, CIG¹⁶ ordinaria e straordinaria e donazione di sangue. Il dato economico sulle retribuzioni ridotte verrà analizzato nei successivi paragrafi 4 e 6.

Nel computo delle settimane caratterizzate da retribuzione ridotta vengono considerate quelle in cui, anche per un solo giorno, non è stata percepita la retribuzione o è stata percepita in misura ridotta a causa di detti eventi, comprese le settimane relative a eventi indennizzati direttamente dall'INPS. Separatamente è indicato, per ogni evento, il numero totale annuo delle settimane per le quali il lavoratore non ha percepito alcuna retribuzione ridotta dal datore di lavoro ("sett.1") e quelle invece caratterizzate da una retribuzione ridotta ("sett.2"), tralasciando in entrambi i casi gli eventi di durata inferiore a 7 giorni. Gli eventi di malattia o di infortunio di durata inferiore ai tre giorni, cd. periodo di carenza, sono desumibili dalle informazioni riportate in questa sezione del modello: in concomitanza con questi eventi viene infatti indicata l'occorrenza di una settimana di retribuzione ridotta e l'importo della retribuzione ridotta corrisposta, mentre vengono lasciate in bianco le caselle "sett.1" e "sett.2" sopra descritte. Non è tuttavia possibile risalire alla durata esatta (numero di giorni) degli eventi protetti se non attraverso una stima ottenuta dal confronto tra la retribuzione ridotta corrisposta e la retribuzione media giornaliera desumibile dai dati riportati nella sezione I del quadro ("competenze correnti" diviso "giorni retribuiti").

Nel riquadro dati previdenziali e assistenziali INPDAP sono contenuti i dati contributivi per gli iscritti all'Istituto Previdenziale. Nel riquadro rientrano tutti i redditi da lavoro dipendente imponibili secondo la vigente normativa¹⁷ relativi all'anno di riferimento della dichiarazione. Ai fini dell'indagine, tra i dati contenuti nel riquadro sono utilizzabili le informazioni sulla corresponsione di indennità sostitutiva del preavviso a carattere risarcitorio: a differenza di quanto avviene per l'indennità sostitutiva del preavviso a carattere retributivo, il cui emolumento entra a far parte della retribuzione fondamentale e viene indicato congiuntamente a questa nella apposita casella denominata "retribuzioni fisse e continuative", l'amministrazione (datore di lavoro) è tenuta ad indicare l'inizio e la fine del periodo (ultimo mese di attività) cui si riferisce l'indennità una tantum erogata a carattere risarcitorio e il relativo ammontare.

L'ultimo riquadro è predisposto per la registrazione dei dati assicurativi relativi all'INAIL, da cui prende la denominazione, e fornisce la posizione assicurativa e il relativo controcodice, i periodi di inclusione del soggetto nella posizione assicurativa di riferimento e gli eventuali periodi in cui l'assicurato ha svolto nello stesso anno attività lavorative riconducibili a diverse posizioni assicurative riferite alla stessa azienda.

Nel **quadro SB** sono indicati i dipendenti ai quali sono stati corrisposti trattamenti di fine rapporto di lavoro dipendente e altre indennità¹⁸ soggette a tassazione separata: indennità di buonuscita, indennità di mancato preavviso, trattamento di fine rapporto e indennità equipollenti percepite da parte degli aventi diritto del lavoratore defunto.

Anche qui, come per il precedente quadro SA, sono indicati il codice fiscale e i dati anagrafici dei beneficiari (nome, cognome, sesso, luogo e data di nascita, il domicilio), la data di inizio e di

dei giorni utili per la determinazione della misura delle prestazioni pensionistiche. Nella sezione III, compilato in relazione agli eventi di malattia o infortunio, malattia ex L. n.88/87 (TBC), donazione di sangue (L. n.107/90), maternità e cassa integrazione guadagni, viene indicato per ogni lavoratore per il quale questi eventi si sono verificati: il totale annuo delle settimane di calendario caratterizzate da una retribuzione ridotta, con riferimento anche al periodo di carenza per il quale l'evento di malattia o infortunio ha avuto durata inferiore ai 3 giorni; l'importo complessivo annuo delle retribuzioni corrisposte nelle settimane di retribuzione ridotta, i periodi di sospensione e/o riduzione dell'orario di lavoro per i quali è stata autorizzata la corresponsione dell'integrazione salariale; l'ammontare complessivo delle retribuzioni che sarebbero spettate al lavoratore se avesse lavorato normalmente.

¹⁶ Nella casella relativa alla CIG vengono indicate le settimane nelle quali è intervenuto il contratto di solidarietà.

¹⁷ D.Lgs. n.314/97 e Circ. n.29/98.

¹⁸ Cfr. art.16 co.1 lett. a del Tuir. Lo stesso quadro contiene informazioni relative alle prestazioni in forma di capitale per le forme pensionistiche complementari corrisposte nell'anno di riferimento della dichiarazione ai lavoratori dipendenti o soci di cooperative di produzione lavoro iscritti alle predette forme assicurative anteriormente al 28 aprile 1993.

cessazione del rapporto di lavoro. Sia per il trattamento fine rapporto che per le altre indennità corrisposte nel 1999 è specificato se si tratta di anticipazione, saldo o acconto, e se queste somme sono state erogate per incentivare l'esodo dei dipendenti (codice D,E caselle 19, 23 denominate "titolo"), caso in cui andrebbero presumibilmente classificate nella funzione disoccupazione piuttosto che in quella vecchiaia. Tra i dati riportati nel quadro sono incluse le informazioni sull'erogazione delle indennità agli eredi in caso di morte del lavoratore dipendente¹⁹ (art.2122 c.c.), tra cui i dati anagrafici del lavoratore deceduto e di ciascun avente diritto o erede, con indicazione separata delle somme eventualmente corrisposte in corso d'anno sia al deceduto che agli eredi. Si fa presente peraltro che gli importi erogati ai superstiti per trattamento fine rapporto del lavoratore deceduto rientrano, ai fini dell'indagine, nella funzione superstiti. Il dato economico sulle liquidazioni di fine rapporto viene analizzato nei successivi paragrafi 5 e 6.

¹⁹ La legge non prevede disposizioni specifiche nelle eventualità in cui il TFR, a seguito del decesso del lavoratore, debba essere corrisposto agli aventi diritto ai sensi dell'art. 2122 c.c. Le indicazioni dell'Agenzia delle Entrate del Ministero dell'Economia e delle Finanze suppliscono a tale carenza informativa e per la determinazione del reddito imponibile e della relativa imposta.

3. Le procedure di sistemazione automatica dei record in un formato fisso leggibile in SAS

L'acquisizione degli archivi mod.770 dalla SOGEI ha comportato la creazione di due archivi sequenziali (relativi a due acquisizioni successive). La SOGEI ha utilizzato un formato standard per l'acquisizione dei dati e quindi le informazioni utili per le elaborazioni devono essere ricostruite secondo un formato più utile agli scopi del processo.

Gli archivi SOGEI sono caratterizzati da record aventi una struttura formata da una parte fissa ed una variabile; questa organizzazione è giustificata dal fatto che le informazioni effettivamente inserite durante la compilazione di un modello 770 sono generalmente una piccola percentuale rispetto a quelle possibili. Un tracciato fisso sarebbe un enorme spreco di spazio in questo caso, quindi si memorizzano solo i valori effettivamente presenti; non potendo prevedere quali informazioni sono effettivamente presenti nell'archivio, non vengono assegnate a priori posizioni specifiche nell'archivio. Per riconoscere i singoli campi si utilizza una struttura di campo virtuale o metacarpo: (<codice campo><tipo campo><lunghezza campo><valore campo>).

Questo metodo di rappresentazione dell'informazione viene comunemente indicata come "codici a serrare".

L'organizzazione degli archivi descritta non è immediatamente leggibile con gli ambienti più utilizzati in Istituto (es. SAS); da qui nasce l'esigenza di ristrutturare gli archivi.

L'analisi degli archivi SOGEI e dei modelli 770 ha portato alla creazione di tre archivi strutturati secondo le esigenze delle analisi statistiche da impostare: quadro A (frontespizio/anagrafica azienda), quadro SA e quadro SB.

I processi di costruzione dell'archivio "quadro A" si rivolgono principalmente alla parte fissa del tracciato e riescono quindi ad essere semplici nella loro struttura.

Al contrario i processi relativi ai quadri SA ed SB, si rivolgono quasi esclusivamente alla parte variabile del tracciato; ciò ha comportato la definizione di due fasi di ricostruzione delle informazioni relative all'unità di analisi (*impresa, dipendente*) ed alla sua validazione.

La prima fase ha il compito di agganciare all'unità di analisi le informazioni presenti in SA ovvero SB ma non in un ordine stabilito a priori. Questa azione di ricostruzione si è resa necessaria avendo verificato (valutazione empirica degli archivi SOGEI) che le informazioni logiche relative ad un dipendente non erano contenute su un solo record fisico e/o contigue tra di loro; inoltre è stato necessario fissare dei criteri per decidere quando le informazioni estratte dall'archivio appartengono ad uno stesso dipendente, in quanto non sempre il codice fiscale è "un separatore" affidabile all'interno dei dati di una singola azienda.

La fase successiva a quella descritta (ricostruzione delle informazioni) è quella relativa alla validazione: con essa si decide, in modo automatico, se le informazioni estratte possono essere rilasciate (scrittura negli archivi finali). Le verifiche si basano sostanzialmente sull'ipotesi che ad ogni unità di analisi (azienda, dipendente) devono essere agganciate informazioni significative.

Fornendo dei pesi ad ogni valore che si può associare ad una unità di analisi, il record ricostruito viene rilasciato se si raggiunge un punteggio minimo;

Queste fasi di ricostruzione e validazione hanno permesso di ridurre significativamente il numero di record incompleti e/o non significativi risultanti da una *estrazione passiva* degli archivi SOGEI.

Negli archivi ricostruiti è stata aggiunta l'informazione relativa all'anno di riferimento dei dati, in modo da poter costruire un archivio storico dei modelli 770.

