

LA MOBILITÀ DEI LAVORATORI DA FONTI AMMINISTRATIVE E DA SURVEYS SULLE FAMIGLIE: UN'ANALISI COMPARATA

P. Baretta, U. Trivellato

1. INTRODUZIONE

Negli ultimi due decenni gli studi incentrati sulla misura e l'analisi dei flussi di lavoratori e di posti di lavoro hanno conosciuto un imponente sviluppo (per recenti rassegne, vedi, tra gli altri, Davis *et al.*, 1996, Hamermesh *et al.*, 1996, Abowd e Kramarz, 1999, Davis e Haltiwanger, 1999). La rilevanza di questa letteratura difficilmente può essere sopravvalutata. Sul piano descrittivo, essa ha documentato livello e *pattern* della mobilità nel mercato del lavoro, mettendo in evidenza importanti fatti stilizzati – innanzitutto, gli elevati tassi di *worker and job flows* che si riscontrano nelle moderne economie di mercato – e numerose, significative regolarità empiriche. Sul terreno analitico-interpretativo, le possibilità offerte – e gli stimoli venuti – dalle statistiche sui flussi nel mercato del lavoro hanno avuto rilievo in molteplici direzioni (per una succinta ricognizione, vedi Davis e Haltiwanger, 1998, pp. 79-81).

Gli studi sui flussi di lavoratori e di posti di lavoro utilizzano una varietà di basi di microdati longitudinali. In sintesi, esse possono essere ricondotte a tre tipologie: (i) *employee data*, alimentati da *surveys* – tipicamente indagini campionarie sulle famiglie – oppure da fonti amministrative – tipicamente archivi previdenziali di lavoratori –; (ii) *employer data*, riferiti a stabilimenti e/o a imprese, alimentati da indagini – censuarie e campionarie – e/o da registri amministrativi; (iii) *employer-employee linked data*, che per l'appunto si caratterizzano per l'integrazione di dati riferiti ad imprese/stabilimenti e ai lavoratori e consentono di misurare contemporaneamente flussi di lavoratori e di posti di lavoro.

Lavoro svolto nell'ambito del progetto di ricerca “Dinamiche e persistenze nel mercato del lavoro italiano ed effetti di politiche (basi di dati, misura, analisi)”, cofinanziato dal MIUR. Il lavoro si è giovato della collaborazione dell'Istat – nel quadro di una convenzione di ricerca tra l'Istituto e il Dipartimento di Scienze Statistiche dell'Università di Padova – e del LABORATORIO R. Revelli, che hanno reso disponibili i *files* di microdati di cui si dà conto nel testo principale. I precedenti sono nella tesi di dottorato di Paolo Baretta (Baretta, 2002), alla quale rimandiamo per dettagli in merito alle caratteristiche delle fonti e alle modalità di loro utilizzo. Una versione preliminare è stata presentata al XVIII Convegno di Economia del lavoro dell'AIEL (Messina, 25-26 settembre 2003). Siamo grati ai partecipanti al convegno e soprattutto a Bruno Contini, Anna de Angelini, Antonio Discenza e Roberto Leombruni per utili consigli e commenti.

L'uso, in un specifico studio, di un certo insieme (o di alcuni insiemi) di dati è dettato innanzitutto dagli obiettivi conoscitivi dello studio in questione. Esso, tuttavia, è anche condizionato dallo spettro delle fonti disponibili: spettro che varia da paese a paese (in particolare per le basi di dati di origine amministrativa, le quali risentono in maniera decisiva di normative e prassi nazionali) e che può essere parecchio limitato per vari motivi. Si tratta, infatti, di insiemi di microdati longitudinali la cui creazione e manutenzione è parecchio costosa. Quand'anche siano prodotti, poi, la possibilità di accedervi per i ricercatori può essere ristretta per motivi di *privacy*.

In Italia le analisi sulla mobilità dei lavoratori e dei posti di lavoro poggiano quasi esclusivamente su dati amministrativi: soprattutto sugli archivi dell'Inps, segnatamente sul "Panel di lavoratori e imprese Inps/LABORatorio R. Revelli" (nel seguito INPS), progressivamente messo a punto da un gruppo di ricercatori coordinato da Bruno Contini¹. I primi studi sono di Contini e Revelli (1987) e (1992) e di Contini *et al.* (1992). Il più recente contributo – un volume a più mani sulla mobilità del lavoro e delle retribuzioni nel nostro paese dal 1985 al 1996 – è Contini (2002).

Le potenzialità ed i limiti della fonte INPS sono noti. Il vantaggio è nella documentazione dettagliata delle storie lavorative, che sono descritte avendo come unità di tempo il mese. Tra le limitazioni, due meritano di essere segnalate. Innanzitutto, la copertura è parziale: circoscritta, in prima approssimazione, al lavoro dipendente nel settore privato extra-agricolo. In secondo luogo, la fonte soffre di scarsa tempestività: in Contini (2002), appena citato, le analisi sulla mobilità dei lavoratori si arrestano al 1996.

La questione che qui affrontiamo è la seguente: è ragionevole proporsi di utilizzare la Rilevazione Trimestrale delle Forze di Lavoro (nel seguito RTFL) come ulteriore fonte corrente per analisi della mobilità dei lavoratori?² Qualora la risposta fosse affermativa, i vantaggi sarebbero palesi. Le analisi potrebbero essere riferite a pressoché tutta l'occupazione³. Inoltre, sarebbero tempestive: prodotte con un *lag* dell'ordine di 4-5 mesi e aggiornate trimestralmente. A ciò si oppone, peraltro, un serio ostacolo: la RTFL non è disegnata per cogliere la mobilità dei lavoratori. Il piano di osservazione risultante dal questionario, infatti, documenta in modo assai parziale i segmenti delle storie lavorative.

I nostri obiettivi sono pertanto due:

¹ Dei non molti studi condotti ancora su basi di dati Inps, ma diverse dal panel appena menzionato, ricordiamo Tattara e Valentini (2002) e (2003). Una diversa fonte è rappresentata dagli archivi amministrativi – su lavoratori e imprese – prodotti nel quadro delle attività di collocamento, tradizionalmente di competenza del Ministero del Lavoro e dal 1997 trasferite a Regioni ed Enti Locali (e operativamente svolte a livello dei Centri per l'impiego). In quest'ambito, spicca l'esperienza che si è sviluppata nel Veneto: vedi Accornero (2000) e Trivellato (2001).

² Va da sé che la RTFL, per la sua natura di indagine campionaria sulle famiglie, non documenta alcunché in materia di flussi dei posti di lavoro.

³ Rispetto all'aggregato di riferimento rappresentato dall'occupazione totale interna mancherebbero gli occupati non residenti e i residenti in convivenze; all'opposto, sarebbero inclusi i residenti occupati temporaneamente all'estero.

- (a) identificare e, per quanto possibile, controllare le ragioni di divergenza fra le misure di mobilità fornite dalle due fonti;
- (b) vagliare se si possa elaborare (almeno) un metodo in grado di riconciliare le evidenze fornite dalle due fonti, e che si presti quindi per essere impiegato per ragionevoli stime della mobilità dei lavoratori dalla RTFL.

L'analisi empirica è condotta intorno all'anno 1995, per il quale disponiamo delle necessarie basi di microdati da entrambe le fonti⁴. Inoltre il 1995 è, opportunamente, un anno neutrale dal punto di vista ciclico – gli effetti della recessione del 1992 cominciano a essere riassorbiti anche sul fronte dell'occupazione – (Altissimo *et al.*, 2000).

L'articolo procede come segue. Nella sez. 2 esaminiamo le caratteristiche delle due fonti e le similarità e le differenze tra le stesse, rilevanti per gli obiettivi dello studio. Nella sez. 3 richiamiamo concisamente concetti e definizioni in tema di misure di mobilità dei lavoratori e presentiamo alcune precisazioni metodologiche al riguardo.

Con la sez. 4 affrontiamo gli aspetti cruciali dello studio comparato. Dapprima illustriamo quattro piani di osservazione delle storie lavorative (nel seguito detti anche 'metodi') che possono essere definiti per la RTFL, utilizzando in vario modo le informazioni – correnti e retrospettive – fornite dalla sequenza delle rilevazioni *cross-section* e il disegno parzialmente longitudinale dell'indagine, e ne esaminiamo la portata informativa rispetto alla misura della mobilità (sez. 4.1). Ci soffermiamo quindi sulla strategia per analisi comparate di mobilità dei lavoratori sulle fonti INPS e RTFL. Essa poggia su due operazioni, in parte interconnesse: (i) la standardizzazione della popolazione, cioè a dire l'individuazione di un comune aggregato di riferimento – costituito grosso modo dai lavoratori dipendenti dei settori dell'industria e dei servizi privati –; (ii) la simulazione dei piani di osservazione propri della RTFL sui *data set* INPS, al fine di condurre analisi di mobilità sulle due fonti comparabili anche sotto il profilo della base informativa sulle storie lavorative (sez. 4.2). Il vaglio della capacità dei diversi metodi nel misurare la mobilità dei lavoratori avviene comparandone le *performance* con quella del piano di osservazione originario della fonte INPS, dato dalla storia lavorativa descritta con unità di tempo il mese (sez. 4.3).

Nella sez. 5 concentriamo l'attenzione sul metodo che si è mostrato soddisfacente, detto 'dell'abbinamento a tre occasioni', il quale utilizza appunto dati abbinati di tre rilevazioni della RTFL nell'arco dell'anno, e ne presentiamo e discutiamo i risultati salienti⁵. L'evidenza empirica segnala la capacità del metodo, e quindi dei corrispondenti *data set* longitudinali della RTFL, di cogliere larga parte della mobilità dei lavoratori documentata dall'INPS: pressoché completamente per quanto attiene a misure di riallocazione – cioè a dire, della frazione di lavoratori interessati

⁴ Ad oggi, è anche l'anno più prossimo per uno studio comparato della mobilità dalle due fonti. Come abbiamo appena segnalato, infatti, l'attuale aggiornamento del panel INPS si ferma al 1996. Ciò preclude di – o comunque rende problematico – condurre l'analisi su tale anno, perché per la fonte INPS non sarebbe possibile distinguere gli episodi lavorativi che si concludono a dicembre 1996 da quelli censurati alla stessa data.

⁵ Sintetici ragguagli sui risultati degli altri tre metodi sono nell'Appendice 3.

da transizioni –; con una sottostima contenuta, dell’ordine di 3 punti percentuali per quanto riguarda misure di *gross worker turnover* – cioè a dire, del numero medio di transizioni operate dai lavoratori –.

Ciò suggerisce che il metodo dell’abbinamento a tre occasioni potrebbe essere impiegato per la stima di soddisfacenti misure di mobilità dei lavoratori basate sull’intero campione della RTFL. Tali misure avrebbero due non trascurabili pregi: essere riferite all’insieme degli occupati delle famiglie residenti – dunque, a un aggregato prossimo all’occupazione totale interna –; essere tempestive, con un *lag* ridotto e aggiornabili trimestralmente. L’opportunità di una utilizzazione corrente della RTFL per analisi della mobilità dei lavoratori è, peraltro, condizionata al fatto che rispetto al 1995, anno del nostro esercizio, non si siano manifestati vistosi incrementi della mobilità, segnatamente delle storie lavorative costellate di molteplici, brevi episodi. Il che è, palesemente, da verificare.

2. LE DUE FONTI DI DATI

Il panel INPS e la RTFL sono largamente noti. Ci limitiamo quindi a richiamarne i tratti essenziali, con riguardo soprattutto agli aspetti attinenti alla loro utilizzazione per analisi comparate della mobilità dei lavoratori, e a dare conto delle operazioni preliminari condotte a tale scopo sulle due basi di dati (per dettagli vedi, tra i molti, Filippi, Pacelli e Villosio, 2002, per l’ INPS e Trivellato, 1997, per la RTFL).

2.1. Il “Panel di lavoratori e imprese Inps/LABORatorio R. Revelli”

Il “Panel di lavoratori e imprese Inps/LABORatorio R. Revelli” risulta dall’integrazione di due archivi amministrativi dell’Inps:

- (a) l’archivio dei lavoratori, alimentato dai moduli 01/M annuali, presentati dai datori di lavoro (registrati nell’archivio sub (b)) per ogni lavoratore alle dipendenze nell’arco dell’anno, che riassumono le informazioni rilevanti per il calcolo dei benefici pensionistici;
- (b) l’archivio delle imprese, che include tutte le imprese private dei settori industriali e dei servizi con almeno un dipendente, alimentato dai moduli DM10 mensili.

Il panel INPS è costituito da tutti i lavoratori nati il giorno 10 dei mesi di marzo, giugno settembre e dicembre di qualunque anno. Per questi lavoratori sono stati estratti tutti i record presenti nell’archivio lavoratori dal 1985 al 1996 e generati i corrispondenti dati longitudinali individuali. Ad ogni lavoratore presente nel campione sono state poi associate le informazioni longitudinali sull’impresa, tratte dal secondo archivio.

In definitiva, si dispone di un *employer-employee panel data set* relativo a un campione casuale di lavoratori dipendenti delle imprese private dell’industria e del terziario (con frazione di campionamento pari a circa 1:90), per il periodo 1985-1996, con informazioni disponibili a cadenza mensile.

Per l'utilizzazione che se ne farà in questa sede, è utile precisare che: (i) di massima, è considerato occupato, nel mese, chi abbia prestato lavoro per almeno un giorno nel mese in questione; (ii) per ogni lavoratore è disponibile la sequenza degli episodi lavorativi (anche sovrapposti, in tutto o in parte, perché svolti con imprese diverse), definiti dal mese e anno di inizio/fine del rapporto di lavoro⁶; (iii) i dati sulla storia lavorativa sono depurati della mobilità spuria conseguente a trasformazioni societarie (secondo la procedura descritta in Pacelli e Villosio, 2002); (iv) fra le informazioni sui singoli episodi lavorativi figurano la posizione nella professione e l'attività economica.

Ai fini di analisi della mobilità dei lavoratori centrate sull'anno 1995, abbiamo costruito una base di dati *ad hoc*, tramite una procedura che può essere così riassunta. Innanzitutto, si sono eliminati gli episodi lavorativi con date di inizio/fine del rapporto di lavoro mancanti o incongrue. In secondo luogo, si sono eliminati i molteplici episodi riferiti a uno stesso lavoratore e coincidenti per mese e anno di inizio/fine, mantenendo quello col maggior numero di giornate retribuite (in caso di parità, col maggior numero di settimane retribuite e, eventualmente, con retribuzione più alta). Infine, si è generato il *data set* per l'anno 1995.

La numerosità campionaria del *data set* INPS per il 1995 è di poco inferiore ai 130.000 individui, ai quali corrispondono poco meno di 142.500 episodi lavorativi (vedi oltre la Tavola 2, riga 2).