4. Le procedure di controllo e correzione dei dati relativi al quadro SA

Le procedure di controllo e correzione dei dati sul quadro SA hanno riguardato i seguenti aspetti:

- 1) verifica della presenza di valori mancanti (*missing value*) nelle 12 variabili di interesse (TMNP) e selezione di *record* contenenti almeno un'informazione di interesse valorizzata;
- 2) verifica della presenza di *record* duplicati (sovrapposizioni totali) e relativa eliminazione;
- 3) verifica della presenza di valori anomali (*outlier*) sulle variabili economiche di interesse e creazione di procedure per l'imputazione di valori corretti;
- 4) verifica della presenza di mancate risposte su alcune variabili di interesse e implementazione di procedure di imputazione del dato mancante;
- 5) verifica della presenza di record "parzialmente duplicati" e calcolo dei valori assunti dalle variabili economiche di interesse a livello di singolo percipiente;
- 6) controllo dei dati relativi alla variabile retribuzioni ridotte e analisi delle variabili ausiliarie: a) durata dell'evento malattia (variabili: *sett.1* e *sett.2* associate alla malattia); durata dell'evento maternità (variabili: *sett.1* e *sett.2* associate alla malattia); differenze retributive per Cigo-Cigs;

L'unità di rilevazione (*record*) del quadro SA (come pure del quadro SB) relativo ai modd. 770 è la "dichiarazione fiscale" a cui è tenuto il datore di lavoro (sostituto d'imposta) per aver corrisposto somme in denaro a favore di percipienti (identificati col relativo codice fiscale). L'unità di analisi dell'indagine relativa ai trattamenti monetari non pensionistici è il "beneficiario" delle prestazioni erogate da una determinata impresa. Di regola vi è una sola dichiarazione associata a ciascun beneficiario dipendente da una determinata impresa. Tuttavia, con riguardo al quadro SA, esistono delle eccezioni dovute al fatto che nel corso dell'anno il lavoratore può cambiare qualifica assicurativa, condizioni contrattuali ovvero può essere iscritto a più Enti previdenziali. In queste circostanze il datore di lavoro è tenuto alla compilazione di tanti quadri SA quante sono le corrispondenti variazioni. Oltre a questi aspetti formali, un datore di lavoro può in realtà fornire più dichiarazioni per uno stesso soggetto, qualora si rendesse conto che i dati trasmessi in precedenza risultino errati.

La prima fase di selezione delle sole imprese campione erogatrici di TMNP ha comportato l'eliminazione di 8.605 imprese con annessi 227.600 record. Le restanti 13.039 imprese accludevano 95.763 record.

La fase di controllo sui record duplicati ha fatto emergere 1.810 record riportanti valori esattamente identici sugli stessi soggetti (sovrapposizione totale), con una percentuale pari all'1,9% (cfr. tabella 4.1). Di conseguenza si è proceduto all'eliminazione dei record ridondanti.

Tabella 4.1 - Le fasi preliminari di trattamento dei dati del quadro SA

CARATTERISTICHE DELL'ARCHIVIO: UNICO-SEPARATO 770	Quadro SA			
	Numero di record	Numero di imprese	Numero di beneficiari con almeno 1 TMNP	Note
Data-set in ingresso	323.363	21.644		53.703 record si riferiscono all'unico; 269.660 record si riferiscono alla dichiarazione separata
<i>Dati che non contengono valori in nessuna delle 12 variabili TMNP (missing)</i>	227.600	8.605	0	
Data-set in ingresso al netto dei missing sulle 12 variabili TMNP	95.763	13.039	92.162	
<i>Numero di record duplicati sul data-set senza missing</i>	1.810			
Data-set senza duplicati e senza missing sui TMNP: "Data-set finale"	93.953	13.039	92.162	La differenza tra il nr. di record e il nr. di beneficiari (1791) fornisce il nr. di dichiarazioni in aggiunta alla prima
- <i>di cui con 1 dichiaraz. per lo stesso beneficiario (1 record)</i>	90.404		90.404	
- <i>di cui con 2 dichiaraz. per lo stesso beneficiario (1 record)</i>	3.456		1.728	
- <i>di cui con 3 dichiaraz. per lo stesso beneficiario (1 record)</i>	81		27	
- <i>di cui con 4 dichiaraz. per lo stesso beneficiario (1 record)</i>	12		3	

4.1 La correzione ed imputazione dei valori delle retribuzioni ridotte e delle competenze correnti

Per l'identificazione e la correzione degli *outlier* relativi agli importi delle retribuzioni ridotte è stato utilizzato un approccio di tipo euristico, ovvero sono state analizzate le tipologie di errori presenti nei dati appartenenti agli estremi della distribuzione dei valori positivi. Tale gruppo di osservazioni ha offerto utili suggerimenti per l'implementazione delle procedure di correzione. Ad esempio, una tipologia di errore rilevata di frequente tra i valori più elevati, è costituita da importi di "retribuzioni ridotte" superiori ai 100 milioni di lire e con le tre cifre finali pari a zero. L'elenco di tutti i potenziali errori sulle retribuzioni ridotte e le conseguenti azioni di correzione sono illustrate nella tabella 4.1.1.

Come si può notare dai valori riportati nell'ultima colonna "*Differenze tra valori originari e corretti*", gli errori che hanno maggiore impatto (positivo o negativo) sulla spesa delle retribuzioni ridotte sono rappresentati dai codici: 2, 4, 8 e 12. In particolare la tipologia di errore "2" è costituita tanto dagli *outlier* posizionati nell'estremo superiore della distribuzione delle retribuzioni ridotte (oltre il miliardo di lire) quanto da valori comunque elevati nelle retribuzioni ridotte (oltre i 100 milioni di lire) che presentano tre zeri finali. Alla presenza di un errore codificato con "2" si procede a correggere il valore originario della retribuzione ridotta dividendolo per un fattore 1000, cioè si ipotizza che la dichiarazione sia stata registrata in lire e non in migliaia di lire. Gli errori codificati con "4" identificano record che presentano contemporaneamente anomalie sulle

retribuzioni ridotte (codici 1, 3)²⁰ ed errori sulle competenze correnti (codice 2). In questo caso il dubbio che vi sia un errore sulle retribuzioni ridotte, come indicato dai codici 1 e 3, si traduce in certezza a causa della presenza di errori (codice 2) sulla variabile competenze correnti relativa alla stessa dichiarazione. L'ipotesi di fondo è che il dichiarante commetta lo stesso tipo di errore (valori espressi in lire) su tutte le variabili economiche, per cui se in almeno una variabile si presenta un errore e nelle restanti vi sono casi sospetti allora tutte le variabili sono ritenute ugualmente errate.

Tabella 4.1.1 - Tipologie di errori su "retribuzioni ridotte" ed azioni conseguenti

Tipologia di errore su retribuzioni ridotte e azioni conseguenti (correzioni/imputazioni):	Codice errore	Nr. casi	% casi	Retribuzion i ridotte originarie migliaia lire	Retribuzion i ridotte corrette migliaia lire	Differenze tra Retr_rid_orig e Retr_rid_cor.
Retribuzioni ridotte di importo nullo	0	21.881	23,30	0	0	0
Retribuzioni ridotte di importo positivo con nessun errore	0	69.377	73,84	78.810.914	78.810.914	0
Retribuzioni ridotte con valori: < 100 milioni di lire e con tre "000" finali (nessuna azione)	1	41	0,04	85.000	85.000	0
Retribuzioni ridotte con valori: (superiori a 100 milioni di lire e tre "000" finali) o (superiori al miliardo di lire) (azione: valore corretto = valore originario /1.000)	2	86	0,09	80.643.515	80.644	80.562.871
Retribuzioni ridotte con valori compresi nell'intervallo: ("1 miliardo - 100 milioni" di lire e senza tre "000" finali) (nessuna azione)	3	0	0,00	0	0	0
Compresenza di errori: tipo "3 o 1" su Retrib. ridotte e tipo "2" sulle variabili Competenze_correnti e Differenze_CIG (azione: valore corretto = valore originario /1.000)	4	22	0,02	11.817.305	11.817	11.805.488
Viola il vincolo: (Retrib. ridotte > Compet. correnti > 0) e dove non si applica la correzione per l'errore di tipo 6 e 8 (azione: valore corretto = Competenze correnti)	5	33	0,04	145.339	90.972	54.367
Viola il vincolo: (1 milione di lire > Retr_rid_settim. > Compet_cor_settim. > 0) (azione: valore corretto = Compet_cor_set * Nr.sett_ret_rid)	6	37	0,04	108.561	78.998	29.563
Viola il vincolo: (Retrib_rid_settim. > 1 milione lire) & (Compet_cor_settim. = 0) (nessuna azione)	7	6	0,01	9.584	9.584	0
Viola il vincolo: (Retrib_rid_settim. > Comp_cor_settim. > 1 milione lire) (azione: valore corretto = Compet_cor_set * Nr.sett_ret_rid)	8	895	0,95	4.234.205	2.298.231	1.935.974
Retrib_ridotte con valori incoerenti rispetto a Compet_corr (azione: correzione ad hoc)	9	2	0,00	517.000	5.170	511.830
(Settimane_retrib_ridotta > 0) & (Retribuzioni_ridotte = 0) e (sett.2_malat_mater_tbc = 0) (nessuna azione)	10	641	0,68	0	0	0
(Settimane_retrib_ridotta > 0) & (Retribuzioni_ridotte = 0) & (sett.2_malat_mater_tbc > 0) & (compet_corr_set = 0) (nessuna azione)	11	14	0,01	0	0	0
(Nr.settim._retrib._ridotte > 0) e (Retribuzioni_ridotte = 0) e (sett.2_malat_mater_tbc > 0) e (Compet_corr_settim. > 0) (azione: valore corretto = Compet_corr_settim. * Nr.sett.2)	12	664	0,71	0	1.330.240	-1.330.240
Retribuzioni_ridotte non attribuibili: (Retrib_ridotte > 0) e (Nr.sett._retrib_rid=0) e (Nr.sett.2=0) & (Nr.sett.1=0) (nessuna azione)	13	164	0,17	61.490	61.490	0
Retrib_ridotte riattribuibili: (Retrib_ridotte > 0) & (Settim._retrib_rid=0) & (sett.2=0) & (sett.1 > 0) (nessuna azione)	14	25	0,03	10.883	10.883	0
Retrib_ridotte < 20 mila lire & (Comp_cor_set < 1.000 lire) (azione: valore corretto = valore originario * 1.000)	15	55	0,06	157	157.000	-156.843
Totale		93.953	100	176.443.953	83.030.942	93.413.011

²⁰ La presenza di casi identificati con i codici "1 (importi che non superano i 100 milioni di lire ma presentano tre zeri finali)" e "3 (importi compresi tra 100 milioni e il miliardo di lire ma senza tre zeri finali)" non determina di per sé alcuna azione di correzione ma segnala solo la presenza di un'anomalia nel dato ovvero una probabile situazione di errore.

In archivio si segnalano solo 3 casi in cui si presentano errori di tipo “2” sulla variabile competenze correnti senza che si evidenzino anomalie sulle retribuzioni ridotte (importi positivi): due record vengono corretti con correzioni “ad hoc” e sono codificati con “9”, il restante contiene valori ammissibili sulle retribuzioni ridotte e non viene pertanto modificato.

Per verificare la robustezza delle procedure implementate, si è voluto testare se gli errori di tipo “2” e “4” venivano commessi in blocco dalla stessa impresa su tutte le dichiarazioni dei propri percettori o se sussistevano casi singoli riferibili a beneficiari con elevate retribuzioni e/o qualifiche (cfr. tabella 4.1.2). Si è appurato che gli errori in questione coinvolgevano 30 imprese e 108 dichiarazioni di retribuzioni ridotte (valorizzate). Segnatamente 28 imprese commettevano gli stessi errori su tutti i record, mentre le restanti due mostravano percentuali di errore pari a 1/3 e a 2/3 delle dichiarazioni. La penultima impresa (casistica B) produce 2 dichiarazioni errate su un totale 6 dichiarazioni. Essa effettua tre doppie dichiarazioni riferite ad altrettanti beneficiari. Per due di essi l’impresa presenta un modulo con valori corretti e l’altro con valori moltiplicati per mille. Infine, l’impresa con una percentuale di errore pari al 66,7%, presenta 3 dichiarazioni per altrettanti percettori: la sola dichiarazione corretta contiene valori ammissibili sulle retribuzioni ridotte ed errori di tipo 2 sulle competenze correnti (cfr. caso citato in precedenza).