2.2. I dati di fonte RTFL

L'universo di riferimento della RTFL è la popolazione nelle famiglie residenti, popolazione che ai fini della partecipazione al lavoro è ristretta alle persone di almeno 15 anni. L'indagine è condotta a cadenza trimestrale, su un campione – ad ogni occasione – dell'ordine di 75.000 famiglie. Il disegno campionario è con rotazione, del tipo 2-2-2⁷. È questo un tratto dell'indagine importante anche ai nostri fini, perché consente di costruire (i) panel a due occasioni a distanza di un anno, con tasso di abbinamento lordo dell'ordine del 50%, nonché (ii) panel a tre occasioni, del tipo $t.I-t.II-t+1.I$ oppure $t.I-t.IV-t+1.I$, con tasso di abbinamento lordo dell'ordine del 25%.

La condizione rispetto al lavoro delle persone intervistate è definita con riguardo alla settimana di riferimento della rilevazione – di massima la prima del mese –. È classificato occupato chi si è dichiarato tale o, comunque, ha dichiarato di aver lavorato almeno un'ora nella settimana.

⁶ Nella ricostruzione degli episodi lavorativi, eventi quali l'assenza per malattia, maternità, cassa integrazione guadagni, ecc. non comportano l'interruzione del rapporto di lavoro (cioè a dire una separazione, seguita da un'associazione presso la stessa impresa a conclusione dell'evento): il lavoratore, infatti, mantiene l'*attachment* all'impresa, continua a ricevere una retribuzione e a versare contributi. La condizione di aver prestato lavoro almeno un giorno in ognuno dei mesi dell'episodio va intesa in questo quadro e vale, quindi, 'di massima'.

⁷ Detto altrimenti, una famiglia che entra a far parte del campione è intervistata per due occasioni d'indagine successive, ne esce temporaneamente per altre due, rientra per due ulteriori occasioni successive, dopo di che esce definitivamente dal campione.

Le ulteriori informazioni sulla storia lavorativa che la RTFL acquisisce sono parecchio parziali, e risultano da risposte a quesiti retrospettivi. Specificamente:

- (a) ad ogni occasione d'indagine, vengono rilevati: (a1) per gli occupati, il mese e l'anno di inizio dell'episodio di lavoro in corso (tramite il quesito 24: “*Quando ha iniziato a lavorare con l'attuale datore di lavoro o in proprio?*”); (a2) per i non occupati con precedenti esperienze lavorative, il mese e l'anno della fine dell'ultimo episodio lavorativo (tramite il quesito 38: “*Quando ha concluso l'ultima occupazione?*”);
- (b) nell'indagine di aprile, è rilevata anche la condizione rispetto al lavoro nello stesso mese dell'anno precedente (tramite il quesito 64: “*Qual era la sua condizione [un anno fa]?*”⁸).

Inoltre, per ogni episodio lavorativo riportato sono rilevate la “*posizione nella professione*” e la “*branca di attività economica*”.

Una ricostruzione meno frammentaria della storia lavorativa si ha, poi, a seguito della creazione – tramite *record linkage* – di panel a due o a tre occasioni, i quali, per ogni individuo che sia e/o sia stato occupato, combinano sequenze di informazioni di tipo (a) (vedi oltre la sez. 4.2).

Per lo studio della mobilità dei lavoratori intorno all'anno 1995, abbiamo preso le mosse dalla serie di *files* della RTFL relativi alle occasioni da 1995.I a 1996.II. La generazione dei corrispondenti *data set* sezionali utili ai nostri fini ha comportato l'eliminazione dei *record* relativi alle persone che risultano non avere né avere avuto alcun episodio lavorativo, cioè a dire (i) con età inferiore ai 15 anni o (ii) mai occupate, né nella settimana di riferimento né precedentemente (l'assenza di occupazione valendo, evidentemente, in ciascuna delle tre occasioni di indagine)⁹.

La dimensione campionaria dei risultanti *data set* sezionali RTFL varia dai 119.500 ai poco più di 123.000 individui (vedi oltre la Tavola 2, riga 1).

2.3. Le differenze tra le due fonti a fini di stima della mobilità dei lavoratori

Caratteristiche e differenze salienti delle due fonti, significative rispetto all'obiettivo della misura della mobilità dei lavoratori, sono nel quadro sinottico della Tavola 1. Il prospetto è largamente autoesplicativo e non abbisogna dunque di molti commenti.

Per un verso vi sono differenze pressoché irriducibili, ma verosimilmente di peso contenuto. Così è, ad esempio, per diversità in aspetti della popolazione di riferimento (occupati presenti in INPS *vs.* occupati di famiglie residenti in RTFL), nella capacità delle due fonti di seguire una popolazione dinamica (adeguata per INPS, mentre lo è solo in parte per RTFL, che poggia su un campione estratto dallo *stock* inizia-

⁸ Si noti che la condizione un anno prima risulta semplicemente dalla risposta a tale domanda – quindi da una sorta di ‘autodichiarazione’ –, e non già dall'articolato insieme di quesiti utilizzato correntemente per classificare le persone come occupate o disoccupate (o, per residuo, non appartenenti alle forze di lavoro).

⁹ La procedura per la generazione dei panel RTFL a due e tre occasioni, logicamente analoga, è più laboriosa, perché disegno con rotazione dell'indagine, mobilità dei lavoratori e standardizzazione della popolazione interferiscono l'uno con l'altro. Per la sua presentazione rinviamo alla sez. 4.2.

TAVOLA 1
Quadro sinottico delle differenze tra le due fonti

Fattori	Fonti		Differenze ed effetti attesi ^a
	INPS	RTFL	
<i>Differenze (in parte) eliminabili / controllabili</i>			
<i>Popolazione di riferimento</i>	Occupati presenti dipendenti da imprese private extra-agricole	Occupati di famiglie residenti	Occupati interni <i>vs.</i> in famiglie residenti [<i>p</i> ; >] Copertura dell'occupazione parziale <i>vs.</i> totale [<i>G</i> ; <]
<i>Piano di osservazione</i>	Storia lavorativa con unità di tempo il mese	<i>Singola occasione:</i> stato corrente + scarse informazioni retrospettive <i>Indagine aprile:</i> + stato un anno prima <i>Abbinamento:</i> + combinazione di informazioni correnti	Piani di osservazione della RTFL variamente incompleti rispetto a quello INPS [<i>G</i> / <i>p</i> ; >]
<i>Differenze irriducibili (ma, plausibilmente, di minor peso)</i>			
<i>Definizione di occupato</i>	Almeno una giornata di lavoro nel mese	Dichiaratosi tale o almeno un'ora di lavoro nella settimana di riferimento	1 giorno/mese <i>vs.</i> 1 ora/settimana [<i>p</i> ; >]
<i>Caratteristiche della fonte</i>	Campione casuale da popolazione dinamica Dati amministrativi connessi a obblighi previdenziali e associate fonti di errore: - errori connessi a evasione o sotto-dichiarazione - errori di registrazione	Campione da popolazione al tempo <i>t</i> (1 [^] occasione) + attrito Dati da <i>survey</i> e associate fonti di errore - non risposta totale - effetti <i>proxy respondent</i> - errori di risposta (inclusi quelli dovuti a memoria) - errori di registrazione	In RTFL sottostima del la mobilità (per attrito; in dati abbinati per disegno) [<i>p</i> ; >] Varie fonti di errore, in larga parte diverse [<i>p</i> ; > / <]

^a Entità: *G* ≡ grande, *p* ≡ piccola. Direzione: INPS > o < RTFL.

le di famiglie residenti¹⁰), nella definizione di occupato (chi ha [versamenti previdenziali per] almeno una giornata di lavoro nel mese in INPS *vs.* chi si è dichiarato tale o ha lavorato almeno un'ora nella settimana di riferimento in RTFL).

Per un altro verso, vi sono due fattori di discrepanza di notevole rilievo.

- (a) È diversa la copertura della popolazione degli occupati. La copertura è parziale – grosso modo limitata ai dipendenti da imprese private extra-agricole – in INPS, mentre è totale in RTFL.
- (b) È diverso il piano di osservazione della mobilità. In INPS si hanno storie lavorative documentate con unità di tempo il mese. Per RTFL si possono definire vari piani di osservazione – a seconda delle informazioni che vengano utilizzate da una singola *cross-section* e dell'eventuale sfruttamento della dimensione longitudinale dell'indagine –, che comunque non ricostruiscono in maniera completa i segmenti delle storie lavorative.

¹⁰ La RTFL risente quindi delle mancate risposte, dovute a irreperibilità o rifiuto nelle singole *cross-section* e inoltre, nei panel a due e più occasioni, al disegno stesso dell'indagine, perché le famiglie non sono seguite quando cambiano residenza. E le mancate risposte interessano in misura maggiore le famiglie/individui con più elevata mobilità lavorativa. Per evidenze in questo senso con riguardo al disegno longitudinale della RTFL, vedi Istat (2002), pp. 6-14.

Rispetto a questi due fattori, sono peraltro possibili interventi volti a eliminare, almeno in parte, le differenze fra le fonti in questione. Questi interventi sono appunto il fuoco dell'esercizio svolto nel seguito.

3. LE MISURE DI MOBILITÀ DEI LAVORATORI: ALCUNE PRECISAZIONI

Prima di procedere con le analisi empiriche, tornano peraltro utili alcune puntualizzazioni sulle misure di mobilità dei lavoratori. Gli eventi elementari rilevanti sono, notoriamente, l'inizio di un rapporto di lavoro – *associazione* – e la fine di un rapporto di lavoro – *separazione* – da parte di un individuo. Indici sintetici di mobilità sono ottenuti riassumendo convenientemente i movimenti – le associazioni e le separazioni – di un insieme di lavoratori in un dato intervallo di tempo, tipicamente l'anno (vedi, per tutti, Davis e Haltiwanger, 1999).

Denotiamo con A_i il numero di associazioni e con S_i il numero di separazioni del generico individuo i in un dato intervallo di tempo, nel nostro caso l'anno, e con ${}_w A_i$ e con ${}_w S_i$ le corrispondenti grandezze riferite solo agli episodi lavorativi in corso all'inizio o alla fine dell'intervallo (escludendo, cioè, gli episodi interni). Denotiamo infine con $N_s = (N_{s_i} + N_{s_f})/2$ e N_f due misure del totale degli occupati, che chiamiamo rispettivamente di *stock* e di *flusso*, date:

- la prima dalla media degli *stock* di occupati all'inizio e alla fine dell'intervallo¹¹;
- l'altra dal totale degli individui che hanno avuto un rapporto di lavoro in corso in un qualsiasi momento dell'intervallo. Palesemente, N_f è dato dallo *stock* iniziale di occupati, più coloro che si sono associati nell'intervallo e sono ancora occupati alla fine dell'intervallo, più i transitati nell'intervallo (vale a dire, gli associati nell'intervallo non presenti alla fine dello stesso).

Indici sintetici di mobilità, innanzitutto il tasso di associazione dei lavoratori (a) e il tasso di separazione dei lavoratori (s) e quindi il *gross worker turnover* (t) – dato dalla somma di a e s –, possono essere definiti in vari modi, a seconda delle grandezze che vengono poste al numeratore e al denominatore. Tre plausibili formulazioni alternative, rispettivamente per il tasso di associazione, per quello di separazione e per il *gross worker turnover* (nel seguito anche *gwt*), sono le seguenti:

$$\begin{aligned}
 a_s &= \frac{1}{N_s} \sum_{i=1}^{N_f} A_i, & {}_w a_s &= \frac{1}{N_s} \sum_{i=1}^{N_f} {}_w A_i, & a_f &= \frac{1}{N_f} \sum_{i=1}^{N_f} A_i; \\
 s_s &= \frac{1}{N_s} \sum_{i=1}^{N_f} S_i, & {}_w s_s &= \frac{1}{N_s} \sum_{i=1}^{N_f} {}_w S_i, & s_f &= \frac{1}{N_f} \sum_{i=1}^{N_f} S_i; \\
 t_s &= a_s + s_s, & {}_w t_s &= {}_w a_s + {}_w s_s, & t_f &= a_f + s_f.
 \end{aligned} \tag{1}$$

¹¹ Possibili alternative sono date dallo *stock* di occupati iniziale o da quello finale.

In letteratura è frequente l'impiego di a_s , s_s e t_s , cioè di indici che rapportano il totale dei movimenti pertinenti ad una misura dello *stock* di occupati nell'intervallo (vedi ancora Davis e Haltiwanger, 1999). A ben vedere, tuttavia, la scelta non è convincente, perché la grandezza al numeratore viene rapportata a una grandezza comprensiva soltanto di una parte (sia pure, di solito, la gran parte) degli individui che si sono associati e/o separati nell'intervallo. L'esigenza di coerenza fra le grandezze che compaiono al numeratore – i movimenti – e al denominatore – gli individui esposti al rischio di mobilità nell'intervallo in questione – suggerisce che sono più appropriati gli altri indici (per indicazioni in tal senso, vedi Anastasia *et al.*, 2000, e Baretta, 2002, pp. 17-26). Specificamente:

- se gli obiettivi dell'analisi, o i vincoli dell'informazione disponibile, portano a considerare i soli lavoratori presenti all'inizio o alla fine dell'intervallo, tornano utili gli indici ${}_w a_s$, ${}_w s_s$ e ${}_w t_s$ (per una formulazione dei tassi di associazione e di separazione al netto di episodi completamente interni al periodo di riferimento, vedi anche Abowd *et al.*, 1999);
- in generale sono di interesse soprattutto gli indici a_f , s_f e t_f , i quali rapportano l'insieme dei movimenti nell'intervallo a quella che abbiamo chiamato la 'popolazione flusso', cioè al totale degli esposti al rischio di mobilità;
- una possibile alternativa è data, poi, dalla stima del *gwt* recentemente proposta da Leombruni e Quaranta (2002), pp. 180-181. Essi ponderano ciascuno dei lavoratori presenti nella popolazione flusso per il periodo di permanenza nella popolazione in questione – cioè a dire, per il numero di mesi rapportato a 12 –. La proposta, metodologicamente persuasiva, non è peraltro utile per il vaglio comparato di misure di mobilità da RTFL e INPS, perché può essere utilizzata soltanto per la fonte INPS¹².

Se portiamo l'attenzione sul numero di lavoratori coinvolti nei movimenti (a prescindere dal numero di transizioni effettate), invece che sul numero di transizioni, abbiamo le corrispondenti misure di *gross worker reallocation*. Denotata con R_i la variabile indicatrice della presenza, per l'individuo i , di almeno un'associazione o almeno una separazione nell'intervallo (e con ${}_w R_i$ la variabile indicatrice con riferimento ai soli episodi lavorativi in corso all'inizio o alla fine dell'intervallo), per il tasso di riallocazione dei lavoratori (r) sono possibili le seguenti alternative:

$$r_s = \frac{1}{N_s} \sum_{i=1}^{N_f} R_i, \quad {}_w r_s = \frac{1}{N_s} \sum_{i=1}^{N_f} {}_w R_i, \quad r_f = \frac{1}{N_f} \sum_{i=1}^{N_f} R_i. \quad (2)$$

Per esse valgono considerazioni analoghe a quelle appena fatte per i tassi di associazione e di separazione e per il *gwt*.

¹² Come anticipato, e come vedremo meglio nel seguito, dalla RTFL non è possibile ricostruire la storia lavorativa mese per mese per tutti i lavoratori.

4. I DIVERSI PIANI DI OSSERVAZIONE E LA STRATEGIA PER I CONFRONTI

4.1. I diversi piani di osservazione

Il piano di osservazione della mobilità più completo coincide con quello originario della fonte INPS, che chiamiamo *metodo del panel*. Esso documenta la storia lavorativa di ogni individuo avendo come unità di tempo il mese; nel nostro studio comparato, registra con questo formato gli eventi ed episodi lavorativi avvenuti nell'anno 1995. È appena ovvio notare che può essere utilizzato solo per la fonte INPS. Esso costituisce il naturale *benchmark* rispetto al quale vagliare la *performance* di più sommarî piani di osservazione che possono essere proposti sulla scorta della base informativa della RTFL.

I piani di osservazione che definiamo a partire dalla RTFL, utilizzando in maniera variamente completa le informazioni fornite dalla sequenza delle rilevazioni *cross-section*, sono quattro. Ne richiamiamo brevemente i tratti distintivi soffermandoci sull'ultimo – il metodo dell'abbinamento –, che è risultato decisamente il più soddisfacente (per dettagli, vedi Baretta, 2002, pp. 33-48).

(1) *Metodo della 'tenure' pregressa*

Introdotta da Gennari e Sestito (1996), utilizza la data di inizio del lavoro corrente, disponibile in ogni *cross-section* per gli occupati, per determinarne la *tenure* pregressa.

Sulla base delle distribuzioni degli occupati per classi di *tenure* in due rilevazioni a distanza di un anno – nel nostro caso 1995.I e 1996.I –, si stimano i (soli) tassi di separazione specifici per classe di *tenure* e quindi il (solo) tasso di separazione dei lavoratori nel complesso¹³.

(2) *Metodo della condizione*

Esplorato da Leombruni e Quaranta (2001), utilizza l'informazione rilevata in una singola *cross-section* – nel nostro caso 1996.I –. Essa consta della condizione lavorativa corrente, più la risposta al quesito sulla data d'inizio del lavoro in corso (per gli occupati) o di conclusione dell'ultima occupazione (per i non occupati precedentemente occupati).

Disponendo di informazioni riferite solamente al rapporto di lavoro più recente, è possibile rilevare al più un'associazione o una separazione per ogni individuo¹⁴.

(3) *Metodo dell'anno prima*

Utilizza le informazioni acquisite con la rilevazione di aprile – nel nostro caso 1996.II – e fa quindi riferimento all'anno compreso tra 1995.II e 1996.II. Le

¹³ Operando su distribuzioni marginali degli occupati per classi di *tenure* a distanza di un anno (e non su dati individuali), il tasso di separazione è ovviamente definito in maniera diversa da come detto nella sez. 3. Lo denotiamo con s_0 ed è $s_0 = (O_1 - O_2) / O_1$, dove $O_1 = \sum_{c=1}^C O_{1c}$ e $O_2 = \sum_{c^*=1}^{C^*} O_{2c^*}$, essendo O_{1c} e O_{2c^*} gli occupati rispettivamente al tempo 1 nella classe di *tenure* c e al tempo 2 nella classe di *tenure* c^* , la quale ha estremi traslati di un anno rispetto a c .

¹⁴ Pertanto, con riferimento alle misure di mobilità è $t_s \equiv r_s \equiv n^t_s \equiv n^r_s$.

informazioni attengono sia alla condizione lavorativa corrente che a quella di un anno prima (rilevata con quesiti retrospettivi), e comprendono inoltre le risposte ai quesiti sul tipo di lavoro svolto (posizione nella professione e branca di attività economica), sempre ai due tempi – corrente e nello stesso mese dell'anno precedente –. Del metodo si danno quattro varianti, che si distinguono per il dettaglio della classificazione degli occupati per attività economica utilizzato per identificare le transizioni.

In questo caso, è possibile rilevare al più un'associazione e una separazione per ogni individuo.

(4) *Metodo dell'abbinamento a tre (due) occasioni*

Il metodo dell'abbinamento sfrutta il disegno con rotazione della RTFL: tramite una procedura di *record linkage*, si generano dati longitudinali a tre (due) occasioni. Per identificare le transizioni operate dagli individui intervistati, si utilizzano le informazioni sia sulla condizione lavorativa corrente sia sulla data di inizio del lavoro corrente (per gli occupati) o di fine dell'ultimo lavoro (per i disoccupati precedentemente occupati), rilevate ad ognuna delle tre (due) occasioni di indagine.

Qualora le due tipologie di informazione – quella sulla condizione corrente e quella retrospettiva sulla data di inizio/fine del lavoro – siano incoerenti, si dà la prevalenza a quella sulla condizione corrente, che appare ragionevole considerare più attendibile, e di conseguenza si corregge, o si trascura, l'altra¹⁵.

Le possibilità di rilevazione di movimenti ed episodi lavorativi consentite dal metodo dell'abbinamento sono ben illustrate dalle Figure 1 e 2.

Per la loro lettura, vale quanto segue.

- Ogni rapporto di lavoro è rappresentato da un segmento a tratto continuo orientato orizzontalmente lungo un asse temporale che va da sinistra a destra, il quale viene intersecato dalle distinte occasioni d'indagine, rappresentate da segmenti disposti verticalmente (nel caso di tre occasioni, per semplicità posti equidistanti, mentre gli intervalli fra le occasioni sono differenti: l'uno di 3 e l'altro di 9 mesi¹⁶).

¹⁵ Nell'abbinamento a tre occasioni, nel collocare le transizioni identificate nell'anno – definito dalle due occasioni estreme – si devono risolvere eventuali ulteriori incoerenze, che emergono perché si dispone anche delle informazioni rilevate nell'occasione d'indagine intermedia. Esse possono riguardare (i) le date di inizio/fine del rapporto di lavoro, riportate in due diverse occasioni e riferite allo stesso episodio, nonché (ii) difformità fra la condizione lavorativa rilevata correntemente in un'occasione e quella risultante dalle durate riportate degli episodi in una successiva occasione. Il criterio seguito per risolvere queste incoerenze è analogo a quello appena enunciato: prevalenza all'informazione sulla condizione corrente, e in subordine alla risposta più prossima al tempo (riportato) dell'evento. Sul problema degli errori di risposta in indagini sul lavoro, connessi in particolare agli effetti della memoria, vedi la rassegna di Bound *et al.* (2001), pp. 3743-3748 e 3784-3811.

¹⁶ Più precisamente, gli intervalli fra le occasioni di indagine sono di 3 e 9 mesi per la sequenza di rilevazioni gennaio-aprile-gennaio (nel nostro caso 1995.I-IV-1996.I) e di 9 e 3 mesi per la sequenza di rilevazioni gennaio-ottobre-gennaio (nel nostro caso 1995.I-IV-1996.I).

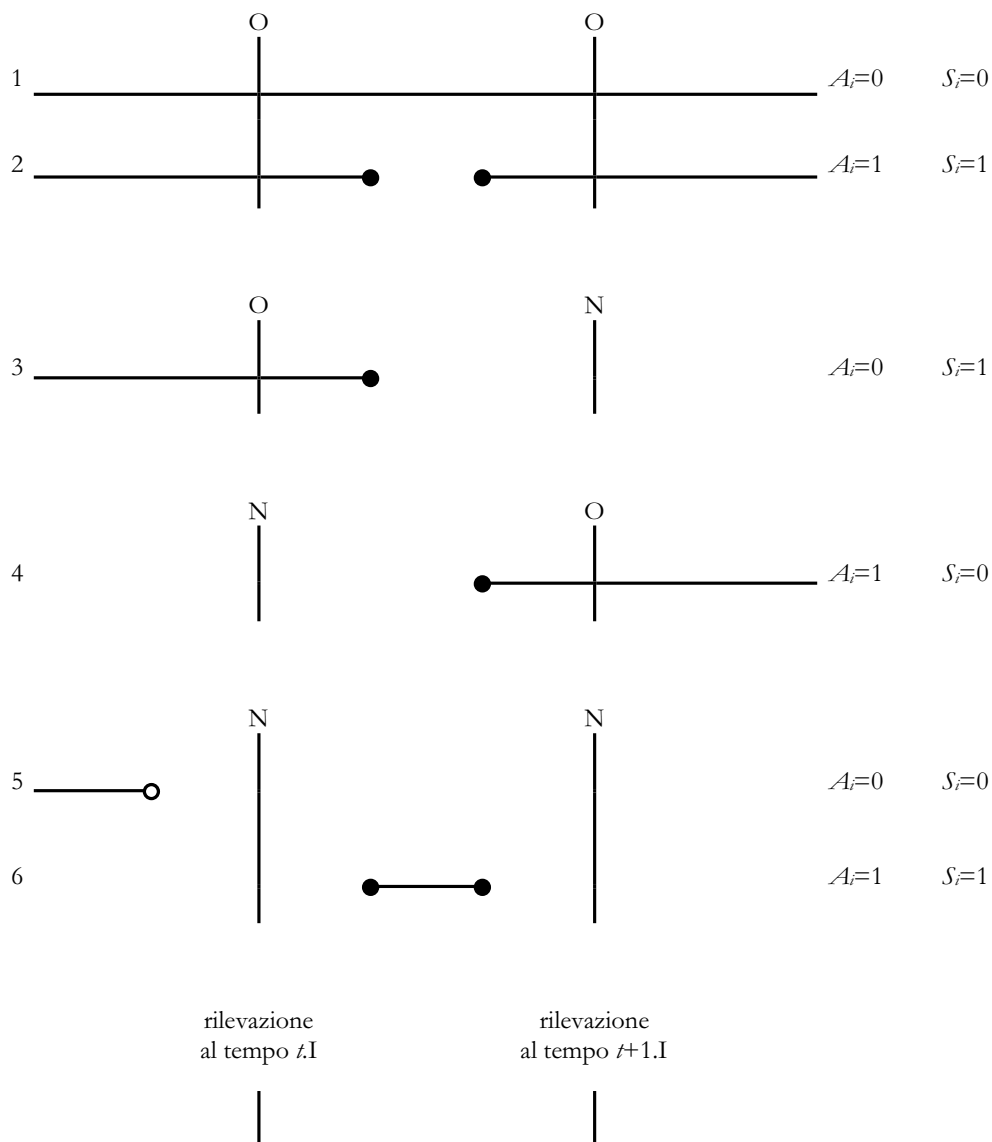


Figura 1 – Schema per l'identificazione degli episodi lavorativi e delle transizioni dall'abbinamento a due occasioni.

- La condizione lavorativa di un individuo rilevata correntemente in ogni occasione d'indagine è contraddistinta dalla lettera 'O' per gli occupati e dalla lettera 'N' per i non occupati, e viene posta in corrispondenza dell'asse relativo all'occasione in questione. Una qualsiasi transizione, sia essa un'associazione o una separazione, effettuata all'interno dell'intervallo di tempo compreso tra le due occasioni estreme – 1995.I e 1996. I – è contraddistinta dal simbolo '•', mentre un'eventuale separazione avvenuta prima dell'intervallo è contraddistinta dal simbolo 'o'.

Già con l'abbinamento a due occasioni si ha un sensibile guadagno nelle capacità di osservazione dei segmenti di storie lavorative (Figura 1).

Ma è soprattutto con l'abbinamento a tre occasioni che diventa possibile una documentazione parecchio più dettagliata dei segmenti di storie lavorative. Com'è evidente dalla Figura 2, la disponibilità di informazioni da una terza

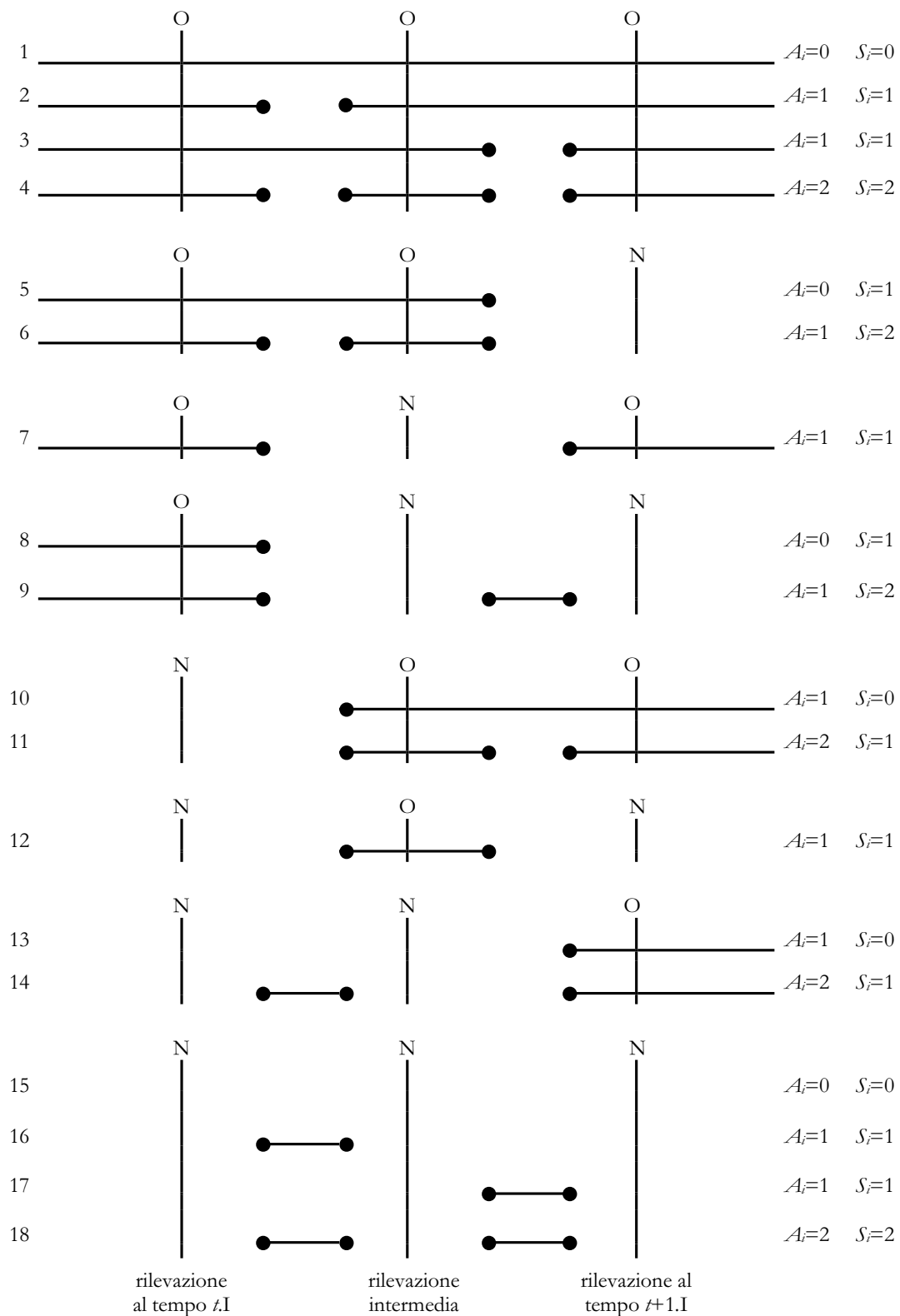


Figura 2 – Schema per l'identificazione degli episodi lavorativi e delle transizioni dall'abbinamento a tre occasioni.

occasione d'indagine – intermedia rispetto alle due a distanza di un anno – aumenta in modo considerevole il numero di movimenti che è potenzialmente accertabile per ogni individuo. Il piano di osservazione implicato dall'abbinamento a tre occasioni consente di identificare sino a due associazioni e due separazioni – e sino a tre episodi di occupazione – per ogni individuo nell'anno.