Tabella 4.1.2 - Distribuzione della percentuale di errori del tipo 2 e 4 (outlier) sulle dichiarazioni rilasciate delle imprese

CASISTICA	Numero di IMPRESE	Numero di DICHIARAZIONI	Percentuale di OUTLIER (flagr=2,4)	Numero di OUTLIER (flagr=2,4)
A	1	6	0,333	2
B	1	3	0,667	2
C	28	104	1,000	104
Totale	30	113		108

La seconda tipologia di errori in ordine di importanza è quella rappresentata dalle retribuzioni ridotte con importi “inammissibili”, cioè superiori in valore alle competenze correnti. Dal momento che le retribuzioni ridotte costituiscono una parte delle competenze correnti, deve essere soddisfatta la condizione: “le retribuzioni ridotte sono in valore inferiori o uguali alle competenze correnti”. Gli errori legati a quest’ultima condizione sono classificati con i codici: 5, 6, 7 e 8 (cfr. tabella 4.1.1). La tipologia più frequente (895 casi) è data dal codice “8” e corrisponde a retribuzioni medie settimanali superiori alle competenze correnti settimanali e di importo superiore al milione di lire. Imputando il valore corretto, definito dal prodotto fra le competenze correnti settimanali (limite superiore dei valori ammissibili) e il numero di settimane a retribuzione ridotta, si riduce la spesa in origine per le retribuzioni ridotte di circa 2 miliardi di lire.

La terza tipologia di errori è data dalla compresenza di “importi nulli sulle retribuzioni ridotte” e di “un totale annuo positivo di settimane a retribuzione ridotta (sett.2)” per gli eventi di malattia, tbc e maternità. Identificati con il codice “11” e “12”, questi errori richiedono un’apposita procedura di imputazione. Quando si è in grado di calcolare le competenze correnti settimanali, ovvero quando queste ultime assumono valori positivi (tipologia “12”), si procede ad imputare il valore delle retribuzioni ridotte con il prodotto dato dalle competenze correnti settimanali per il totale annuo di settimane a retribuzione ridotta di malattia, tbc e maternità. Diversamente, non si procede con alcun tipo di imputazione (tipologia “11”).

In ultimo, si fa notare la presenza di errori nella coda inferiore della distribuzione delle retribuzioni ridotte. Segnatamente riguardano i casi (tipologia 15) in cui da un lato, gli importi delle retribuzioni ridotte sono compresi nell’intervallo 1-20 mila lire e, dall’altro, le competenze correnti settimanali assumono valori positivi inferiori alle mille lire. A queste variabili si imputa un fattore

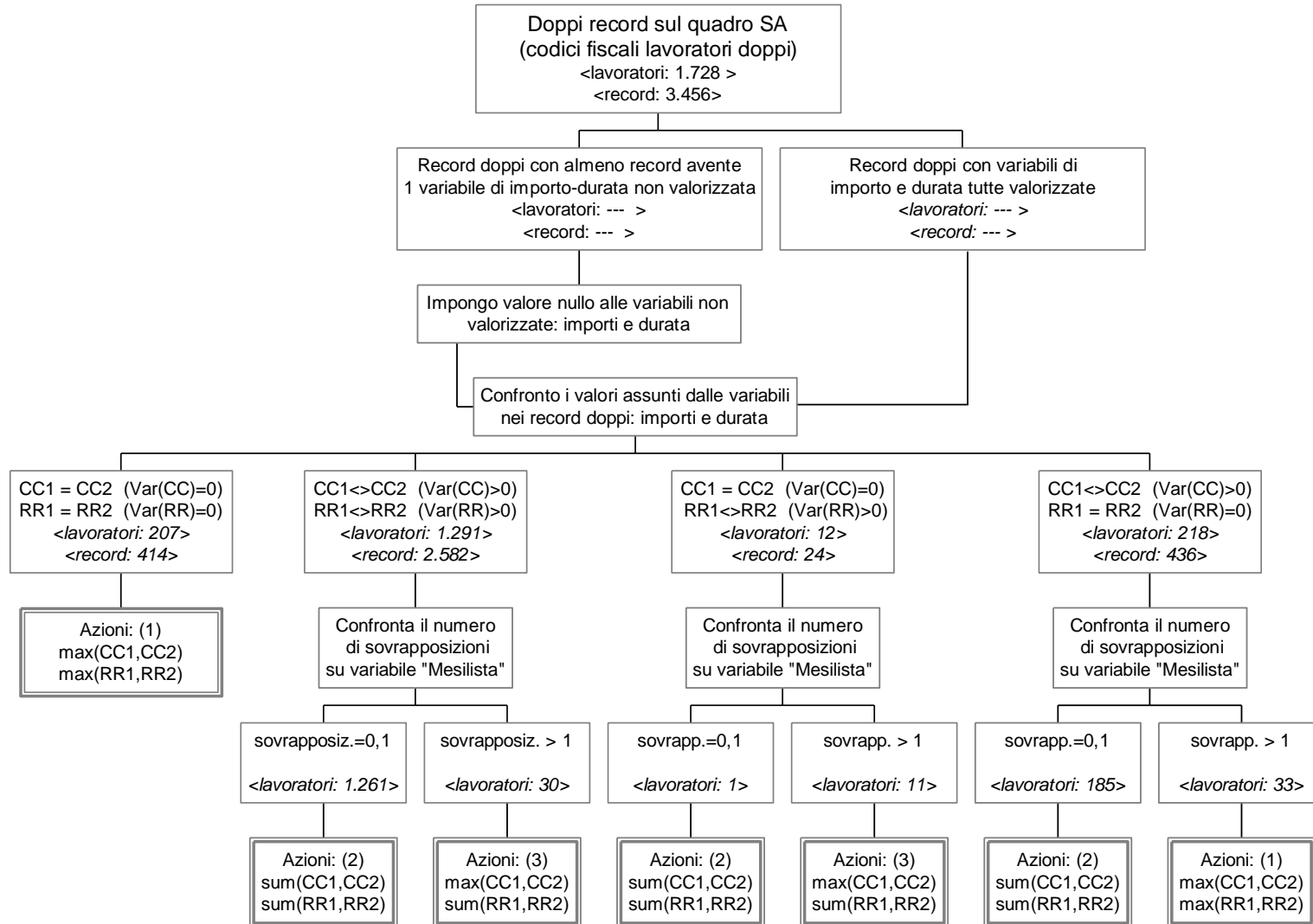
di correzione costante dato da 1.000, sottintendendo che il datore di lavoro abbia fornito dati economici espressi in milioni di lire anziché in migliaia.

4.2 *Le procedure di correzione degli errori di duplicazione parziale sulle retribuzioni ridotte e sulle competenze correnti*

Una volta eseguite le procedure di correzione dei valori assunti dalle retribuzioni ridotte, la fase successiva consiste nell'identificazione delle dichiarazioni doppie o plurime sulle retribuzioni ridotte riferite allo stesso beneficiario e nell'eliminazione degli eventuali doppi conteggi. In determinate circostanze (cfr. par.1) il datore di lavoro è tenuto alla compilazione di più quadri analitici (SA) relativi allo stesso percipiente. Così ad esempio, se nel corso dell'anno si verifica una variazione nella qualifica professionale di un lavoratore, il datore di lavoro è tenuto a rilasciare due distinte dichiarazioni, ciascuna relativa al periodo in cui il lavoratore occupa una determinata qualifica. In questi casi, il sostituto d'imposta può alternativamente: 1) registrare informazioni identiche sui due quadri SA (record), con l'esclusione del campo relativo alla qualifica professionale; 2) riempire un quadro con tutti i dati completi (anagrafici ed economici) ed un altro (integrazione) in cui sono valorizzati i soli dati anagrafici e la nuova qualifica professionale; 3) fornire due dichiarazioni separate e complete di tutte le informazioni del quadro SA, attribuendo al soggetto la retribuzione e/o la prestazione di competenza di ciascun periodo. Questi tre casi richiedono l'applicazione di distinte procedure per il calcolo della retribuzione ridotta totale spettante al soggetto. Nel primo caso, la somma delle retribuzioni ridotte dello stesso soggetto, registrate su più quadri (record), genera un errore di doppio conteggio. Per cui è necessario prendere come dato rappresentativo uno dei due (o più) valori (ad es. il valore massimo). Nel secondo caso, una volta azzerati gli importi non dichiarati (*i valori missing sono identificati con il codice "-1"*), non si commette alcun errore nel processo di aggregazione delle somme corrisposte. Infine nel terzo caso, in presenza di importi distinti ma valorizzati, non si commettono errori di duplicazione nel sommare i diversi valori.

Le procedure di aggregazione dei dati sulle retribuzioni ridotte operano su due livelli: *i*) in presenza di doppie dichiarazioni; *ii*) in presenza di dichiarazioni plurime (triple, quadruple, ecc.). La figura 2 illustra il procedimento di correzione e di aggregazione dei "doppi record" sulle variabili retribuzioni ridotte e competenze correnti. Le linee guida per l'identificazione dei doppi record consistono in: *a*) confrontare i valori assunti sui due record dalle retribuzioni ridotte e dalle competenze correnti; *b*) verificare la possibilità di sovrapposizioni sulla variabile MESILISTA contenente informazioni relative alle 12 mensilità (e valorizzate "0" o "1" a seconda che il soggetto abbia percepito o meno una retribuzione mensile). Quando al passo *a*) si presentano valori identici (o al più con uno scarto compreso tra ± 2 mila lire) tanto nelle retribuzioni ridotte quanto nelle competenze correnti, allora i due record sono considerati "sovrapposti", per cui si prende un solo valore (valore massimo) tra quelli dichiarati rispettivamente per le due variabili economiche (*azione 1*). Viceversa se al passo *a*) le dichiarazioni sulle retribuzioni ridotte differiscono allora si procede a sommarle, mentre si rinvia al passo successivo *b*) il calcolo della variabile aggregata relativa alle competenze correnti. In quest'ultima fase, quando, tra i due, record vi sono due o più sovrapposizioni lungo le 12 componenti della variabile MESILISTA si impone come valido il valore massimo delle competenze correnti dichiarate (*azione 3*). Altrimenti, in presenza di valori inferiori o uguali all'unità, cioè di periodi lavorativi diversi tra i due record, si sommano i valori delle due competenze correnti (*azione 2*). Infine, la combinazione data da retribuzioni ridotte uguali e competenze correnti diverse affida al passo successivo *b*) la decisione finale per il calcolo del dato individuale aggregato su ambedue le variabili economiche.