4.2. *La strategia per analisi comparate di mobilità*

Come abbiamo visto nella sez. 2, le due fonti differiscono per l'aggregato di riferimento: le popolazioni dei lavoratori, di cui i *data set* INPS e RTFL sono campioni rappresentativi, sono molto diverse. Inoltre, abbiamo appena documentato che i piani di osservazione definiti a partire dalla RTFL colgono solo parzialmente la storia lavorativa registrata da INPS.

Per operare confronti informativi sulle misure di mobilità dei lavoratori desumibili dalle due fonti, sono pertanto cruciali due operazioni:

- (a) la standardizzazione della popolazione, cioè a dire l'individuazione di un comune aggregato di riferimento, in modo da eliminare differenze sistematiche nella mobilità dei lavoratori dovute alla diversa copertura della popolazione;
- (b) la simulazione dei metodi di osservazione propri della RTFL sui *data set* INPS, al fine di ottenere stime della mobilità dalle due fonti comparabili anche sotto il profilo della base informativa sulle storie lavorative e poterle quindi comparare col *benchmark* rappresentato dal metodo del panel sulla fonte INPS.

4.2.1. *Standardizzazione della popolazione*

L'obiettivo di standardizzare la popolazione viene raggiunto eliminando gli episodi lavorativi relativi a categorie di attività economica non comparabili, perché nel panel INPS – la fonte con copertura limitata – sono assenti o comunque presenti in misura parziale, selezionata. L'esito è l'individuazione di una comune popolazione di riferimento costituita dai lavoratori dipendenti di gran parte dell'industria e dei servizi privati.

La standardizzazione della popolazione è condotta – giova sottolinearlo – con riferimento agli episodi lavorativi, non agli individui. Infatti, è la natura dell'episodio lavorativo che decide della sua inclusione in INPS (o, all'opposto, della sua esclusione). Sul fronte della RTFL, peraltro, per i primi due piani di osservazione – della *tenure* pregressa e della condizione – l'eliminazione di un episodio coincide con l'eliminazione di un individuo, perché essi permettono di identificare un solo episodio lavorativo per individuo.

Il processo di standardizzazione avviene in più passi.

- (a) Innanzitutto, dalla RTFL sono esclusi gli episodi di lavoro indipendente.
- (b) Sono inoltre esclusi da entrambe le fonti:
 - (b1) gli episodi lavorativi nei rami 'Agricoltura' e 'Pubblica amministrazione', perché quelli presenti in INPS sono una frazione piccola, e non rappresentativa, del corrispondente totale;

(b2) gli episodi lavorativi in una serie di divisioni di attività economica soprattutto (ma non solo) nell'ambito dei rami 'Istruzione e sanità' e 'Altri servizi pubblici e sociali', per lo stesso motivo¹⁷.

Maggiori dettagli sulle modalità di standardizzazione della popolazione sono presentati nell'Appendice 1. I risultati salienti, per la RTFL riferiti ai metodi che si basano su rilevazioni *cross-section*, sono nella Tavola 2.

TAVOLA 2

Numerosità campionarie dei data set INPS e RTFL ai vari passi della procedura di standardizzazione della popolazione (in corsivo numero di episodi lavorativi)

Popolazione di riferimento	INPS 1995	RTFL ^a				
		1995.I	1995.II	1995.III	1995.IV	1996.I
Totale occupati	–	119.545	123.289	123.144	123.209	122.926
– indipendenti	129.710 <i>142.428</i>	94.314	97.030	97.044	97.063	96.951
– ATECO 01,10	115.473 <i>125.850</i>	81.915	84.396	84.810	84.841	84.713
– ATECO 11,12 ^b	108.365 <i>118.082</i>	69.289	71.563	71.900	71.743	71.640
Pop. standardizzata	104.162 <i>113.386</i>	65.934	68.042	68.516	68.225	68.097

^a Per i metodi che si basano su singole rilevazioni *cross-section*.

^b Fatte salve le divisioni 91 e 93 (vedi l'Appendice 1).

¹⁷ La procedura è sensibile a errori nella classificazione per professione e per attività economica, potenzialmente rilevanti soprattutto per la RTFL. Essi possono produrre sia falsi positivi (inclusione di episodi lavorativi che si sarebbero dovuti escludere) sia falsi negativi (esclusione di episodi che, all'opposto, si sarebbero dovuti escludere). Per contenerli, abbiamo scelto di usare la classificazione per professione e per attività economica nella misura più parsimoniosa possibile, in un duplice senso. (i) Innanzitutto, la condizione dichiarata è utilizzata soltanto per la standardizzazione della popolazione – mantenendo o eliminando gli opportuni episodi lavorativi –, e a tale scopo si usa l'informazione tratta da *una singola occasione* (per gli episodi lavorativi in corso è l'occasione che per prima li rileva, per gli episodi lavorativi conclusi quella più prossima al tempo (riportato) della fine dell'episodio). Essa non serve, dunque, per individuare associazioni e separazioni, per le quali si fa riferimento unicamente alle informazioni riguardanti la condizioni lavorativa corrente e la data di inizio/fine del lavoro (l'eccezione è il metodo dell'anno prima). (ii) Inoltre, la condizione dichiarata è utilizzata al livello di disaggregazione minimo richiesto dalle regole di standardizzazione (vedi l'Appendice 1).

In assenza di *pattern* sistematici negli errori di classificazione *in chiave sezionale* – quella rilevante per i nostri scopi –, falsi positivi e falsi negativi si distribuiscono casualmente e, di conseguenza, non hanno effetti distorsivi sulle misure di mobilità. In quest'ottica, è ragionevole congetturare che la partizione lavoro dipendente/indipendente sia robusta. Vi è qualche incertezza, invece, sulla casualità degli errori nelle risposte sull'attività economica, essenzialmente in conseguenza della minore attendibilità delle risposte in particolare per episodi conclusi, quindi rilevati con quesiti retrospettivi (vedi ancora Bound *et al.*, 2001, pp. 3802-3805). Il tema è largamente inesplorato per la RTFL. Un recente, interessante contributo è di de Angelini e Giraldo (2003). Nel quadro di un confronto fra misure di mobilità dalla RTFL e da Netlabor (l'archivio amministrativo dei centri per l'impiego) per il Veneto, esse conducono un'analisi della coerenza fra data di inizio del lavoro e condizione dichiarata *in chiave dinamica*, su panel a due occasioni – a distanza di un trimestre e di un anno – costruiti dalla RTFL. La conclusione cui giungono è la seguente: “Gli errori nella dichiarazione della posizione nella professione e/o nell'attività possono contribuire a sovrastimare notevolmente il numero di associazioni. Per un alto verso, gli errori di memoria relativi alla data di inizio del rapporto di lavoro dichiarata dall'intervistato sembrerebbero piuttosto contenuti” (de Angelini e Giraldo, 2003, pag. 24). Palesemente, essa conforta la scelta che abbiamo operato.

La procedura di standardizzazione della popolazione è concettualmente analoga, ma più laboriosa quando si usi il metodo dell'abbinamento, perché per la RTFL standardizzazione della popolazione, processi di mobilità e disegno campionario con rotazione interagiscono. In tal caso si procede come segue. (i) Innanzitutto, per la RTFL vengono abbinati i *data set* originari, riferiti cioè all'intero campione, con la procedura di *record linkage* di Paggiaro e Torelli (1999)¹⁸. Lo scopo è evidente: non perdere individui che abbiano riportato brevi episodi lavorativi – al limite un solo episodio lavorativo – nell'anno, quindi di interesse per le analisi di mobilità¹⁹. (ii) In secondo luogo, dal *file* abbinato della RTFL si rimuovono gli individui per i quali non è riportato alcun episodio lavorativo nell'anno. (iii) Infine, per giungere alla popolazione standardizzata vengono individuati i singoli episodi da eliminare nelle due fonti secondo la procedura già illustrata, cioè a dire esaminando la posizione nella professione (dipendente/indipendente) e l'attività economica relative all'episodio lavorativo in questione²⁰.

I risultati sono riassunti nella Tavola 3. Merita di essere notato come per la RTFL il meccanismo di rotazione del campione, in combinazione con la procedura di standardizzazione della popolazione, comporti una sensibile riduzione nella dimensione del campione della popolazione finale di interesse – gli individui con almeno un episodio lavorativo in corso, nelle attività economiche pertinenti, anche solo in parte durante l'anno –. Ciò non avviene, invece, per il campione INPS. È questa l'ovvia conseguenza del fatto che in INPS la riduzione della numerosità campionaria è dovuta unicamente alla procedura di standardizzazione²¹.

4.2.2. La simulazione sui data set INPS dei metodi di osservazione definiti a partire dalla RTFL

Per svolgere analisi di mobilità sulle due fonti comparabili anche sotto il profilo della base informativa sulle storie lavorative, i piani di osservazione propri della RTFL – riassunti nei quattro metodi prima menzionati – vengono simulati sui dati INPS.

¹⁸ Nei suoi lineamenti essenziali, tale procedura è stata adottata dall'Istat per la produzione di *files* abbinati, preliminare alla stima delle matrici di transizione delle forze di lavoro (Istat, 2002). Per generare *files* abbinati a tre occasioni semplicemente iteriamo la procedura, applicandola per il *record linkage* fra il *file* abbinato a due occasioni e quello della terza. Per il riporto all'universo utilizziamo i pesi campionari della prima occasione di indagine. La scelta è pressoché obbligata alla luce delle informazioni disponibili (per contrasto, vedi il procedimento, esigente quanto a informazioni esogene sui movimenti demografici e laborioso, utilizzato in Istat, 2002, per produrre la cosiddetta “popolazione longitudinale” e conseguentemente i pesi longitudinali). D'altra parte, la distorsione indotta da questa scelta semplificatrice è verosimilmente modesta.

¹⁹ Tali individui potrebbero non essere considerati se si restringesse il *record linkage* a sottoinsiemi selezionati sulla base di una qualche condizione lavorativa: ad esempio, lo condizione di occupato – in corso o precedentemente – in una delle *cross-section*.

²⁰ Per la RTFL, a fronte di incoerenze nella posizione nella professione e/o nell'attività economica riportate – per lo stesso episodio – in occasioni di indagine diverse, si segue il criterio enunciato nella nota 14.

²¹ Infatti, il *data set* INPS comprende soltanto individui che hanno avuto rapporti di lavoro nell'anno. Quindi, quando vi sia simulato il piano di osservazione ‘abbinamento a tre occasioni’, non risente di alcun effetto di rotazione.

TAVOLA 3

Numerosità campionarie dei data set INPS e RTFL abbinati a tre occasioni ai vari passi della procedura di standardizzazione della popolazione

Popolazione di riferimento	INPS		RTFL	
	95.I-II-96.I	95.I-IV-96.I	95.I-II-96.I	95.I-IV-96.I
Totale occupati	–	–	13.600	14.024
- indipendenti	129.710	129.710	9.950	10.366
- ATECO 01,10	115.409	115.414	8.431	8.799
- ATECO 11,12 ^a	108.273	108.278	6.518	6.752
Popol. standardizzata	104.046	104.058	6.028	6.272

^a Fatte salve le divisioni 91 e 93 (vedi l'Appendice 1).

Con riferimento a ciascuno dei primi tre metodi, in sostanza per ciascun individuo del panel INPS si simula un'osservazione che si colloca nel mese della corrispondente rilevazione *cross-section* RTFL ed ha le medesime caratteristiche di quest'ultima (ad esempio, per il metodo della condizione: condizione lavorativa corrente e data di inizio del corrente rapporto di lavoro o di fine dell'ultimo rapporto di lavoro)²².

La simulazione del metodo dell'abbinamento richiede qualche ulteriore precisazione. (i) Per ciascun mese corrispondente a un'occasione d'indagine RTFL, per ciascun individuo del panel INPS si simula una 'osservazione' con le caratteristiche proprie del metodo della condizione. L'individuo viene considerato occupato se ha un rapporto di lavoro in corso all'inizio del mese in questione; altrimenti viene considerato non occupato (indipendentemente dal fatto che abbia concluso un rapporto di lavoro in precedenza o che ne abbia iniziato uno in seguito). (ii) Quando per un individuo si abbiano molteplici episodi lavorativi, vengono eliminati tutti meno uno, così da disporre di un unico rapporto di lavoro per ogni individuo a ogni occasione. L'episodio lavorativo che viene mantenuto è quello considerato predominante: per gli individui occupati all'inizio del mese si tratta del rapporto di lavoro in corso iniziato per primo e, in caso di parità, finito per ultimo; per gli individui non occupati all'inizio del mese e precedentemente occupati si tratta del rapporto di lavoro – tra quelli terminati in precedenza – finito per ultimo e, in caso di parità, iniziato per primo.

4.3. Uno schema sintetico per i confronti

In vista delle successive analisi empiriche comparate, è utile disporre di uno schema che fornisca indicazioni riassuntive sulle differenze nelle misure di mobilità dei lavoratori risultanti dalle due fonti, in particolare sull'incidenza dei due principali fattori di discrepanza:

- (a) la popolazione di riferimento (o, altrimenti detto, la composizione);
- (b) il piano di osservazione delle storie lavorative.

In questa sezione suggeriamo, appunto, uno schema che risponde allo scopo.

²² Per maggiori ragguagli, e per alcune particolarità relative al metodo della *tenure* progressa, vedi Baretta (2002), pp. 49-53.

Denotiamo con:

y_p e $d y_p$ le stime di una data misura di mobilità riferite alla fonte INPS e basate sul piano di osservazione proprio di tale fonte – il metodo del panel (p) –, calcolate rispettivamente sul campione totale e su quello standardizzato quanto a composizione (d);

x_m e $d x_m$ le stime della medesima misura di mobilità riferite alla fonte RTFL e basate su uno dei quattro piani di osservazione utilizzabili per tale fonte (m), calcolate rispettivamente sul campione totale e su quello standardizzato (d);

y_m e $d y_m$ le stime della medesima misura di mobilità riferite alla fonte INPS e basate sulla simulazione di uno dei quattro piani di osservazione definiti a partire dalla RTFL (m), calcolate rispettivamente sul campione totale e su quello standardizzato (d).

È immediato notare che si possono evidenziare le seguenti differenze di interesse:

- differenza originaria: $\alpha = y_p - x_m$;
- differenza a parità di composizione: $\beta = d y_p - d x_m$;
- differenza a parità di piano di osservazione: $\gamma = y_m - x_m$;
- differenza a parità di composizione e di piano di osservazione: $\delta = d y_m - d x_m$.

Su questa base è possibile decomporre la differenza originaria α , evidenziando gli effetti relativi marginali della composizione e del piano di osservazione, nonché l'effetto relativo congiunto dei due fattori. I risultati sono convenientemente riassunti nel prospetto-tipo della Figura 3, che – con riferimento a una data differenza originaria associata all'impiego per la fonte RTFL di un dato metodo – evidenzia: (i) l'effetto marginale della composizione, dato da $1 - \beta/\alpha$; (ii) l'effetto marginale del metodo, dato da $1 - \gamma/\alpha$; (iii) l'effetto congiunto della composizione e del metodo, dato da $1 - \delta/\alpha$ ²³.

I criteri di lettura per tabelle costruite con questo formato, utili per informare sinteticamente sulla *performance* dei vari metodi, sono riassumibili in poche, scarse proposizioni. I risultati sono tanto migliori quanto più: (i) è possibile utilizzare misure di mobilità del tipo t_f e r_f ; (ii) l'effetto congiunto è prossimo a 1; (iii) l'effetto marginale del metodo è prossimo a 0 e, all'opposto, l'effetto marginale della composizione è prossimo a 1.

<i>Differenza originaria: α</i>	Effetto marginale del metodo: $1 - \gamma / \alpha$
Effetto marginale della composizione: $1 - \beta / \alpha$	Effetto congiunto: $1 - \delta / \alpha$

Figura 3 – Schema di decomposizione della differenza originaria α fra misure ‘grezze’ di mobilità dei lavoratori da INPS e RTFL.

²³ Si noti che i diversi effetti relativi non sono costretti entro l'intervallo [0,1]; in particolare, gli effetti marginali possono essere negativi, mentre l'effetto congiunto può essere negativo o, all'opposto, maggiore di 1. Lo schema può essere poi ulteriormente articolato, per evidenziare anche effetti relativi condizionati del metodo (della composizione) al netto della composizione (del metodo): vedi Baretta (2002), pp. 103-107. La portata informativa di questo affinamento è tuttavia trascurabile.

5. EVIDENZE EMPIRICHE PER IL METODO DELL'ABBINAMENTO A TRE OCCASIONI

5.1. I risultati col metodo del panel per i dati di fonte INPS

Il *benchmark* per le analisi comparate è rappresentato dalle misure di mobilità stimate sulla fonte INPS con il metodo del panel, cioè a dire utilizzando pienamente l'informazione sulla storia lavorativa a cadenza mensile che la fonte offre. Per il calcolo delle diverse misure di mobilità, abbiamo applicato ed esteso la procedura di Leombruni e Quaranta (2002) per la correzione dei flussi di associazioni e separazioni agli estremi dell'anno di calendario. (Per dettagli, vedi l'Appendice 2.)

Segnaliamo inoltre che, per questa come per le successive analisi empiriche incentrate sul metodo dell'abbinamento a tre occasioni, alle stime puntuali delle misure di mobilità affianchiamo gli intervalli di confidenza al 95% stimati tramite *bootstrapping* (Efron e Tibshirani, 1986), con 200 replicazioni. Essi tornano utili, in particolare, per valutazioni statisticamente fondate sulla 'vicinanza' o meno delle misure di mobilità tratte dalle due fonti. Aggiungiamo, infine, che in generale presentiamo i risultati per l'intera batteria di indicatori di mobilità dei lavoratori – a , s , t e r –, ma in sede di commento portiamo l'attenzione soprattutto sul *gross worker turnover* e sul tasso di riallocazione.

I risultati riferiti all'anno 1995, per la popolazione originaria del *data set* INPS e per quella standardizzata, sono nella Tavola 4. Va notato che le misure di mobilità calcolate sulla popolazione standardizzata sono generalmente inferiori a quelle ottenute sulla popolazione originaria: ad esempio, il tasso t_f passa dal 52,4% al 51,1%. La ragione è presto detta. La standardizzazione della popolazione porta all'esclusione di lavoratori/episodi lavorativi nella pubblica amministrazione, nell'istruzione e nella sanità, che sono presenti negli archivi INPS in misura assai limitata e per situazioni particolari (lavoratori non di ruolo, selezionati episodi di lavoratori di ruolo), e hanno un comportamento mediamente caratterizzato da un'alta mobilità, dunque ben lontano da quello della restante occupazione negli stessi comparti – notoriamente, la pubblica amministrazione è il ramo meno mobile dell'economia italiana –.

TAVOLA 4

Misure di mobilità dei lavoratori (%) da INPS, anno 1995, utilizzando pienamente il piano di osservazione (in corsivo intervalli di confidenza al 95%)

Aggregato di riferimento	Tassi di associazione				Tassi di separazione			
	a_s	$u a_s$	a_f	a_{fp}^a	s_s	$u s_s$	s_f	s_{fp}^a
Occ. dipendenti	33,3	19,5	26,6	30,2	32,3	19,8	25,8	29,3
	<i>32,8;33,7</i>	<i>19,3;19,8</i>	<i>26,3;26,9</i>	<i>29,9;30,6</i>	<i>31,9;32,7</i>	<i>19,5;20,1</i>	<i>25,5;26,1</i>	<i>29,0;29,7</i>
Pop. standarizzata	32,2	19,6	26,2	29,8	30,6	19,2	24,9	28,4
	<i>31,7;32,6</i>	<i>19,3;20,0</i>	<i>25,8;26,5</i>	<i>29,4;30,2</i>	<i>30,2;31,0</i>	<i>18,9;19,5</i>	<i>24,6;25,2</i>	<i>28,0;28,8</i>
	<i>Gross worker turnover</i>				Tassi di riallocazione			
	t_s	$u t_s$	t_f	t_{fp}^a	r_s	$u r_s$	r_f	
Occ. dipendenti	65,5	39,3	52,4	59,6	41,0	31,5	32,8	
	<i>64,8;66,2</i>	<i>38,9;39,8</i>	<i>51,9;52,9</i>	<i>59,0;60,2</i>	<i>40,6;41,4</i>	<i>31,1;31,8</i>	<i>32,5;33,1</i>	
Pop. standarizzata	62,8	38,8	51,1	58,2	40,6	31,5	33,0	
	<i>62,1;63,6</i>	<i>38,3;39,2</i>	<i>50,5;51,7</i>	<i>57,4;58,9</i>	<i>40,2;41,0</i>	<i>31,1;31,8</i>	<i>32,7;33,3</i>	

^a Indicatori calcolati secondo la formulazione di Leombruni e Quaranta (2002), pp. 180-181.

Per la popolazione standardizzata – alla quale soprattutto guardiamo, in vista del confronto delle misure di mobilità dalle due fonti –, il *gwt* è di poco superiore al 51%²⁴. Ciò significa che in media e in condizioni di stato stazionario un lavoratore opera una transizione da un lavoro a un altro in poco meno di due anni. Per il tasso di riallocazione si ha un valore del 33%, il che significa che, sempre in condizioni di stato stazionario, circa un lavoratore su tre è mediamente coinvolto in almeno una transizione durante l'anno.

Questi risultati si affiancano a quelli di parecchi altri analisti (da ultimo, vedi Leombruni e Quaranta, 2002) e suonano a conferma della marcata mobilità del mercato del lavoro italiano nei settori considerati. L'elevato valore del *gwt* e, soprattutto, la sua notevole distanza dal tasso di riallocazione chiamano in causa il ruolo giocato dalla presenza di individui con più di un rapporto di lavoro in corso anche solo in parte durante l'anno. Gli individui con episodi multipli sono il 9,0% del totale, e sono responsabili del 35% circa del *gwt* e del 18% circa del tasso di riallocazione.

5.2. I risultati col metodo dell'abbinamento a tre occasioni

I risultati salienti per il metodo dell'abbinamento a tre occasioni sono riportati nelle Tavole 5, 6 e 7. Per la RTFL essi riguardano sia entrambi i campioni abbinati che possono essere costruiti per il 1995, rispettivamente sulle occasioni 95.I-II-96.I e 95.I-IV-96.I, sia il campione risultante dal *pooling* dei due precedenti. Si è di fronte, infatti, a due campioni disgiunti e indipendenti dalla stessa popolazione, che si distinguono per la collocazione della rilevazione intermedia – rispettivamente ad aprile e ad ottobre –; si possono quindi aggregare e si ottiene un campione *pooled* di numerosità grosso modo doppia, che fornisce stime della mobilità dei lavoratori parecchio più precise. Parallelamente, per INPS si presentano i risultati riferiti alla simulazione dei piani di osservazione dati dalle due sequenze di tre occasioni appena menzionate e, più sinteticamente, quelli relativi alla loro media.

Le evidenze salienti possono essere riassunte in due ordini di considerazioni. In primo luogo, per la popolazione standardizzata e a parità di piano di osservazione vi è una notevole vicinanza delle stime della mobilità fornite dalle due fonti. I tassi di riallocazione sono pressoché uguali: intorno al 33%, con differenze assolute che non superano i 0,4 punti percentuali, quindi dell'ordine dell'1%. Gli intervalli di confidenza per le stime di fonte RTFL, nettamente più ampi a motivo della più ridotta dimensione campionaria (e, comunque, apprezzabilmente accorciati per la

²⁴ Salva diversa segnalazione, qui è nel seguito facciamo riferimento a t_f e in generale a misure che rapportano i movimenti alla popolazione flusso (vedi la sez. 3). La Tavola 4 riporta anche le misure di mobilità secondo la formulazione di Leombruni e Quaranta (2002). Per definizione, esse si collocano entro l'intervallo dato dalle corrispondenti misure che hanno come denominatore la popolazione flusso (estremo inferiore) e quella di *stock* (estremo superiore). Empiricamente, è interessante notare che in generale sono parecchio più vicine alle misure riferite allo *stock* di popolazione (ad esempio, per la popolazione standardizzata il *gwt* calcolato con la procedura di Leombruni e Quaranta è 58,2%, a fronte di $t_f = 51,1\%$ e di $t_s = 62,8\%$) e che questa maggiore prossimità è accentuata per i tassi di separazione.

TAVOLA 5

Misure di mobilità dei lavoratori (%) da RTFL e INPS, col metodo dell'abbinamento a tre occasioni
(in corsivo intervalli di confidenza al 95%)

Abbinamento e pop. di riferimento	RTFL				INPS			
	<i>a_f</i>	<i>s_f</i>	<i>t_f</i>	<i>r_f</i>	<i>a_f</i>	<i>s_f</i>	<i>t_f</i>	<i>r_f</i>
<i>95.I-II-96.I</i>								
Totale occupati	16,9	15,8	32,7	20,6	-	-	-	-
	<i>16,1;17,7</i>	<i>15,0;16,6</i>	<i>31,2;34,1</i>	<i>19,7;21,4</i>				
- lav. indipendenti	20,2	18,7	38,9	24,9	23,9	22,5	46,3	32,1
	<i>19,3;21,0</i>	<i>17,7;19,6</i>	<i>37,4;40,3</i>	<i>24,0;25,7</i>	<i>23,6;24,1</i>	<i>22,2;22,7</i>	<i>45,8;46,8</i>	<i>31,8;32,4</i>
Pop. standardizzata	24,7	22,8	47,5	32,8	23,9	22,0	45,9	32,4
	<i>23,4;26,0</i>	<i>21,6;24,1</i>	<i>45,4;50,0</i>	<i>31,5;34,1</i>	<i>23,6;24,2</i>	<i>21,7;22,3</i>	<i>45,4;46,4</i>	<i>32,1;32,7</i>
<i>95.I-IV-96.I</i>								
Totale occupati	17,4	16,7	34,1	21,4	-	-	-	-
	<i>16,7;18,2</i>	<i>16,0;17,5</i>	<i>32,9;35,5</i>	<i>20,6;22,1</i>				
- lav. indipendenti	20,5	19,7	40,2	25,4	25,0	22,9	47,9	32,6
	<i>19,6;21,5</i>	<i>18,7;20,6</i>	<i>38,5;41,8</i>	<i>24,5;26,4</i>	<i>24,7;25,3</i>	<i>22,6;23,2</i>	<i>47,4;48,4</i>	<i>32,3;32,9</i>
Pop. standardizzata	24,9	23,1	48,0	33,1	24,8	22,3	47,1	32,8
	<i>23,5;26,0</i>	<i>21,8;24,1</i>	<i>45,8;50,0</i>	<i>31,6;34,3</i>	<i>24,5;25,0</i>	<i>22,0;22,6</i>	<i>46,6;47,6</i>	<i>32,5;33,1</i>
<i>pooled</i>								
Totale occupati	17,2	16,3	33,4	21,0				
	<i>16,7;17,7</i>	<i>15,8;16,8</i>	<i>32,5;34,5</i>	<i>20,5;21,5</i>				
- lav. indipendenti	20,3	19,2	39,6	25,2				
	<i>19,7;21,0</i>	<i>18,5;19,9</i>	<i>38,2;40,7</i>	<i>24,4;25,8</i>				
Pop. standardizzata	24,8	23,0	47,8	32,9				
	<i>23,9;25,5</i>	<i>21,9;23,9</i>	<i>46,1;49,2</i>	<i>32,0;33,7</i>				

stima *pooled*), contengono completamente gli intervalli delle corrispondenti stime di fonte INPS, il che sta a indicare che le differenze riscontrate nelle stime puntuali non sono statisticamente significative. Per il *gwt* si hanno risultati simili, anche se con distanze fra le stime dalle due fonti lievemente maggiori. Per la RTFL i valori si attestano sul - o poco al di sotto del - 48% , superiori ai corrispondenti tassi da INPS di 1-1,5 punti percentuali, dunque del 2-3% in termini relativi. Anche in questo caso, peraltro, gli intervalli di confidenza delle stime di fonte INPS sono sistematicamente contenuti entro i corrispondenti intervalli di fonte RTFL (ciò vale tanto per il *gwt* che per le sue due componenti, i tassi di associazione e di separazione).