Figura 2 - Procedura di eliminazione dei doppi record nel quadro SA del modd.770



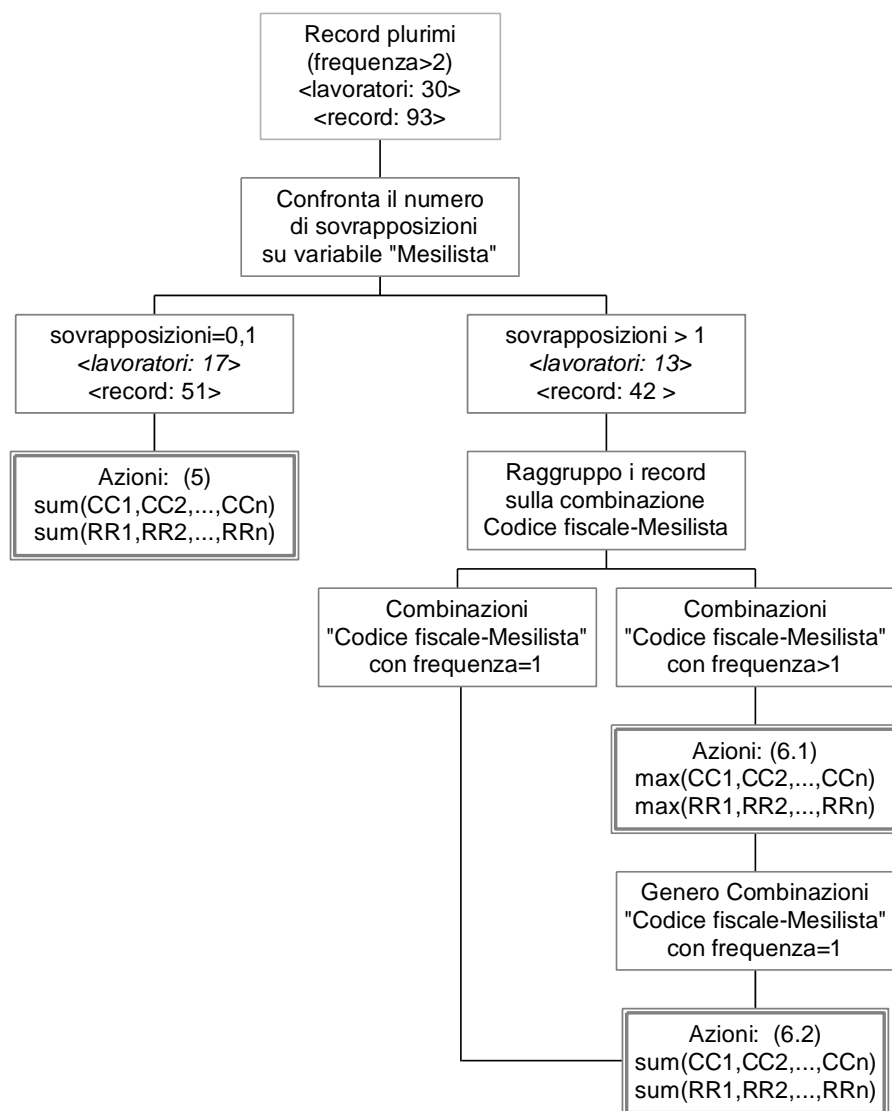
In presenza di dichiarazioni multiple, si procede dapprima alla scomposizione della variabile MESILISTA nelle dodici variabili indicatrici mensili (0=mese retribuito,1=mese non retribuito) e, successivamente, al calcolo della somma dei valori assunti dalle suddette variabili lungo il vettore delle dichiarazioni ripetute (per uno stesso percipiente). Si veda a titolo esemplificativo la seguente figura 3. In questo esempio i casi sovrapposti sono rappresentati dalla contemporanea presenza di valori “0” (mese retribuito) lungo il vettore delle dichiarazioni ripetute. Ai casi sovrapposti corrispondono somme con valori nulli o unitari.

Figura 3 - Procedura di identificazione delle sovrapposizioni sui record multipli

Record	Codice_fiscale	MESILISTA						
		mese_1	mese_2	mese_3	mese_4	mese_5	mese_6	mese...
rec_1	XXX111AAA	0	1	1	0	1	0	
rec_2	XXX111AAA	0	1	1	0	0	1	
rec_3	XXX111AAA	1	0	1	0	1	1	
SOMMA		1	2	3	0	2	2	

In analogia con il criterio adottato al passo *b*) per la correzione dei doppi record, si definiscono come “diverse” le dichiarazioni che contengono al più una sovrapposizione lungo le dodici componenti mensili della variabile MESILISTA. In quest’ultima circostanza, si procede a sommare i valori delle retribuzioni ridotte (competenze correnti). Viceversa se le dichiarazioni plurime presentano due o più sovrapposizioni, allora si procede a raggruppare i record ripetuti (codice fiscale) anche in base alle modalità (combinazioni) della variabile MESILISTA. Due o più record sono considerati identici quando presentano la stessa combinazione di valori nelle 12 componenti di MESILISTA, per cui si prende il valore massimo delle variabili retribuzioni ridotte e delle competenze correnti. Record aventi diverse combinazioni vengono invece definiti come distinte dichiarazioni, per cui si sommano tanto i valori delle retribuzioni ridotte quanto quelli delle competenze correnti.

Figura 4 - Procedura di eliminazione degli errori di duplicazione sui record plurimi nel quadro SA del modd.770



A seguito dell'eliminazione dei doppi conteggi (sui record doppi e plurimi) il valore della spesa erogata per le retribuzioni ridotte decresce di un ammontare pari a 139 milioni di lire (cfr. tabella 4.2.1).

Tabella 4.2.1 - Impatto delle procedure di eliminazione degli errori di duplicazione sui record doppi e plurimi

<i>Eliminazione duplicati su Retribuzioni ridotte</i>	<i>Nr. duplicati</i>	<i>Nr. casi</i>	<i>% casi</i>	<i>RR_corr_ con duplicati</i>	<i>RR_corr_ senza duplicati</i>	<i>Differenze RR_con-dup RR_no_dup</i>
dichiarazioni singole	0	90.404		80.770.990	80.770.990	0
dichiarazioni doppie	120	1.728		2.186.065	2.056.736	129.329
dichiarazioni plurime (3+)	12	30		73.888	63.765	10.123
Totale	132	92.162		83.030.943	82.891.491	139.452

Ai fini dell'indagine sui trattamenti monetari non pensionistici si è tenuto conto solo di quella parte delle retribuzioni ridotte (circa 64 miliardi di lire) che forniscono copertura assicurativa da parte del datore di lavoro per gli eventi malattia e maternità²¹.

Tabella 4.2.2 - Distribuzione delle retribuzioni ridotte per evento coperto

Eventi	Retribuzioni ridotte importi in migliaia di lire	Composizione %
<i>Malattia</i>	60.158.363	93,74
<i>Maternità</i>	4.015.417	6,26
Totale	64.173.780	100

5. Le procedure di controllo e correzione dei dati sul quadro SB

Le procedure di controllo e correzione dei dati sul quadro SB hanno riguardato i seguenti aspetti:

- 1) verifica della presenza di valori significativi (diversi da zero) nelle 7 variabili economiche riferite alle liquidazioni di fine rapporto²² e selezione dei record contenenti almeno un'informazione di interesse;
- 2) verifica della presenza di record duplicati (sovrapposizioni totali) e relativa eliminazione;
- 3) controllo di coerenza sui valori assunti dalle variabili economiche di interesse, attraverso l'implementazione di regole di compatibilità, e controlli di validità sugli importi totali delle liquidazioni (o di range) attraverso la fissazione di valori soglia e di percentuale di errore ammissibili;
- 4) verifica della presenza di mancate risposte su alcune variabili di interesse e implementazione di procedure di imputazione del dato;
- 5) verifica della presenza di parziali sovrapposizioni del dato sulle dichiarazioni plurime per lo stesso percipiente e calcolo dei valori individuali per le variabili economiche;

Il quadro SB è compilato dai sostituti d'imposta che nel corso dell'anno hanno corrisposto trattamenti di fine rapporto di lavoro dipendente (TFR), nonché altre indennità equipollenti di cui all'art.16, comma 1, lett.a, del D.P.R. 917/86. Nella fattispecie le indennità equipollenti riguardano: l'indennità di mancato preavviso, gli incentivi all'esodo, le liquidazioni in capitale di forme pensionistiche complementare. Per definire il complesso di queste prestazioni useremo d'ora in poi il termine "liquidazioni di fine rapporto (LFR)".

La prima fase di selezione (filtro) dei record con valori significativi sulle 7 variabili relative alle liquidazioni di fine rapporto ha comportato l'eliminazione di appena 6 imprese con relative 85 dichiarazioni (cfr. tabella 5.1). Le restanti 11.185 imprese accludevano 76.605 dichiarazioni (record).

La fase di controllo sui record duplicati ha fatto emergere 1.174 dichiarazioni riportanti valori esattamente identici sugli stessi soggetti (sovrapposizione totale), con una percentuale pari all'1,5% (cfr. tabella 5.1). Di conseguenza si è proceduto all'eliminazione dei record contenenti le medesime informazioni (ridondanze).

²¹ Le retribuzioni ridotte connesse all'applicazione della cassa integrazione guadagni (18,7 miliardi di lire) hanno natura di retribuzioni e quindi sono escluse dal campo di osservazione.

²² Trattasi rispettivamente dell'importo totale della liquidazione di fine rapporto erogato al lavoratore, dell'importo riferito alle cinque componenti di quest'ultima e dell'importo delle liquidazioni erogato ai superstiti del lavoratore.

Tabella 5.1 - Le fasi preliminari di trattamento dei dati del quadro SB: liquidazioni di fine rapporto

CARATTERISTICHE DELL'ARCHIVIO: UNICO-SEPARATO 770	Quadro SB			
	Numero di record	Numero di imprese	Numero di beneficiari con almeno 1 TMNP	Note
Data-set in ingresso	76.690	11.191		66.579 record si riferiscono all'unico; 10.111 record si riferiscono alla dichiaraz. separata
<i>Dati che non contengono valori in nessuna delle 7 variabili LFR (missing)</i>	85	6	0	
Data-set in ingresso al netto dei missing sulle 7 variabili LFR	76.605	11.185	68.459	
<i>Numero di record duplicati sul data-set senza missing</i>	1.174			
Data-set senza duplicati e senza missing sui TMNP: "Data-set finale"	75.431	11.185	68.459	La differenza tra il nr. di record e il nr. di beneficiari (6.792) fornisce il nr. di dichiarazioni in aggiunta alla prima
- di cui con 1 dichiaraz. per lo stesso beneficiario (1 record)	62.823		62.823	
- di cui con 2 dichiaraz. per lo stesso beneficiario (1 record)	9.168		4.584	
- di cui con 3 dichiaraz. per lo stesso beneficiario (1 record)	2.478		826	
- di cui con 4 dichiaraz. per lo stesso beneficiario (1 record)	716		179	
- di cui con 5 dichiaraz. per lo stesso beneficiario (1 record)	185		37	
- di cui con 6 dichiaraz. per lo stesso beneficiario (1 record)	54		9	
- di cui con 7 dichiaraz. per lo stesso beneficiario (1 record)	7		1	

5.1 I controlli di coerenza sugli importi delle liquidazioni di fine rapporto

I controlli di coerenza sull'importo totale delle liquidazioni di fine rapporto ha richiesto l'individuazione di alcune regole di compatibilità, sotto-forma di relazioni matematiche, che devono essere soddisfatte da ogni unità statistica. Ad ogni situazione di incompatibilità (violazione del vincolo di compatibilità) segue una corrispondente azione di correzione che assegna un valore

corretto sulla base di procedure di imputazione di tipo deterministico . Utilizzando le informazioni associate ad ogni singola dichiarazione (record), si è proceduto a confrontare il valore assunto dalla variabile “liquidazioni totali” (TOLFR) con quello relativo alla somma delle sue componenti (SCOMP). In tal modo sono state individuate 7 casistiche di cui si fornisce una presentazione in tabella 5.2.

La prima condizione (di compatibilità) si riferisce alla circostanza in cui i valori delle due variabili TOTLFR e SCOMP si discostano al massimo per 2 mila lire, mostrando cioè differenze nulle oppure scostamenti minimi attribuibili ad arrotondamenti. In tale circostanza si ritiene corretto il valore assunto dal saldo delle liquidazioni TOLFR . La relazione è rispettata nel 99,5% dei casi, il che dimostra un buon grado di regolarità dei valori assunti dalle liquidazioni di fine rapporto presso l’archivio modd.770.

La seconda condizione (di incompatibilità) indica la compresenza di valori nulli o mancanti sulla variabile TOLFR e di valori significativi (superiori alle 2 mila lire) nella somma delle componenti. In questo caso si procede ad imputare il valore della somma delle componenti delle liquidazioni al saldo totale. Nel complesso si verificano appena 144 casi di errori, pari all’0,2% delle osservazioni, che tuttavia hanno un impatto considerevole in termini di livello degli importi erogati: +61,7 miliardi circa (+8,2% rispetto al dato grezzo sulla variabile TOLFR originaria).

La terza condizione (di incompatibilità) segnala la presenza di differenze positive, anche se contenute (inferiori al milione di lire), fra la somma delle componenti e il totale delle liquidazioni (avente valore positivo). Si ipotizza pertanto che il saldo totale sottostimi il dato vero rappresentato dalla somma delle componenti. Questa tipologia di errore si presenta in appena 44 osservazioni (0,1%) ed ha un impatto trascurabile sull’importo complessivo delle liquidazioni: +13 milioni di lire (0,002%).

La quarta condizione (di incompatibilità) è diametralmente opposta alla precedente, cioè indica differenze negative, anche se contenute (in valore assoluto non superiori al milione di lire), fra la somma delle componenti ed il saldo totale. In quest’ultima circostanza si ipotizza che qualche componente delle liquidazioni sia nulla o sottostimata e che pertanto il saldo totale è considerato valido.

La quinta condizione (di incompatibilità) indica la compresenza di valori significativi nel saldo totale delle liquidazioni (TOLFR) e di valori nulli nella somma delle componenti (SCOMP). Anche qui si ipotizza che il saldo totale (positivo) è considerato valido mentre le sue componenti sono tutte nulle o mancanti.

La sesta condizione è rappresentata da una combinazione di due relazioni; l’una definisce l’esistenza di differenze, in valore assoluto, significative (superiori al milione di lire) fra la somma delle componenti e il saldo totale, l’altra stabilisce che la variabile “somma delle componenti” fornisce una migliore approssimazione del valore teorico delle liquidazione di fine rapporto rispetto alla variabile “saldo totale”. Il valore teorico delle liquidazioni TOLFR* o TOLFR** viene determinato distintamente per TOLFR (saldo totale) e SCOMP (somma delle componenti), attraverso due modelli di regressione che legano rispettivamente ciascuno di questi due aggregati ad un regressore definito dall’imposta pagata²³. In pratica il criterio adottato stabilisce che se, in riferimento ad una unità statistica, il residuo del primo modello di regressione (variabile dipendente SCOMP) è inferiore al quello calcolato sul secondo modello (variabile dipendente TOLFR) allora la variabile somma delle componenti approssima meglio il valore teorico del complesso delle liquidazioni. L’azione conseguente consiste nell’imputare il valore della somma delle componenti al saldo totale TOLFR. La condizione di incompatibilità descritta si verifica in appena 38 casi (0,05%) e ha un impatto piuttosto modesto sul livello delle liquidazioni: +1,1 miliardi di lire (+0,15%).

La settima condizione denota invece una migliore approssimazione della variabile saldo totale (TOLFR) al valore teorico delle liquidazione, di cui in precedenza, sempre in presenza di differenze

²³ L’imposta pagata rappresenta una *proxy* del totale delle liquidazioni di fine rapporto, in quanto risulta fortemente correlata con quest’ultima.

significative tra i due aggregati. In tale circostanza il valore del saldo non viene corretto in quanto ritenuto valido.

Tabella 5.2 - Controllo e correzione degli errori di incoerenza

Relazioni di (in)compatibilità su: Saldo LFR (TOLFR) e relative Componenti (SCOMP) (AZIONI: correzioni/imputazioni):	Codice	Nr. casi	% casi	TOLFR_orig. (migliaia lire)	TOLFR_cor1 (migliaia lire)	Differenze: TOLFR_orig - TOLFR_cor1
-2≤(SCOMP-TOLFR)≤2 (azione nessuna il saldo LFR è valido)	1	75.086	99,54	753.188.031	753.188.031	0
(TOLFR=0 & SCOMP>2) (azione: TOLFR=Somma(COMPONENTI_LFR))	2	142	0,19	0	61.688.918	-61.688.918
2<SCOMP-TOLFR(+)&≤1000 (azione: TOLFR=Somma(COMPONENTI_LFR))	3	43	0,06	332.045	345.007	-12.962
-1000< SCOMP-TOLFR ≤-2 (azione: nessuna il saldo LFR è valido mentre qualche componente è sottostimata)	4	57	0,08	1.461.239	1.461.239	0
(TOLFR>0 & SCOMP=0) (azione: nessuna il saldo LFR è valido mentre le componenti sono tutte nulle)	5	30	0,04	5.154.773	5.154.773	0
(TOLFR(+)-SCOMP >1000) & (TOLFR-TOLFR* > SCOMP-TOLFR**) (azione: TOLFR=Somma(COMPONENTI_LFR))	6	38	0,05	4.161.674	5.296.857	-1.135.183
(TOLFR(+)-SCOMP >1000) & (TOLFR-TOLFR* < SCOMP-TOLFR**) (azione: nessuna il saldo LFR è valido)	7	35	0,05	4.916.136	4.916.136	0
TOTALE		75.431	100	769.213.898	832.050.961	-62.837.063

5.2 I controlli di validità sugli importi delle liquidazioni di fine rapporto

I controlli di validità hanno riguardato l'implementazione di un complesso di regole per identificare la presenza di valori anomali (che cadono al di fuori di un range prestabilito) sui saldi delle liquidazioni di fine rapporto dichiarati dalle imprese. Una volta fissati i criteri per definire un valore come anomalo, si è proceduto a verificare le percentuali di errori di validità commesse dalle stesse imprese sul complesso delle loro dichiarazioni. Per convenzione si è stabilito di procedere alla correzione dei valori anomali sulla coda superiore della distribuzione solo nel caso in cui la percentuale di casi superava il 30% del complesso delle dichiarazioni della stessa impresa.

Gli errori appartenenti alla coda superiore della distribuzione dei valori delle liquidazioni sono stati classificati in base alle seguenti tipologie: 1) importi con tre zeri finali e valori inferiori a 300 milioni; 2) valori superiori a 1,5 miliardi; 3) importi con tre zeri finali e valori compresi fra 300 milioni e 1,5 miliardi; 4) importi con le ultime tre cifre diverse da "000" e valori compresi fra 300 milioni e 1,5 miliardi.

In particolare, le condizioni 2) e 3) rappresentano situazioni fortemente indiziate di errore, mentre le rimanenti condizioni palesano solo casi sospetti. Le percentuali di casi anomali presenti nelle dichiarazioni delle imprese sono calcolate solo in base al numero di errori del tipo 2) e 3). Nella tabella 5.3 è riportata una sintesi dei risultati conseguiti con l'applicazione del criterio di correzione sopra menzionato. Il numero totale di dichiarazioni con importi diagnosticati come anomali è pari a 43 (0,6% dei casi osservati). Ad essi è associato un incremento di spesa per le liquidazioni di circa 375 miliardi, un importo corrispondente al 82% della spesa totale depurata

dagli errori. In conclusione, si può affermare che una percentuale modesta di casi anomali (appena lo 0,6%) determina variazioni significative sull'ammontare di spesa per le liquidazioni di fine rapporto.

Tabella 5.3 - Controlli di validità sugli importi delle liquidazioni di fine rapporto e relativa correzione

Tipologia di outlier su saldo LFR (AZIONI: correzioni/imputazioni):	Codice errore	Nr. casi	% casi	LFR_cor1	LFR_cor2	Differenze (LFR_cor1 - LFR_cor2)
nessun outlier	0	75.371	99,92	457.164.058	457.164.058	0
outlier coda superiore: >30% dichiarazioni con importi LFR: (>1,5 mld lire) & ("000" finali"> 300 milioni lire) (azione: LFR corretto = LFR originario/1000)	1	43	0,06	374.886.821	374.887	374.511.934
outlier coda inferiore: >80% dichiarazioni con importi LFR ≤ 20 mila lire & età titolare > 40 anni (azione: LFR corretto = LFR originario*1000)	2	17	0,02	82	82.000	-81.918
Totale		75.431	100	832.050.961	457.620.945	374.430.016

Il problema delle dichiarazioni affette da errori del tipo "outlier sulla coda superiore" riguarda un totale di 15 imprese (cfr. tabella 5.4). Di esse nove presentano errori del tipo 2) e 3) su tutte le dichiarazioni (casistica E), mentre altre quattro (casistica A, B, C, D) riportano dichiarazioni tutte contenenti errori del tipo 2), 3) e 4). Un'ultima impresa riporta una dichiarazione esente da errori ed un'altra con un valore anomalo (tipo 2-3-4). In realtà, ad una successiva analisi si è potuto osservare che le due dichiarazioni si riferiscono allo stesso beneficiario: una presenta valori corretti e l'altra registra valori moltiplicati per mille (valori corretti + tre zeri finali). In quest'ultima circostanza, la procedura impone un valore corretto all'importo con tre zeri finali in eccesso e un valore ridotto di 1.000 volte all'altra osservazione corretta. Il valore finale, ottenuto come somma dei due importi, sovrastima il dato reale con una percentuale dello 0,1%. Pertanto si può affermare che la procedura applicata consegue risultati molto soddisfacenti.