In secondo luogo, la disaggregazione delle misure di mobilità per area geografica, genere e classe d'età dei lavoratori non segnala l'emergere di diffuse difformità tra le due fonti (vedi Tavola 6, colonne 3-6 per i raffronti ancora riferiti alla popolazione standardizzata e al metodo dell'abbinamento a tre occasioni). Naturalmente i divari aumentano, ma si tratta di una risultanza attesa: vi è infatti una maggiore variabilità delle stime, conseguente alla più ridotta dimensione dei subcampioni sui quali sono di volta in volta calcolate - fenomeno questo che è marcato soprattutto per la RTFL (l'intero campione *pooled* supera di poco i 12mila lavoratori) -. Le differenze fra le stime fornite dalle due fonti, infatti, in molti casi restano statisticamente non significative, com'è evidente dal fatto che i rispettivi intervalli di confidenza si sovrappongono. Divari sensibili, statisticamente significativi, si hanno tuttavia per due segmenti: con riguardo all'area geografica, per il Sud; con riguardo alle caratteristiche demografiche, per i lavoratori anziani di entrambi i generi, ma in particolare donne - il che si riflette anche in un apprezzabile divario per l'insieme delle donne -.

TAVOLA 6
 Misure di mobilità (%) da RTFL e INPS per area geografica, genere e classe di età
 (in corsivo intervalli di confidenza al 95%)

	INPS medio, popolazione standardizzata				RTFL <i>pooled</i>			
	Metodo del panel		Met. abb. 3 occ.		Pop. standardizzata		Pop. totale	
	t_f	r_f	t_f	r_f	t_f	r_f	t_f	r_f
Totale	51,1 <i>50,5;51,6</i>	33,0 <i>32,7;33,3</i>	46,5 <i>46,1;46,8</i>	32,6 <i>32,4;32,7</i>	47,8 <i>46,1;49,1</i>	32,9 <i>31,8;33,8</i>	33,4 <i>32,5;34,4</i>	21,0 <i>20,5;21,6</i>
<i>Ripartiz. geografica</i>								
Nord	50,2 <i>49,5;50,9</i>	31,5 <i>31,1;31,9</i>	44,2 <i>43,8;44,6</i>	30,5 <i>30,3;30,7</i>	42,8 <i>40,5;44,8</i>	29,1 <i>27,8;30,5</i>	31,0 <i>29,6;32,2</i>	19,5 <i>18,7;20,3</i>
Centro	48,0 <i>46,8;49,1</i>	32,6 <i>32,0;33,2</i>	45,7 <i>45,0;46,3</i>	33,1 <i>32,7;33,6</i>	45,5 <i>42,3;48,6</i>	31,9 <i>30,0;33,9</i>	28,8 <i>27,1;30,9</i>	18,5 <i>17,6;19,6</i>
Sud	55,8 <i>54,6;57,2</i>	39,9 <i>39,3;40,8</i>	53,4 <i>52,9;54,0</i>	39,2 <i>38,9;39,7</i>	62,5 <i>59,0;65,0</i>	43,7 <i>41,5;45,3</i>	40,9 <i>39,2;42,9</i>	25,3 <i>24,3;26,2</i>
<i>Genere</i>								
Maschi	52,0 <i>51,3;52,7</i>	32,8 <i>32,4;33,2</i>	47,0 <i>46,6;47,4</i>	32,6 <i>32,4;32,8</i>	46,0 <i>44,1;47,5</i>	31,9 <i>30,7;32,8</i>	29,8 <i>28,7;30,9</i>	18,7 <i>18,1;19,4</i>
Femmine	49,3 <i>48,3;50,3</i>	33,4 <i>32,8;33,9</i>	45,4 <i>44,9;45,9</i>	32,4 <i>32,1;32,7</i>	51,4 <i>48,7;54,1</i>	35,1 <i>33,6;36,9</i>	39,7 <i>38,0;41,5</i>	25,1 <i>24,2;26,1</i>
<i>Classe d'età</i>								
14-19	117,4 <i>114,7;120,1</i>	77,0 <i>75,5;78,3</i>	98,1 <i>96,5;99,8</i>	64,5 <i>63,5;65,4</i>	91,3 <i>83,4;98,3</i>	61,7 <i>57,7;65,8</i>	96,2 <i>90,7;101,9</i>	66,0 <i>62,8;69,5</i>
20-24	86,1 <i>84,3;87,7</i>	55,2 <i>54,2;56,0</i>	72,9 <i>71,9;74,0</i>	49,0 <i>48,4;49,5</i>	69,1 <i>62,7;73,9</i>	46,9 <i>43,8;49,8</i>	67,7 <i>63,5;71,6</i>	43,1 <i>41,1;45,5</i>
25-34	50,9 <i>50,0;51,8</i>	32,5 <i>32,0;33,0</i>	46,3 <i>45,8;46,8</i>	32,2 <i>31,8;32,5</i>	41,9 <i>39,4;43,8</i>	29,3 <i>27,8;30,6</i>	33,4 <i>31,5;35,0</i>	20,5 <i>19,4;21,5</i>
35-44	36,0 <i>35,0;36,9</i>	22,9 <i>22,3;23,4</i>	34,7 <i>34,1;35,2</i>	24,8 <i>24,4;25,1</i>	38,0 <i>35,3;41,2</i>	26,8 <i>25,1;28,8</i>	21,1 <i>19,6;22,5</i>	12,8 <i>11,9;13,6</i>
45-54	28,7 <i>27,7;29,7</i>	18,6 <i>18,0;19,3</i>	29,2 <i>28,7;29,9</i>	21,8 <i>21,4;22,2</i>	36,0 <i>32,9;38,8</i>	26,8 <i>24,5;28,6</i>	19,9 <i>18,5;21,6</i>	13,2 <i>12,3;14,1</i>
55+	40,0 <i>37,8;41,9</i>	29,6 <i>28,2;31,3</i>	39,5 <i>38,4;40,7</i>	32,6 <i>31,8;33,5</i>	64,5 <i>59,9;69,6</i>	48,1 <i>44,2;52,3</i>	40,0 <i>37,6;42,3</i>	28,8 <i>27,2;30,3</i>
<i>Genere e classe età (≥45)</i>								
Maschi 45-54	28,8 <i>27,6;30,2</i>	18,3 <i>17,6;19,0</i>	29,3 <i>28,7;30,0</i>	21,8 <i>21,4;22,3</i>	33,4 <i>30,2;37,1</i>	24,4 <i>22,1;27,0</i>	17,5 <i>15,8;19,2</i>	11,8 <i>10,7;12,8</i>
55+	40,1 <i>37,8;42,3</i>	29,1 <i>27,7;30,7</i>	39,5 <i>38,0;40,8</i>	32,5 <i>31,4;33,3</i>	63,8 <i>58,1;70,8</i>	47,3 <i>43,8;51,3</i>	37,2 <i>34,6;40,5</i>	26,7 <i>24,7;28,8</i>
Femmine 45-54	28,4 <i>27,0;30,5</i>	19,5 <i>18,5;20,8</i>	29,2 <i>28,1;30,4</i>	21,9 <i>21,2;22,7</i>	43,1 <i>37,1;48,3</i>	33,2 <i>28,7;36,9</i>	24,5 <i>21,8;27,1</i>	16,1 <i>14,4;17,5</i>
55+	39,6 <i>34,7;44,2</i>	31,5 <i>28,4;35,3</i>	39,4 <i>36,9;41,9</i>	33,3 <i>31,5;35,2</i>	69,5 <i>56,4;82,6</i>	53,4 <i>45,4;62,0</i>	48,3 <i>42,8;54,1</i>	35,1 <i>31,8;38,7</i>

- Quanto al Sud, si hanno stime dei tassi di riallocazione e soprattutto di *turnover* più alte dalla fonte RTFL (t_f e r_f pari rispettivamente a 62,5% e a 43,7% contro il 53,4% e il 39,2% dall' INPS). Fra i fattori che concorrono a spiegare il divario, vi è la difformità nelle definizioni accolte dalle due fonti – occupati presenti, secondo il luogo di lavoro, per l' INPS *vs.* occupati residenti per la RTFL –²⁵. Verosimilmente, vi contribuiscono poi le più marcate divergenze fra

²⁵ Per la RTFL abbiamo utilizzato l'informazione sulla residenza della famiglia, e non quella sulla regione di lavoro, per due ragioni: (i) il quesito sulla regione di lavoro è posto soltanto per gli episodi lavorativi in corso alla data dell'indagine, non per quelli rilevati retrospettivamente; (ii) inoltre, per gli episodi lavorativi in corso nella grande maggioranza dei casi regione di lavoro e di residenza coincidono (nelle cinque rilevazioni che vanno dal 1995.I al 1996.I, ciò accade in media per il 97,3% delle risposte valide, mentre le mancate risposte al quesito sulla regione di lavoro si attestano sul 2,7%).

le due fonti che si manifestano nel Sud in tema di mobilità dei lavoratori anziani, delle quali diciamo nel punto immediatamente seguente.

- Ancora più netto, e insieme di meno agevole lettura, è il quadro che si presenta per i lavoratori anziani, soprattutto dell'ultima classe di età – 55 anni e più –. È ancora la fonte RTFL a fornire stime della mobilità parecchio più elevate²⁶. Per le donne il fenomeno si manifesta già a partire dalla classe di età 35-44 anni, e si fa netto per quella successiva – 45-54 anni –. Per l'ultima classe di età, poi, i divari sono molto forti per entrambi i generi, con le stime di fonte RTFL superiori a quelle tratte dall'INPS di oltre il cinquanta per cento (guardando al *gwt*, 63,8% contro 39,5% per gli uomini, 69,5% contro 39,4% per le donne). Il fatto che il numero di lavoratori-campione in questa classe di età, uomini e soprattutto donne, sia piuttosto esiguo contribuisce certo a dare conto del fenomeno. Ma l'evidenza di intervalli di confidenza disgiunti, parecchio lontani l'un l'altro, chiama in causa anche altre spiegazioni. Un ragionevole congettura è che parte dei lavoratori in età avanzata, con rapporti di lavoro poco stabili – in larga misura pensionati –, non compaia nell'archivio dei lavoratori dipendenti dell'Inps²⁷ (o comunque vi compaia per una frazione ridotta degli episodi lavorativi), e dichiarare invece gli episodi lavorativi (o comunque li dichiarare in misura sensibilmente superiore) in sede di RTFL. Le differenze riscontrate tra i generi – con divari nelle stime della mobilità anticipati quanto ad età e maggiori per le donne, – sono coerenti con questa congettura: l'età di pensionamento delle donne è infatti inferiore a quella degli uomini; inoltre, per le ultra trentacinquenni – che (ri)entrano nel mercato del lavoro dopo l'impegno dedicato ai figli – è plausibilmente maggiore l'incidenza di esperienze di 'lavoro sommerso' *tout court* (e non solo a valle del pensionamento)²⁸.

Pur con i *caveat* suggeriti da queste ultime evidenze, l'insieme dei raffronti ci consente di concludere che, a parità di popolazione di riferimento e di piano di osservazione – quello del metodo dell'abbinamento a tre occasioni –, le stime della mobilità dei lavoratori tratte dalla due fonti non sono significativamente diverse, in particolare a livello aggregato. Risulta dunque corroborato quanto affermato nella sez. 2, e riassunto nella Tavola 1, in sede di ricognizione delle differenze tra

²⁶ È interessante notare che per le prime tre classi di età il divario è di segno opposto, con le stime della mobilità da INPS sistematicamente superiori alle corrispondenti da RTFL, peraltro in generale in misura statisticamente non significativa. L'eccezione sono i maschi in età 25-34 anni, per i quali le stime sia del *turnover* che del tasso di riallocazione sono significativamente differenti, e per l'appunto maggiori in INPS.

²⁷ Altrimenti detto, per essi vi sarebbe evasione contributiva totale. Con riguardo alla stima dell'occupazione dipendente privata extra-agricola (non, dunque, alla mobilità dei lavoratori), l'esistenza di un divario fra Inps e RTFL è ben documentata, ad esempio, in Zanzottera (2003) per la Lombardia: le stime della fonte Inps sono sistematicamente più basse e lo scarto tra le due fonti è crescente al crescere dell'età, in particolare per gli ultracinquantenni.

²⁸ Come già segnalato, queste divergenze sono poi più marcate nel Sud, dove il peso del 'lavoro sommerso' è notoriamente maggiore. Ci limitiamo a riportare le differenze, in punti percentuali, nelle stime del *gwt* tratte dalle due fonti (RTFL – INPS) per le due ultime classi di età, per il Sud e, tra parentesi, per il Centro-Nord. Nella classe 45-54 anni abbiamo: per il totale 14,4 (contro 4,5); per i maschi 12,1 (0,9), per le femmine 29,4 (12,1). Nella classe di età 55 anni e oltre, abbiamo poi: per il totale 28,0 (23,4); per i maschi 25,9 (23,1), per le femmine 49,6 (26,2).

le due fonti: le numerose altre differenze, pressoché irriducibili, hanno un peso trascurabile e non precludono lo svolgimento di analisi comparate della mobilità.

Naturalmente, per rispondere all'interrogativo cruciale sull'utilizzabilità della RTFL per soddisfacenti misure della mobilità dei lavoratori, occorre confrontare le stime tratte dalla RTFL, col metodo dell'abbinamento a tre occasioni, con le stime ottenute dalla fonte INPS utilizzandola al meglio. Una prima, confortante indicazione è data dalla Tavola 7, che presenta gli indicatori di decomposizione delle differenze nelle misure di mobilità t_f e r_f calcolate sulle popolazioni originarie di INPS e RTFL, rispettivamente con il metodo del panel e con quello dell'abbinamento a tre occasioni. Essi documentano un effetto congiunto della composizione e del metodo di poco superiore a 1, il peso dominante dell'effetto della composizione e, all'opposto, il peso abbastanza contenuto dell'effetto del metodo. Inoltre, v'è da notare che i risultati sono migliori per i tassi di riallocazione che non per il *gwt*.