Tabella 5.4 - Distribuzione dei casi outlier della spesa delle liquidazioni di fine rapporto per singola impresa (imprese con percentuale di casi 2) e 3 superiori al 30% delle dichiarazioni).

CASISTICA	IMPRESE	BENEFICIARI	FREQ. OUTLIER (flagr=2,3)	NR. OUTLIER (flagr=2,3)	FREQ. OUTLIER tipo 2,3,4 (flagr=2,3,4)	NR. OUTLIER tipo 2,3,4 (flagr=2,3,4)
A	2	4	0,500	2	0,500 1,000	1 2
B	2	6	0,667	4	1,000	4
C	1	6	0,833	5	1,000	5
D	1	11	0,818	9	1,000	10
E	9	16	1,000	16	1,000	16
Totale	15	43	-	36	-	42

Le correzioni sono state applicate alle seguenti tipologie di errore ed hanno comportato una riduzione della spesa di 375 miliardi di lire.

Tabella 5.5 - Distribuzione per tipologie di errore dei casi anomali corretti e dei rispettivi valori prima e dopo la correzione

TIPOLOGIE DI ERRORE SU LFR	Frequenze Assolute	LFR_cor1	LFR_cor2	Differenze (LFR_cor1 – LFR_cor2)
0: senza_errori	1	11.730	12	11.718
2: LFR > 1,5 mld	30	362.567.269	362.567	362.204.702
3: 0,3mld<LFR_000<1,5 mld	6	5.680.000	5.680	5.674.320
4: 0,3mld<LFR_<>000	6	6.627.822	6.628	6.621.194
Totale	43	374.886.821	374.887	374.511.934

Per altro verso, l'applicazione delle procedure di identificazione e correzione dei valori anomali sulla componente delle liquidazioni spettanti ai superstiti non ha segnalato alcun tipo di errore (cfr. tab. 5.6).

Tabella 5.6 - Controlli di validità sugli importi totali delle liquidazioni fine rapporto spettanti ai superstiti

Tipologia di outlier su saldo LFRS (Superstiti) (AZIONI: correzioni/imputazioni):	Codice errore	Nr. casi	% casi	LFRS_cor1	LFRS_cor2	Differenze (LFRS_cor1-LFRS_cor2.)
nessun outlier	0	209	100	2.825.994	2.825.994	0
outlier coda superiore: >30% dichiarazioni con importi LFRS: (>1,5 mld lire) & (“000” finali”>300 milioni lire) (azione: LFRS corretto = LFRS originario/1000)	1	0		0	0	0
outlier coda inferiore: >80% dichiarazioni con importi LFR≤ 20 mila lire & età titolare> 40 anni (azione: LFRS corretto =LFRS originario*1000)	2	0		0	0	0
Totale		209	100	2.825.994	2.825.994	0

5.3 Le procedure di correzione degli errori di duplicazione parziale sulle liquidazioni di fine rapporto

Una volta eseguiti i controlli di coerenza e di validità sugli importi delle liquidazioni, il passo successivo consiste nell'identificare le dichiarazioni doppie o plurime sulle liquidazioni riferite allo stesso beneficiario e nell'eliminare gli eventuali doppi conteggi.

La procedura di identificazione ed eliminazione dei “doppi” record presenti nelle dichiarazioni sulle liquidazioni di fine rapporto utilizza un indice di variabilità assoluta (varianza) per discriminare il caso di sovrapposizione (doppio conteggio) da quello di non sovrapposizione. L'indice viene calcolato sulla coppia di valori osservati (dichiarazioni) per lo stesso beneficiario. Per convenzione si è stabilito di considerare come sovrapposte le dichiarazioni che presentano varianza inferiore a 2 sui valori delle liquidazioni, cioè in altri termini mostrano una differenza in valore assoluto fra i due importi non superiore alle 2 mila lire. In presenza di doppio conteggio si procede a considerare come valido il valore massimo dei due importi dichiarati e ad eliminare la rimanente osservazione. Nel caso di non sovrapposizione si procede a sommare gli importi riferiti allo stesso beneficiario.

La procedura di identificazione ed eliminazione dei record plurimi si articola in due fasi. Nella prima si genera un ordinamento dei record (dichiarazioni) riferiti allo stesso beneficiario in funzione del mese di erogazione delle liquidazioni. Nella seconda si calcolano le differenze prime

lungo la successione dei valori (importi delle liquidazioni) così ordinati. Il criterio applicato definisce un'osservazione come sovrapposta quando, purché riferita allo stesso mese di erogazione, assume un valore non significativamente diverso dalla precedente (in valore assoluto inferiore alle 2 mila lire).

In definitiva l'applicazione delle due procedure di correzione sui record doppi o plurimi produce una riduzione di 222 milioni sulla spesa per le liquidazioni di fine rapporto erogate ai lavoratori, con un'incidenza dello 0,05% sulla spesa non corretta da errori di duplicazione.

Tabella 5.7 - Impatto delle procedure di eliminazione degli errori di duplicazione su record doppi e plurimi sugli importi delle liquidazioni di fine rapporto erogate ai lavoratori.

<i>Eliminazione duplicati su LFR_lavoratori</i>	<i>Nr. duplicati</i>	<i>Nr. casi</i>	<i>% casi</i>	<i>LFR_cor2_ con duplicati</i>	<i>LFR_cor2_ senza duplicati</i>	<i>Differenze (LFR_con-d LFR_no_d)</i>
dichiarazioni singole	0	62.823	83,29	416.450.164	416.450.164	0
dichiarazioni doppie	94	9168	12,15	39.763.846	39.619.124	144.722
dichiarazioni plurime (3+)	79	3040	4,03	1.406.935	1.329.039	77.896
Totale	173	75.431	100	457.620.945	457.398.327	222.618

Infine, non si riscontra alcun caso di sovrapposizione tra le dichiarazioni relative agli importi liquidati ai superstiti (cfr. tab. 5.8).

Tabella 5.8 - Impatto delle procedure di eliminazione degli errori di duplicazione su record doppi e plurimi sugli importi delle liquidazioni di fine rapporto erogate ai superstiti

<i>Eliminazione duplicati su LFR_superstiti</i>	<i>Nr. duplicati</i>	<i>Nr. casi</i>	<i>% casi</i>	<i>LFRS_cor2_ con duplicati</i>	<i>LFRS_cor2_ senza duplicati</i>	<i>Differenze LFRS_con-d LFRS_no_d</i>
dichiarazioni singole	0	207	99,04	2.755.557	2.755.557	0
dichiarazioni doppie	0	2	0,96	70.437	70.437	0
dichiarazioni plurime (3+)	0	0	0,00	0	0	0
Totale	0	209	100	2.825.994	2.825.994	0

La spesa complessiva di tutte le liquidazioni di fine rapporto, erogate sia a lavoratori che ai superstiti, ammonta a circa 460 miliardi di lire. I beneficiari delle liquidazioni raggiungono le 68.459 unità.

6. Le stime campionarie sulle liquidazioni di fine rapporto e retribuzioni ridotte

6.1 Il piano di campionamento

Per l'estrazione del campione di imprese (identificate da codice fiscale/partita IVA) presenti nella lista fornita da ASIA è stato utilizzato un piano di campionamento stratificato ad uno stadio. Allo scopo sono state prese come variabili di stratificazione: la divisione ateco a due cifre (23 raggruppamenti) e la classe dimensionale espressa in termini di addetti totali (3 classi). In particolare, per la prima variabile sono state selezionate le divisioni più rilevanti in termini di quota di dipendenti totali presenti (superiore al 2%), mentre le rimanenti divisioni sono state accorpate all'interno della sottosezione o sezione di appartenenza (cfr. tavola 6.1).

Tavola 6.1 - Formazione degli strati relativi alla variabile ateco a 2 cifre.

Ateco_2	Sezioni/Strati
10-14	C
15	DA_15
17	DB_17
18	DB_18
24	DG
26	DI
28	DJ_28
29	DK
36	DN_36
16-19-20-21-22-23-25-27-30-31-32-33-34-35-37	Altro_D
40-41	E
45	F
52	G_52
51	G_51
50	G_50
55	H
60-61-62-63-64	I
65-66-67	J
74	K_74
70-71-72-73	K_70-73
80	M
85	N
90-92-93	O

In relazione alla classe dimensionale dell'impresa sono stati utilizzati tre raggruppamenti: *i*) 1-9 addetti (piccole imprese); *ii*) 10-99 addetti (medie imprese); *iii*) 100 addetti e oltre (grandi imprese). Pertanto il numero complessivo di strati è risultato pari a 69.

Data la finalità esplorativa dell'indagine sui modelli 770, si è ritenuto opportuno lavorare su una numerosità totale di imprese pari al 2% del totale imprese (con almeno un addetto) contenute in ASIA, ripartite negli strati definiti in modo proporzionale (utilizzando cioè la stessa frazione di campionamento) e comunque garantendo la presenza di almeno 2 imprese nel singolo strato.

La procedura descritta ha portato alla selezione di 22.446 imprese che, sulla base delle informazioni contenute nel registro di imprese ASIA, includono circa 172 mila addetti.

L'obiettivo principale dell'indagine è la stima del numero totale di beneficiari e dell'importo complessivamente erogato per le due tipologie di prestazioni. Le espressioni riferite al punto 1) descrivono rispettivamente: la stima del parametro della popolazione "totale per strato", le varianze e il coefficiente di variazione della stima dei totali per ogni singolo strato. In particolare, N_h rappresenta il numero totale di osservazioni che ricadono nello strato h -esimo, n_h riproduce la numerosità campionaria nel medesimo strato, y_{hi} indica la i -esima osservazione che ricade nello strato h -esimo, mentre infine \bar{y}_h denota la media aritmetica delle osservazioni campionarie che ricadono nello strato h -esimo.

$$\hat{Y}_h = \sum_{i=1}^{n_h} \frac{N_h}{n_h} y_{hi}$$

$$V(\hat{Y}_h) = \frac{N_h(N_h - n_h)}{n_h} s_h^2 = \frac{N_h(N_h - n_h)}{n_h} \left[\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 / (n_h - 1) \right] \quad (1)$$

$$CV(\hat{Y}_h) = \frac{\sqrt{V(\hat{Y}_h)}}{\hat{Y}_h}$$

Per altro verso le espressioni al punto 2) riproducono rispettivamente: la stima del parametro della popolazione "totale generale", nonché la varianza e il coefficiente di variazione di tale stima (cfr. Cecchitelli et al. 1992).

$$\hat{Y}_{ST} = \sum_{h=1}^L \hat{Y}_h$$

$$V(\hat{Y}_{ST}) = V\left(\sum_{h=1}^L \hat{Y}_h\right) = \sum_{h=1}^L V(\hat{Y}_h) \quad (2)$$

$$CV(\hat{Y}_{ST}) = \frac{\sqrt{V(\hat{Y}_{ST})}}{\hat{Y}_{ST}}$$

Tavola 6.2 - Percentuale di campionamento, importo Totale stimato delle liquidazioni di fine rapporto (LFR) e coefficienti di variazione (CV) per strato e complessivo (dati assoluti in lire)