TAVOLA 7

*Indicatori di decomposizione delle differenze nelle misure di mobilità:
risultati per il metodo dell'abbinamento a tre occasioni (misure di mobilità %)*

<i>Gross worker turnover t_f</i>		<i>Tasso di riallocazione r_f</i>	
Abbinamento 95.I-II-96.I			
INPS: 52,4	RTFL: 32,7	INPS: 32,8	RTFL: 20,6
19,7	0,309	12,2	0,058
0,818	1,083	0,981	1,032
Abbinamento 95.I-IV-96.I			
INPS: 52,4	RTFL: 34,1	INPS: 32,8	RTFL: 21,4
18,3	0,245	11,4	0,016
0,831	1,051	1,010	1,031
RTFL <i>pooled</i> vs. INPS medio			
INPS: 52,4	RTFL: 33,4	INPS: 32,8	RTFL: 21,0
19,0	0,278	11,8	0,038
0,825	1,068	0,995	1,032

Le evidenze di diretto interesse vengono, ovviamente, dal confronto fra le stime della mobilità fornite dalle due fonti con riferimento alla popolazione standardizzata e utilizzando il miglior piano di osservazione disponibile per ciascuna (vedi Tavola 4 per le stime dall' INPS, la Tavola 5 per le stime dalla RTFL). Il confronto mette in luce, appunto, le differenze nelle misure della mobilità dovute al piano di osservazione – metodo del panel per l' INPS *vs.* metodo dell'abbinamento a tre occasioni per la RTFL –, a parità di popolazione. Restringendo l'attenzione, per la RTFL, ai risultati per il campione *pooled*, abbiamo che:

- per il tasso di riallocazione non vi sono differenze apprezzabili: le stime puntuali di r_f , rispettivamente per INPS e per RTFL, sono 33,0% e 32,9%; l'intervallo di confidenza della prima stima è completamente contenuto in quello della seconda;

- per il *gross worker turnover*, misurato dall'indicatore t_f , vi è una certa differenza, significativa ma contenuta: le stime puntuali sono, rispettivamente, 51,1% e 47,8%; i corrispondenti intervalli di confidenza sono disgiunti ((50,5-51,7) vs. (46,1-49,2)), il che porta appunto a concludere che le due stime sono significativamente diverse; la sottostima della mobilità da parte della RTFL, in termini di transizioni/episodi per lavoratore, è di 3,3 punti percentuali, dell'ordine del 6,5%, tutto sommato non vistosa.

Ulteriori, indicazioni interessanti sul *pattern* – e indirettamente sulle ragioni – alla base di questa sottostima sono offerte dalle Figure 4 e 5. Esse mostrano la frazione di episodi lavorativi (rispettivamente di transizioni) documentati nell'originario *data set* INPS che non sono rilevati quando vi sia simulato il metodo dell'abbinamento a tre occasioni, per numero di episodi (rispettivamente di transizioni) per lavoratore. Dalla loro ispezione si desume che:

- non si perdono transizioni singole – o in numero dispari –, ma coppie, quadruple, sestuple di transizioni (e così via): cioè a dire, si perdono episodi completi di breve durata interni al periodo di tre e/o di nove mesi, in cui il piano di osservazione dell'abbinamento a tre occasioni suddivide l'anno di riferimento;
- la frazione di episodi che si perdono cresce quanto più la storia lavorativa individuale ne è ricca. Tra i lavoratori con due episodi lavorativi nell'anno, in poco più del 20% dei casi si perde un episodio. Tra i lavoratori con tre episodi lavorativi, solo nel 15% dei casi tutti e tre gli episodi sono rilevati, mentre nella maggioranza dei casi – oltre il 70% – si perde un episodio. E via dicendo. In sintesi, si perdono episodi brevi in storie lavorative accidentate (che, al 1995, sono comunque una piccola frazione del totale: il 92% dei lavoratori ha un solo episodio nell'anno).

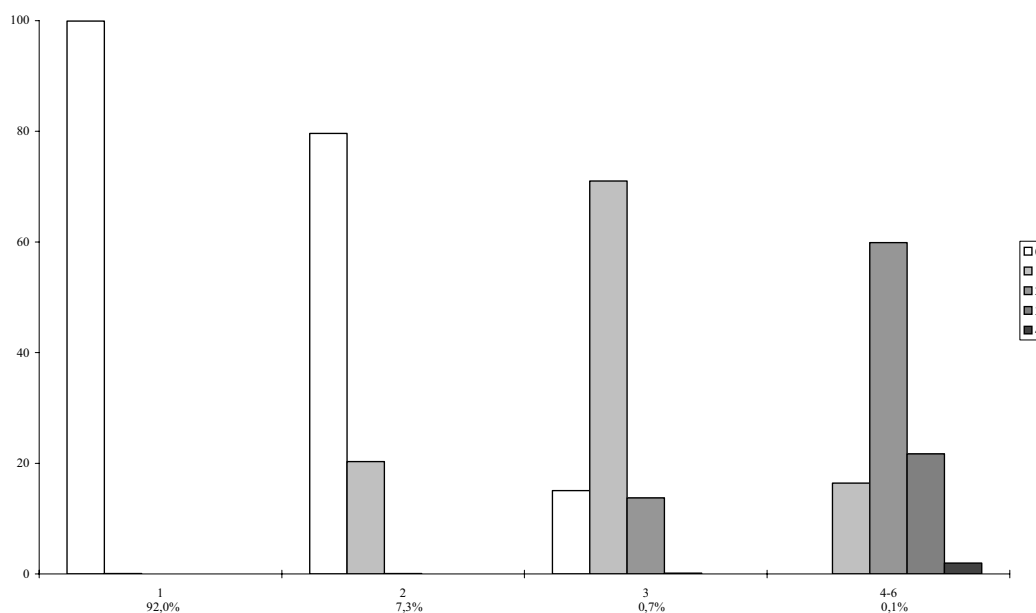


Figura 4 – Frazione di episodi lavorativi del *data set* INPS non rilevati dal piano di osservazione 'abbinamento a tre occasioni', per numero di episodi per lavoratore (anno 1995; popolazione standardizzata; totale lavoratori: 104.162).

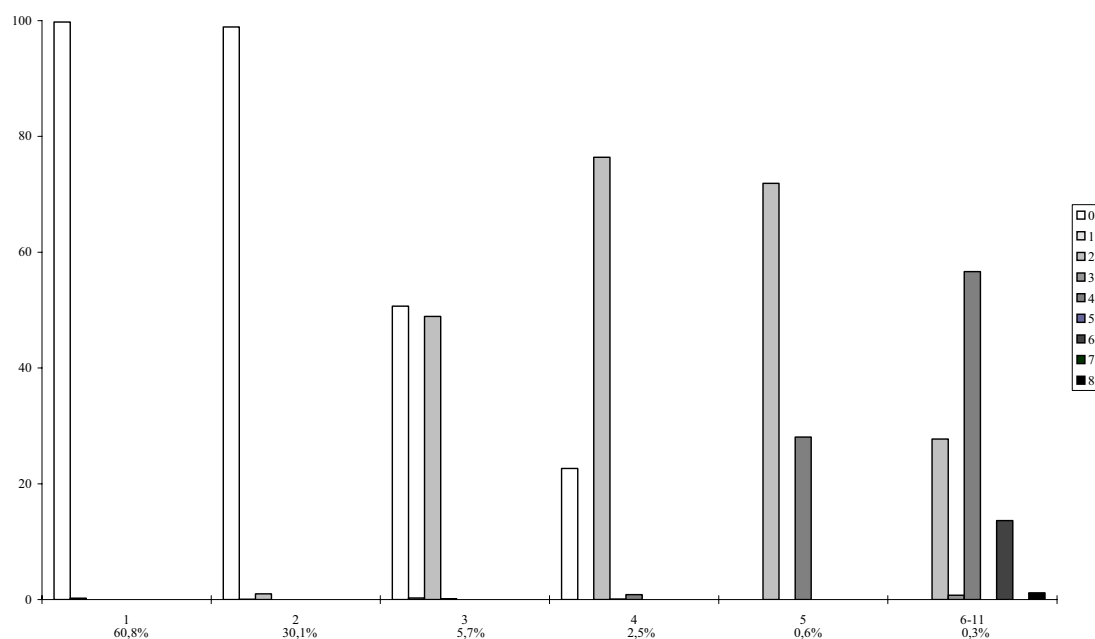


Figura 5 – Frazione di transizioni del *data set* INPS non rilevate dal piano di osservazione ‘abbinamento a tre occasioni’, per numero di transizioni per lavoratore (anno 1995; popolazione standardizzata; totale lavoratori con transizioni: 37.31).

6. CONSIDERAZIONI CONCLUSIVE

La Tavola 8 riassume i risultati salienti dell’intero esercizio, e ben si presta per considerazioni conclusive.

Analisi empiriche condotte sommariamente, senza tenere conto della differente popolazione di riferimento delle due fonti e senza sfruttare la dimensione longitudinale della RTFL, forniscono misure di mobilità assai distanti (Tavola 8, prima riga) e suggeriscono, a nostro avviso erroneamente, che la RTFL non è utilizzabile per stime della mobilità dei lavoratori.

Quando si sia standardizzata la popolazione di riferimento e al piano di osservazione originario per l’ INPS si contrapponga l’abbinamento a tre occasioni per la RTFL, le differenze nelle misure di mobilità fra INPS e RTFL sono abbastanza contenute (Tavola 8, quarta riga). In sostanza, rispetto al *benchmark* rappresentato dall’ INPS, con la RTFL cogliamo completamente la mobilità in termini di frazione di lavoratori interessati da transizioni, mentre sottostimiamo in misura contenuta il numero medio di transizioni operate dai lavoratori. La sottostima è concentrata sulla ridotta frazione (al 1995) di lavoratori con almeno due episodi nell’anno, e riguarda rapporti di lavoro brevi in storie lavorative accidentate.

La RTFL consente poi di estendere le misure di mobilità all’insieme degli occupati delle famiglie residenti (Tavola 8, quinta riga). Il quadro che ne emerge è sensibilmente diverso: il *gwt* scende dal 47,8% per la popolazione standardizzata i lavoratori dipendenti nell’industria e nel terziario privato – al 33,4% per l’intera occupazione; il tasso di riallocazione scende parallelamente dal 32,9% al 21,0%. È

TAVOLA 8

Sintesi dei risultati delle analisi comparate di mobilità, con il metodo dell'abbinamento a tre occasioni
(misure di mobilità %)

Popolazione di riferimento, piano di osservazione e tipo di misura	Gross worker turnover		Tasso di riallocazione	
	INPS	RTFL	INPS	RTFL
misure 'grezze' ^a	65,5	17,6	41,0	17,6
piano di osservazione e misure 'ottimali' ^b	52,4	–	32,8	–
confronti 'normalizzati' ^c	46,5	47,8	32,6	32,9
solo popolazione standardizzata ^d	51,1	47,8	33,0	32,9
solo metodo 'normalizzato' ^e	47,1	33,4	32,3	21,0

^a Popolazione diversa, piano di osservazione diverso (per RTFL da *cross-section*, metodo della condizione), misure t_i e r_i (per RTFL è $t_i \equiv r_i \equiv w_i t_i \equiv w_i r_i$).

^b Piano di osservazione della fonte INPS (metodo del panel), misure t_j e r_j .

^c Popolazione standardizzata, metodo dell'abbinamento a tre occasioni, INPS medio e RTFL *pooled*, misure t_j e r_j .

^d Popolazione standardizzata, piano di osservazione diverso (ottimale per ciascuna fonte: per INPS metodo del panel, per RTFL metodo dell'abbinamento a tre occasioni), RTFL *pooled*, misure t_j e r_j .

^e Popolazione diversa, metodo dell'abbinamento a tre occasioni, INPS medio e RTFL *pooled*, misure t_j e r_j .

inoltre interessante notare che, passando dalla popolazione standardizzata al totale degli occupati, si hanno apprezzabili modificazioni nel *pattern* di mobilità (vedi Tavola 6, ultime quattro colonne): i divari territoriali, pur restando marcati – con la mobilità parecchio più elevata nel Sud –, si riducono in maniera evidente (nel Sud il *gwt* scende dal 62,5 al 40,9%, mentre nel Nord la contrazione è dal 42,8 al 31,0%); guardando alla mobilità per genere e classe di età, la contrazione interessa soprattutto le classi di età mature e anziane (per gli ultra cinquantacinquenni il *gwt* scende dal 64,5 al 40,0%), ed è comparativamente più forte per gli uomini (dal 63,8 al 37,2%) che per le donne (dal 69,5 al 48,3%)²⁹.

Un'ultima, breve riflessione si impone, infine, in un'ottica di confronti internazionali. Certo, non è nei propositi di questo articolo una valutazione comparata del grado di mobilità dei lavoratori in Italia rispetto ai maggiori paesi sviluppati. I nostri dati suggeriscono, comunque, perlomeno due osservazioni: (i) nel metodo, servono attenzione e appropriate cautele nel raffrontare stime riferite a diversi paesi, che spesso differiscono per natura della fonte – *surveys* su famiglie *vs.* archivi amministrativi – e copertura della popolazione; (ii) nel merito, l'Italia è caratterizzata da una mobilità prossima a quella dei maggiori paesi europei – Regno Unito escluso – e, invece, sensibilmente inferiore a quella di USA, Canada e, appunto, Regno Unito³⁰.

In definitiva, l'evidenza empirica segnala la capacità del piano di osservazione 'abbinamento a tre occasioni', e quindi dei corrispondenti *data set* longitudinali della RTFL, di cogliere larghissima parte della mobilità dei lavoratori. In prospetti-

²⁹ Per la lettura di questi risultati valgono, naturalmente, le cautele suggerite dalle osservazioni precedenti in tema di misure di mobilità disaggregate. È da tenere presente, inoltre, che essi risentono in larga misura di effetti di composizione, cioè a dire della diversa distribuzione degli occupati – per area geografica, genere e classe di età – tra lavoro dipendente e indipendente e per branca di attività economica.

³⁰ Un compatto quadro di confronti internazionali è in Leombruni e Quaranta (2002), pp. 114-116. V'è da aggiungere che questi riscontri vanno letti in chiave descrittiva, senza associarvi alcuna connotazione valutativa. Non vi è infatti alcun motivo – se non un *a priori* ideologico – per attribuire un significato comunque positivo a una maggiore mobilità.

va, ciò potrebbe giustificare l'impiego del metodo dell'abbinamento a tre occasioni per la stima di soddisfacenti misure di mobilità dei lavoratori basate sull'intero campione della RTFL. E, palesemente, tali misure avrebbero due non trascurabili pregi: essere riferite all'insieme degli occupati delle famiglie residenti (dunque, in prima approssimazione, all'occupazione totale interna) ed essere tempestive, con un *lag* ridotto e aggiornabili trimestralmente. La condizione perché ciò valga è evidente: che rispetto al 1995, anno del nostro esercizio, non ci sia stato – e non ci sia – un vistoso aumento della mobilità, in particolare in termini di storie lavorative accidentate, costellate di episodi lavorativi di durata molto breve.

*Dipartimento di Scienze statistiche
Università di Padova*

PAOLO BARETTA
UGO TRIVELLATO

APPENDICE 1: LA STANDARDIZZAZIONE DELLA POPOLAZIONE SULLA BASE DELLA CLASSIFICAZIONE ATECO 91

Per tutte le analisi di mobilità svolte, la classificazione per attività economica degli episodi lavorativi in corso nei periodi in esame avviene prendendo come riferimento la classificazione ATECO 91. Essa suddivide le attività economiche in 17 sezioni (e 16 sottosezioni) e in 60 divisioni, 222 gruppi, 512 classi, 874 categorie, codificate rispettivamente con una (e due) lettere e con due, tre, quattro e cinque cifre. Ne è data, poi, anche una classificazione in 12 rami.