ATECO_2	DIMENSIONE									Totale		
	piccola			media			grande			% (nH/NH)	LFR ^Yh	LFR C.V. %
	% (nh/Nh)	LFR ^Yh	LFR C.V. %	% (nh/Nh)	LFR ^Yh	LFR C.V. %	% (nh/Nh)	LFR ^Yh	LFR C.V. %			
C	2,3%	6.954.967	30,1%	1,4%	27.269.378	30,5%	14,3%	37.339.750	14,5%	2,0%	71.564.095	14,2%
DA_15	2,0%	212.407.863	37,5%	1,5%	277.665.577	12,2%	1,9%	502.451.297	26,4%	2,0%	992.524.738	16,0%
DB_17	2,1%	55.172.807	15,7%	1,6%	344.291.185	13,9%	2,0%	150.629.367	24,5%	1,9%	550.093.358	11,1%
DB_18	2,0%	98.993.203	14,5%	1,6%	305.956.878	23,3%	2,0%	181.382.163	39,2%	1,9%	586.332.244	17,3%
DG	2,2%	10.342.490	26,3%	1,7%	108.820.636	18,0%	2,0%	1.652.064.748	82,7%	2,0%	1.771.227.874	77,1%
DI	2,1%	51.703.879	13,4%	1,6%	293.087.223	14,5%	1,8%	82.357.171	33,4%	2,0%	427.148.273	12,0%
DJ_28	2,2%	205.395.412	8,2%	1,5%	614.202.990	7,1%	1,9%	292.992.092	38,5%	2,0%	1.112.590.493	11,0%
DK	2,2%	103.687.971	15,3%	1,6%	561.664.827	10,3%	2,0%	609.390.832	32,7%	2,0%	1.274.743.629	16,3%
DN_36	2,1%	75.437.787	12,6%	1,5%	253.199.623	13,8%	1,9%	154.184.241	60,7%	2,0%	482.821.651	20,8%
Altro_D	2,1%	349.686.413	6,8%	1,6%	1.607.067.711	11,2%	2,0%	2.661.372.960	40,8%	2,0%	4.618.127.084	23,8%
E	2,0%	5.513.224	61,9%	1,9%	70.580.311	75,0%	2,0%	38.765.713	80,7%	2,0%	114.859.248	53,6%
F	2,0%	929.925.127	27,3%	1,4%	1.794.437.028	55,1%	2,1%	234.680.275	47,4%	1,9%	2.959.042.430	34,7%
G_50	2,0%	177.661.022	11,0%	1,5%	186.581.657	10,7%	3,3%	19.574.190	42,1%	2,0%	383.816.870	7,6%
G_51	2,1%	429.540.467	7,4%	1,5%	717.024.981	9,1%	1,7%	129.546.753	20,9%	2,0%	1.276.112.200	6,1%
G_52	2,0%	481.492.325	18,5%	1,4%	348.649.342	9,8%	2,1%	263.420.833	68,5%	1,9%	1.093.562.500	18,7%
H	2,0%	300.638.213	21,3%	1,3%	308.789.398	12,5%	1,8%	69.630.338	28,1%	1,9%	679.057.949	11,4%
I	2,0%	180.319.729	9,6%	1,6%	667.550.599	39,2%	2,0%	1.498.033.514	78,2%	2,0%	2.345.903.842	51,2%
J	2,0%	142.214.116	50,1%	1,5%	155.769.418	27,0%	2,0%	1.482.101.361	45,7%	1,9%	1.780.084.895	38,3%
K	2,0%	173.162.440	23,7%	1,5%	254.169.935	25,0%	2,0%	231.516.195	36,5%	2,0%	658.848.570	17,2%
K_74	2,0%	336.379.578	4,7%	1,5%	395.671.524	8,2%	1,9%	891.835.028	23,0%	2,0%	1.623.886.130	12,8%
M	1,9%	14.182.190	30,1%	2,0%	13.067.666	23,0%	25,0%	440.540	63,4%	2,0%	27.690.396	18,9%
N	2,0%	66.270.386	16,9%	1,7%	105.146.986	29,3%	1,8%	105.697.191	23,4%	2,0%	277.114.562	14,8%
O	2,0%	125.144.619	10,5%	1,5%	263.980.906	33,6%	1,5%	49.387.308	79,3%	1,9%	438.512.833	22,3%
Totale	2,0%	4.532.226.230	6,7%	1,5%	9.674.645.775	11,0%	2,0%	11.338.793.859	19,8%	1,9%	25.545.665.864	9,8%

Tavola 6.3 - Percentuale di campionamento, importo Totale stimato del numero di beneficiari LFR e coefficienti di variazione (CV) per strato e complessivo

ATECO_2	DIMENSIONE									Totale		
	piccola			media			grande			%	Beneficiari	Beneficiari
	%(nh/Nh)	Beneficiari ^Yh	Beneficiari C.V. %	%(nh/Nh)	Beneficiari ^Yh	Beneficiari C.V. %	%(nh/Nh)	Beneficiari ^Yh	Beneficiari C.V. %			
C	2,3%	2.082	26,6%	1,4%	4.898	38,9%	14,3%	1.512	2,4%	2,0%	8.492	23,3%
DA_15	2,0%	51.038	18,9%	1,5%	57.140	17,7%	1,9%	15.583	24,3%	2,0%	123.761	11,7%
DB_17	2,1%	10.796	13,1%	1,6%	40.284	12,4%	2,0%	13.668	25,1%	1,9%	64.747	9,6%
DB_18	2,0%	22.324	11,7%	1,6%	43.183	9,5%	2,0%	10.201	38,2%	1,9%	75.709	8,3%
DG	2,2%	2.386	13,2%	1,7%	9.812	16,3%	2,0%	26.803	61,8%	2,0%	39.001	42,7%
DI	2,1%	11.801	11,9%	1,6%	38.328	14,9%	1,8%	9.810	29,4%	2,0%	59.939	10,9%
DJ_28	2,2%	43.624	5,9%	1,5%	93.160	7,9%	1,9%	24.963	50,5%	2,0%	161.747	9,2%
DK	2,2%	18.393	9,0%	1,6%	60.109	7,4%	2,0%	45.208	53,5%	2,0%	123.711	19,9%
DN_36	2,1%	15.414	11,2%	1,5%	43.717	20,7%	1,9%	27.114	89,4%	2,0%	86.244	30,1%
Altro_D	2,1%	65.163	5,2%	1,6%	186.894	6,3%	2,0%	79.554	18,6%	2,0%	331.611	5,8%
E	2,0%	345	46,9%	1,9%	2.190	43,7%	2,0%	2.107	58,6%	2,0%	4.642	33,8%
F	2,0%	225.578	3,7%	1,4%	209.897	12,2%	2,1%	20.598	12,7%	1,9%	456.074	6,0%
G_50	2,0%	33.916	4,9%	1,5%	23.330	11,1%	3,3%	1.470	31,2%	2,0%	58.717	5,3%
G_51	2,1%	56.089	6,8%	1,5%	72.240	8,0%	1,7%	6.728	43,6%	2,0%	135.056	5,5%
G_52	2,0%	81.795	3,2%	1,4%	68.573	8,7%	2,1%	23.327	23,7%	1,9%	173.694	4,9%
H	2,0%	169.599	4,9%	1,3%	137.775	9,9%	1,8%	25.842	47,9%	1,9%	333.216	6,1%
I	2,0%	43.327	6,1%	1,6%	69.996	16,9%	2,0%	63.342	36,2%	2,0%	176.665	14,7%
J	2,0%	7.533	12,8%	1,5%	6.854	20,2%	2,0%	21.471	31,3%	1,9%	35.858	19,3%
K	2,0%	23.496	9,2%	1,5%	24.812	16,6%	2,0%	7.902	31,5%	2,0%	56.210	9,4%
K_74	2,0%	68.926	5,7%	1,5%	98.407	7,1%	1,9%	1.010.922	48,1%	2,0%	1.178.255	41,3%
M	1,9%	3.685	26,5%	2,0%	12.215	47,1%	25,0%	36	27,2%	2,0%	15.936	36,6%
N	2,0%	10.617	12,0%	1,7%	22.252	22,2%	1,8%	23.396	25,5%	2,0%	56.265	14,0%
O	2,0%	48.993	6,8%	1,5%	47.680	19,6%	1,5%	8.576	47,0%	1,9%	105.249	10,2%
Totale	2,0%	1.016.921	1,8%	1,5%	1.373.746	3,0%	2,0%	1.470.133	33,3%	1,9%	3.860.801	12,7%

Tavola 6.4 - Percentuale di campionamento, importo Totale stimato delle retribuzioni ridotte e coefficienti di variazione (CV) per strato e complessivo (dati assoluti in lire)

ATECO_2	DIMENSIONE									Totale		
	piccola			media			grande			% (nH/NH)	Retr. Rid. ^Yh	Retr. Rid. C.V. %
	% (nh/Nh)	Retr. Rid. ^Yh	Retr. Rid. C.V. %	% (nh/Nh)	Retr. Rid. ^Yh	Retr. Rid. C.V. %	% (nh/Nh)	Retr. Rid. ^Yh	Retr. Rid. C.V. %			
C	2,3%	1.615.444	23,6%	1,4%	7.105.102	23,1%	14,3%	1.499.337	29,2%	2,0%	10.219.883	17,0%
DA_15	2,0%	13.001.012	11,7%	1,5%	42.282.804	13,5%	1,9%	39.889.531	36,7%	2,0%	95.173.346	16,6%
DB_17	2,1%	7.523.606	10,5%	1,6%	68.846.780	12,3%	2,0%	59.472.120	29,8%	1,9%	135.842.506	14,5%
DB_18	2,0%	11.979.937	11,6%	1,6%	61.189.796	10,4%	2,0%	47.043.629	61,4%	1,9%	120.213.361	24,6%
DG	2,2%	1.666.145	18,4%	1,7%	25.174.072	18,5%	2,0%	56.133.679	47,6%	2,0%	82.973.896	32,7%
DI	2,1%	12.124.052	11,0%	1,6%	54.726.467	15,2%	1,8%	30.384.791	37,6%	2,0%	97.235.310	14,6%
DJ_28	2,2%	46.034.359	7,5%	1,5%	180.209.526	6,6%	1,9%	61.713.560	29,7%	2,0%	287.957.445	7,7%
DK	2,2%	15.203.537	8,6%	1,6%	94.805.311	7,4%	2,0%	137.253.768	52,5%	2,0%	247.262.615	29,3%
DN_36	2,1%	13.289.276	9,1%	1,5%	57.551.693	12,7%	1,9%	19.839.170	60,3%	2,0%	90.680.139	15,5%
Altro_D	2,1%	49.059.620	5,3%	1,6%	274.657.278	5,5%	2,0%	242.270.666	27,4%	2,0%	565.987.564	12,0%
E	2,0%	61.443	52,9%	1,9%	685.626	54,2%	2,0%	-	-	2,0%	747.070	49,9%
F	2,0%	123.658.151	3,9%	1,4%	229.023.197	8,8%	2,1%	44.919.274	43,3%	1,9%	397.600.622	7,2%
G_50	2,0%	25.233.611	7,4%	1,5%	35.459.054	8,6%	3,3%	2.904.750	54,6%	2,0%	63.597.415	6,1%
G_51	2,1%	38.743.329	5,5%	1,5%	90.027.542	9,9%	1,7%	21.818.893	23,7%	2,0%	150.589.763	7,0%
G_52	2,0%	48.819.642	4,8%	1,4%	58.070.748	9,4%	2,1%	30.980.446	33,9%	1,9%	137.870.837	8,8%
H	2,0%	23.138.501	6,2%	1,3%	36.176.258	13,0%	1,8%	14.759.341	12,4%	1,9%	74.074.100	7,1%
I	2,0%	25.396.738	13,4%	1,6%	61.583.673	15,7%	2,0%	493.759.915	87,5%	2,0%	580.740.327	74,4%
J	2,0%	3.949.994	26,5%	1,5%	8.834.128	30,4%	2,0%	31.512.186	47,9%	1,9%	44.296.308	34,7%
K	2,0%	13.579.581	18,1%	1,5%	26.109.540	27,4%	2,0%	10.684.150	65,9%	2,0%	50.373.270	20,5%
K_74	2,0%	39.296.112	4,4%	1,5%	63.621.038	8,1%	1,9%	154.810.297	22,2%	2,0%	257.727.447	13,5%
M	1,9%	1.760.825	23,2%	2,0%	2.248.179	25,9%	25,0%	176.172	40,5%	2,0%	4.185.176	17,1%
N	2,0%	6.622.101	11,7%	1,7%	31.211.993	18,1%	1,8%	38.038.865	44,2%	2,0%	75.872.959	23,4%
O	2,0%	12.452.539	8,1%	1,5%	32.804.790	28,3%	1,5%	12.579.786	61,6%	1,9%	57.837.115	21,0%
Totale	2,0%	534.209.554	1,8%	1,5%	1.542.404.595	2,6%	2,0%	1.552.444.324	28,9%	1,9%	3.629.058.473	12,4%

Tavola 6.5 - Percentuale di campionamento, importo Totale stimato del numero di beneficiari di retribuzioni ridotte e coefficienti di variazione (CV) per strato e complessivo

ATECO_2	DIMENSIONE									Totale		
	piccola			media			grande			%	Beneficiari	Beneficiari
	%(nh/Nh)	Beneficiari ^Yh	Beneficiari C.V. %	%(nh/Nh)	Beneficiari ^Yh	Beneficiari C.V. %	%(nh/Nh)	Beneficiari ^Yh	Beneficiari C.V. %			
C	2,3%	1.861	14,9%	1,4%	7.584	21,5%	14,3%	1.834	51,5%	2,0%	11.279	16,9%
DA_15	2,0%	17.698	6,8%	1,5%	47.404	10,7%	1,9%	32.102	39,2%	2,0%	97.204	14,0%
DB_17	2,1%	13.302	10,7%	1,6%	75.265	9,9%	2,0%	45.900	23,6%	1,9%	134.467	9,9%
DB_18	2,0%	18.158	9,6%	1,6%	79.457	7,8%	2,0%	31.159	52,3%	1,9%	128.774	13,6%
DG	2,2%	2.476	12,7%	1,7%	21.239	16,5%	2,0%	41.115	53,2%	2,0%	64.829	34,2%
DI	2,1%	13.864	10,1%	1,6%	57.121	11,8%	1,8%	28.889	37,0%	2,0%	99.874	12,7%
DJ_28	2,2%	54.800	4,7%	1,5%	178.854	5,3%	1,9%	54.270	33,5%	2,0%	287.923	7,2%
DK	2,2%	19.249	8,6%	1,6%	99.676	6,7%	2,0%	133.591	60,3%	2,0%	252.515	32,0%
DN_36	2,1%	20.358	8,4%	1,5%	69.499	12,1%	1,9%	20.889	60,5%	2,0%	110.746	13,8%
Altro_D	2,1%	66.244	5,1%	1,6%	298.943	4,4%	2,0%	182.207	17,9%	2,0%	547.394	6,4%
E	2,0%	148	0,0%	1,9%	678	60,5%	2,0%	-	-	2,0%	826	49,6%
F	2,0%	130.538	2,1%	1,4%	179.180	6,8%	2,1%	26.688	24,7%	1,9%	336.406	4,2%
G_50	2,0%	35.526	4,7%	1,5%	44.061	7,4%	3,3%	3.180	44,6%	2,0%	82.767	4,7%
G_51	2,1%	56.499	3,4%	1,5%	114.889	7,5%	1,7%	16.388	27,3%	2,0%	187.775	5,3%
G_52	2,0%	76.799	3,4%	1,4%	91.654	8,3%	2,1%	34.260	28,9%	1,9%	202.712	6,3%
H	2,0%	45.653	4,5%	1,3%	56.072	10,0%	1,8%	21.024	10,8%	1,9%	122.749	5,2%
I	2,0%	27.637	4,8%	1,6%	61.648	12,0%	2,0%	153.766	72,0%	2,0%	243.051	45,6%
J	2,0%	5.388	17,9%	1,5%	6.991	24,8%	2,0%	17.799	44,9%	1,9%	30.178	27,3%
K	2,0%	18.395	11,7%	1,5%	28.835	25,8%	2,0%	6.569	66,3%	2,0%	53.799	16,5%
K_74	2,0%	58.686	2,2%	1,5%	89.131	7,1%	1,9%	251.697	30,9%	2,0%	399.514	19,5%
M	1,9%	3.025	16,1%	2,0%	3.821	16,7%	25,0%	116	4,2%	2,0%	6.962	11,6%
N	2,0%	13.039	9,8%	1,7%	40.876	15,1%	1,8%	44.235	39,2%	2,0%	98.151	18,8%
O	2,0%	25.628	6,5%	1,5%	39.410	19,5%	1,5%	11.993	49,0%	1,9%	77.031	12,7%
Totale	2,0%	724.971	1,2%	1,5%	1.692.288	2,0%	2,0%	1.159.667	14,4%	1,9%	3.576.925	4,8%

7. Conclusioni

La finalità della presente ricerca è verificare la possibilità d'impiego della fonte amministrativa "dichiarazioni fiscali modd.770" ai fini della produzione di statistiche sui TMNP. Pur limitando l'analisi alle liquidazioni di fine rapporto e alle integrazioni economiche nel periodo malattia e maternità, erogate ai lavoratori dipendenti di un campione di imprese tratte dall'archivio Archivio Statistico delle Imprese Attive (ASIA), si può ritenere soddisfacente la qualità dei dati di questa fonte fiscale e tale da essere impiegata correntemente nell'implementazione dell'indagine sui TMNP. Ad esempio si è potuto dimostrare l'esistenza di una buona concordanza tra l'archivio fiscale modd.770 e la fonte ASIA (fonte ufficiale ISTAT) rispetto al numero medio annuo di lavoratori per impresa. Tuttavia, a causa dell'indisponibilità di altre fonti esterne amministrative o statistiche (*benchmark*) che rilevano il dato economico sui trattamenti esaminati al livello di singola impresa dei dati, non siamo in grado di effettuare un controllo esterno sui valori rilevati sulle unità campionate. La fonte fiscale delle dichiarazioni modd. 770 costituisce al riguardo l'unica base dati per la rilevazione delle prestazioni sopra citate. Per quanto riguarda l'aspetto della qualità del dato riferita agli importi dei trattamenti, si è potuto ricostruire il profilo degli errori e valutare il peso di ciascuna componente rispetto al totale della spesa. Nel caso delle liquidazioni di fine rapporto, ad esempio, sono state impiegate procedure di controllo e correzione deterministiche che hanno permesso di individuare un numero ridotto di casi anomali (0,6%) sugli importi (*outlier*) che, tuttavia, determinano distorsioni significative sulla spesa complessivamente erogata (+82%). Ciò suggerisce l'impiego di procedure molto accurate e selettive nella fase di individuazione degli *outlier*. La presenza di duplicati sulle dichiarazioni fiscali trasmesse dai datori di lavoro ad integrazione/sostituzione delle dichiarazioni precedenti impone un'analisi rigorosa di tutte le variabili presenti nella struttura del record ai fini della giusta classificazione delle informazioni ripetute. Al riguardo l'utilizzo combinato della variabile "*mesilista*" (formata da 12 componenti indicanti i mesi lavorati nell'anno) e delle variabili riferite ai valori dei trattamenti individuali erogati (analisi delle differenze tra i valori ripetuti) forniscono una ottima base per guidare il ricercatore al trattamento delle informazioni ripetute nella fase di controllo e correzione dei dati. L'analisi sperimentale presentata in questa sede costituisce il primo passo da parte dell'Istat verso il consolidamento di informazioni di fonte amministrativa (fiscale) sulle prestazioni TMNP (importi e numero di beneficiari). Le prossime fasi da implementare riguarderanno il consolidamento delle informazioni attinenti alle caratteristiche dei percettori (età, sesso, luogo di residenza, qualifica professionale) di interesse ai fini dell'indagine TMNP e l'analisi dei dati economici sui restanti trattamenti individuali presenti in archivio (assegni al nucleo familiare, indennità di malattia e maternità, integrazioni salariali, per la quota parte anticipata dai datori di lavoro per conto dell'Inps). In merito al primo punto sarà necessario valutare la coerenza delle informazioni contenute nel codice fiscale, dove è possibile ricavare l'età anagrafica e il genere del percettore, e verificare la robustezza di altre variabili identificative del luogo di residenza e della qualifica professionale. Per quanto concerne il successivo punto sarà indispensabile provvedere alla realizzazione di un "*linkage*", a livello di singola unità di impresa (codice fiscale/partita IVA, matricola Inps), tra la fonte amministrativa modd. 770 e l'archivio DM10 dell'Inps, al fine di valutare la coerenza del dato fiscale sul valore dei trattamenti esaminati rispetto al dato vero dell'Inps (*benchmark*).

Riferimenti bibliografici

Barcaroli G., D'Aurizio L., Luzi O., Mannari A. Pallara A., 1999, Metodi e software per il controllo e la correzione dei dati, Documenti Istat, Roma, n. 1/1999

Cicchitelli G., Herzel A. e Montanari G. E., 1992, Il campionamento statistico, Il Mulino, Bologna

Consolini P., 2000, Le prestazioni sociali monetarie non pensionistiche: aspetti istituzionali e classificazioni statistiche.

Consolini P. e R. De Carli R., 2002, Le prestazioni sociali monetarie non pensionistiche: unità di analisi, fonti e rappresentazione statistica dei dati, Documenti Istat, Roma, n. 1/2002

EUROSTAT, 1996, ESSPROS Manual 1996, Luxemburg.

Falorsi P.D., Pallara A., Russo A., 2003, Temi di ricerca ed esperienze sull'utilizzo a fini statistici di dati di fonte amministrativa, Franco Angeli, Milano.