Gli occupati nei *data set* RTFL sono un campione rappresentativo degli occupati delle famiglie residenti in Italia; possono quindi svolgere una qualsiasi attività economica, sia essa dipendente o autonoma. Al contrario, gli scopi previdenziali delle denunce che sono alla base della costruzione del panel INPS comportano l'esclusione da tale insieme di tutti i lavoratori indipendenti e di buona parte dei lavoratori appartenenti ai rami 01 (agricoltura, caccia, silvicoltura e pesca), 10 (pubblica amministrazione, difesa, assicurazione sociale obbligatoria), 11 (istruzione, sanità e altri servizi sociali) e 12 (altri servizi pubblici, sociali e alle persone).

Per effettuare un confronto omogeneo quanto a popolazione di riferimento, si sono eliminati dai *data set* di fonte RTFL tutti i rapporti di lavoro indipendente e dai *data set* di entrambe le fonti tutti i rapporti di lavoro appartenenti alle seguenti divisioni di attività economica (a fianco della codifica a due cifre della divisione vi è la codifica a una lettera della sezione di appartenenza):

- 01 – A agricoltura, caccia e relativi servizi
- 02 – A silvicoltura e utilizzazione di aree forestali e servizi connessi
- 05 – B pesca, piscicoltura e servizi connessi
- 60 – I trasporti terrestri; trasporti mediante condotte
- 64 – I poste e telecomunicazioni
- 71 – K noleggio di macchinari e attrezzature senza operatore e di beni per uso personale e domestico
- 73 – K ricerca e sviluppo
- 75 – L pubblica amministrazione e difesa; assicurazione sociale obbligatoria

- 80 – M istruzione
- 85 – N sanità e altri servizi sociali
- 90 – O smaltimento dei rifiuti solidi, delle acque di scarico e simili
- 92 – O attività ricreative, culturali e sportive
- 95 – P servizi domestici presso famiglie e convivenze
- 99 – Q organizzazioni ed organismi extraterritoriali.

In definitiva, vengono completamente eliminati tutti gli episodi lavorativi afferenti ai rami 01, 10 e 11, mentre vengono parzialmente eliminati quelli relativi ai rami 07 (divisioni 60 e 64), 09 (divisioni 71 e 73) e 12 (tutte, a meno delle divisioni 91: attività di organizzazioni associative non classificate altrove e 93: altre attività dei servizi).

APPENDICE 2: LA PROCEDURA PER LA CORREZIONE DEI FLUSSI DI ASSOCIAZIONI E SEPARAZIONI AGLI ESTREMI DELL'ANNO DI CALENDARIO

Il panel INPS presenta alcuni problemi per la stima delle associazioni nel mese di gennaio e delle separazioni nel mese di dicembre, che mostrano dei picchi in parte anomali. Ciò è dovuto essenzialmente a ritardi nell'aggiornamento del panel, i quali inducono la registrazione di false separazioni (associazioni) relative a rapporti di lavoro ancora (già) esistenti. Ad esempio, se in un dato anno t non risultano pervenuti all'Inps i moduli delle denunce annuali 01/M di un'impresa per i propri dipendenti, presenti invece nel panel nell'anno precedente e in quello successivo, ne risultano apparenti separazioni nel mese di dicembre dell'anno $t-1$ e apparenti associazioni nel mese di gennaio dell'anno $t+1$.

Per ovviare a questo inconveniente, Leombruni e Quaranta (2002), pp. 182-187, propongono una procedura di correzione dei flussi osservati per i mesi estremi dell'anno di calendario, che sfrutta informazioni esogene tratte dalla RTFL. Noi adottiamo tale procedura per le stime della mobilità dei lavoratori da INPS per il 1995.

In sostanza, si tratta di riprodurre per la fonte INPS i *pattern* mensili delle associazioni e delle separazioni riscontrati in RTFL, tramite dei coefficienti che 'estrapolano' all'intero anno rispettivamente le associazioni di febbraio-dicembre e le separazioni di gennaio-novembre. Le ipotesi sottostanti sono, palesemente, che la distribuzione per mese dei flussi rilevati dalla RTFL sia attendibile e che gli errori nei flussi INPS siano concentrati appunto a dicembre (per le separazioni) e a gennaio (per le associazioni).

Denotando con A_m , $m = 1, \dots, 12$, le associazioni e con S_m , $m = 1, \dots, 12$, le separazioni nell'anno 1995, così come desunte dalle risposte ai Quesiti 24 e 38 della RTFL, i coefficienti di riporto ai flussi annuali sono dati rispettivamente da:

$$k_A = \frac{\sum_{m=1}^{12} A_m}{\sum_{m=2}^{12} A_m} \quad \text{e} \quad k_S = \frac{\sum_{m=1}^{12} S_m}{\sum_{m=1}^{11} S_m}$$

e sono pari, nell'ordine, a 1,15 e a 1,19. In altre parole, le associazioni di gennaio sono poste pari al 15% del totale delle associazioni effettuate nel periodo febbraio-dicembre, e le separazioni di dicembre al 19% del totale delle separazioni effettuate nel periodo gennaio-novembre.

Estendiamo poi la procedura anche per il calcolo del tasso di riallocazione, per il quale associazioni e separazioni vanno trattate simmetricamente, e otteniamo un coefficiente di riporto k_R pari a 1,17.

APPENDICE 3: ALCUNI RISULTATI PER GLI ALTRI METODI

La Tavola A1 riassume i principali risultati, largamente insoddisfacenti, per gli altri piani di osservazione: della condizione, della *tenure* pregressa e dell'anno prima, nonché dell'abbinamento a due occasioni.

Per il metodo della *tenure* pregressa e per quello della condizione le evidenze sono, palesemente, negative: effetto congiunto, della composizione e del piano di osservazione, nello spiegare la differenza originaria fra le misure di mobilità dalle due fonti addirittura negativo; effetto marginale del metodo negativo.

Il metodo dell'anno prima (del quale è qui presentata la variante in cui la classificazione per attività economica è disaggregata per sezioni e sottosezioni) fa necessariamente riferimento a una misura di *gross total turnover* del tipo wt_f , dunque più sommaria di t_f , con riguardo alla quale le differenze fra le due fonti si riducono sensibilmente. Peraltro, anche per esso si registrano risultati complessivamente insoddisfacenti: l'effetto congiunto sale a oltre 1,5, mentre l'effetto marginale del metodo resta parecchio lontano da 0, a significare una cospicua perdita di informazioni associata alla simulazione di tale piano di osservazione sui dati INPS.

TAVOLA A1

Sintesi dei risultati per gli altri metodi (misure di mobilità %)

<i>Metodo della 'tenure' pregressa, s_0 vs. u_5</i>		<i>Metodo della condizione, u_5</i>	
INPS: 19,8 ^a ; 26,1 ^b	RTFL: 9,0 ^a ; 11,1 ^b	INPS: 31,5 ^a ; 37,4 ^b	RTFL: 17,6 ^a ; 22,0 ^b
10,8	-0,725	13,9	-0,420
0,249	-0,388	0,319	-0,107
<i>Metodo dell'anno prima, u_5</i>		<i>Metodo dell'abbinam. a due occasioni^d, t_f</i>	
INPS: 39,3 ^a ; 35,4 ^b	RTFL: 34,7 ^a ; 38,0 ^b	INPS: 52,4 ^a ; 43,7 ^b	RTFL: 26,3 ^a ; 36,7 ^b
4,6	0,227	26,1	0,308
0,826	1,571	0,450	0,732

^a Misura di mobilità calcolata con il metodo del panel (per INPS) e con il metodo in esame (per RTFL), per il totale degli individui.

^b Misura di mobilità calcolata con il metodo in esame per la popolazione standardizzata.

^c Risultati per la variante in cui la classificazione per attività economica è disaggregata per sezioni e sottosezioni. Per RTFL, il periodo annuale è riferito ad aprile 95-aprile 96; per INPS, il periodo annuale è riferito all'anno solare 1995.

^d Abbinamento 95.I-96.I.

Quanto al metodo dell'abbinamento a due occasioni d'indagine a distanza di un anno, valutato con riguardo a t_f fornisce risultati di una qualche plausibilità, ma ancora inadeguati. L'effetto congiunto si colloca attorno a 0,7 ed è affiancato da effetti marginali della composizione e del metodo pari rispettivamente a 0,45 ed a 0,3. La distanza rispetto ai paralleli risultati del metodo dell'abbinamento a tre occasioni (vedi Tavola 6) è netta. Ciò conferma l'importanza di disporre di una rilevazione intermedia entro il periodo di

riferimento di un anno, per ottenere dalla RTFL misure di mobilità dei lavoratori ragionevolmente prossime a quelle fornite dall'INPS.

RIFERIMENTI BIBLIOGRAFICI

- J.M. ABOW., P. CORBEL, F. KRAMARZ, (1999), *The entry and exit of workers and the growth of employment: an analysis of French establishments*, "The Review of Economics and Statistics", 38 (2), pp. 115-133.
- J.M. ABOWD, F. KRAMARZ, (1999), *The analysis of labor markets using matched employer-employee data*, in O. ASHENFERTER, D. CARD (eds.), *Handbook of Labor Economics, Vol. 3B*, North-Holland, Amsterdam, pp. 2629-2710.
- A. ACCORNERO (a cura di), (2000), *Solo una grande giostra? La diffusione del lavoro a tempo determinato*, Franco Angeli, Milano.
- F. ALTISSIMO, D.J. MARCHETTI, G.P. ONETO, (2000), *The Italian business cycle: coincident and leading indicators and some stylized facts*, Temi di discussione n. 377, Banca d'Italia, Roma.
- B. ANASTASIA, M. GAMBENZA, M. RASERA, (2000), *La mobilità del mercato del lavoro e l'incidenza dei contratti a tempo determinato*, in A. ACCORNERO (a cura di), *cit.*, pp. 109-137.
- P. BARETTA, (2002), *Analisi comparata delle misure di mobilità dei lavoratori con metodi di stima per fonti amministrative e indagini campionarie in Italia*, Tesi di dottorato di ricerca in Statistica Applicata alle Scienze Economiche e Sociali, Dipartimento di Scienze Statistiche, Università degli Studi di Padova (mimeo).
- J. BOUND, C. BROWN, N. MATHIOWETZ, (2001), *Measurement error in survey data*, in J.J. HECKMAN, D. LEAMER (eds.), *Handbook of Econometrics. Volume 5*, North Holland, Amsterdam, pp. 3705-3843.
- B. CONTINI (a cura di), (2002), *Osservatorio sulla mobilità del lavoro in Italia*, Il Mulino, Bologna.
- B. CONTINI, A. GAVOSTO, R. REVELLI, P. SESTITO, (1992), *Creazione e distruzione di posti di lavoro in Italia*, Temi di discussione n. 177, Banca d'Italia, Roma.
- B. CONTINI, R. REVELLI, (1987), *The process of job creation and job destruction in the Italian economy*, "Labour", 1 (3), pp. 121-144.
- B. CONTINI, R. REVELLI, (1992), *Imprese, occupazione e retribuzioni al microscopio*, Il Mulino, Bologna.
- S.J. DAVIS, J. HALTIWANGER, (1998), *Measuring gross worker and job flows*, in J. HALTIWANGER, M.E. MANSER, R. TOPEL (eds.), *Labor statistics. Measurement issues*, The University of Chicago Press, Chicago, pp. 77-119.
- S.J. DAVIS, J. HALTIWANGER, (1999), *Gross job flows*, in O. ASHENFERTER, D. CARD (eds.), *Handbook of Labor Economics, Vol. 3B*, North Holland, Amsterdam, pp. 2711-2805.
- S.J. DAVIS, J. HALTIWANGER, S. SCHUH, (1996), *Job Creation and Destruction*, The MIT Press, Cambridge (MA).
- A. DE ANGELINI, A. GIRALDO, (2003), *La mobilità dei lavoratori nel Veneto. Confronto fra misure su dati RTFL e su dati Netlabor*, Progetto MIUR 'Dinamiche e persistenze nel mercato del la-

- avoro italiano ed effetti di politiche', Working paper n. 61, Dipartimento di Scienze Statistiche, Università di Padova.
- B. EFRON, R. TIBSHIRANI, (1986), *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy*, "Statistical Science", 1 (1), pp. 54-77.
- M. FILIPPI, L. PACELLI, C. VILLOSIO, (2002), *Le basi dati di fonte Inps*, in B. CONTINI (a cura di), *cit.*, pp. 381-400.
- P. GENNARI, P. SESTITO, (1996), *Durata dell'attività lavorativa e turn-over in Italia*, paper presentato all'11^o Convegno nazionale dell'AIEL, Napoli, 3-5 ottobre 1996 (mimeo.).
- D.S. HAMERMESH, W.H.J. HASSINK, J.C. VAN OURS, (1996), *Job turnover and labor turnover: a taxonomy of employment dynamics*, "Annales d'Economie et de Statistique", No. 41/42, Part 1, pp. 21-40.
- ISTAT, (2002), *Le matrici di transizione della rilevazione trimestrale sulle forze di lavoro. Nota metodologica*, Istat, Roma (scaricabile da: <http://www.istat.it/Lavoro/Storico/index.htm>).
- R. LEOMBRUNI, C. MALPEDE, (2002), *Aspetti istituzionali del mercato del lavoro italiano*, in B. CONTINI (a cura di), *cit.*, pp. 371-379.
- R. LEOMBRUNI, R. QUARANTA, (2001), *Stime della mobilità del lavoro da dati Inps 'mimando' l'indagine trimestrale forze di lavoro*, paper presentato allo Workshop "Occupazione e disoccupazione in Italia: misura e analisi dei comportamenti", Bressanone 15-16 gennaio 2001 (mimeo.).
- R. LEOMBRUNI, R. QUARANTA, (2002), *Mobilità dei lavoratori in Italia, 1985-1996: andamento ciclico e cambiamenti strutturali*, in B. CONTINI (a cura di), *cit.*, pp. 111-203.
- L. PACELLI, C. VILLOSIO, (2002), *Trasformazioni societarie: definizione, rilevazione e calcolo dell'impatto sulle statistiche di mobilità delle imprese e dei lavoratori*, in B. CONTINI (a cura di), *cit.*, pp. 401-411.
- A. PAGGIARO, N. TORELLI, (1999), *Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro*, Progetto MURST 'Lavoro e disoccupazione: questioni di misura e di analisi', Working paper n. 15, Dipartimento di Scienze Statistiche, Università di Padova.
- G. TATTARA, M. VALENTINI, (2002), *Who is first to sit down, who is left standing when the music stops?*, paper presented at the EALE 14th Annual Conference, Paris, 19-22 September 2002 (mimeo.).
- G. TATTARA, M. VALENTINI, (2003), *Un mercato del lavoro molto mobile? Sì, ma con molti se e molti ma*, "Economia e Società Regionale", 81, pp. 24-54.
- U. TRIVELLATO, (1997), *Le misure della partecipazione al lavoro nel quadro comunitario*, in L. FREY (a cura di), *Le informazioni sul lavoro in Italia: significato e limiti delle informazioni provenienti da indagini sulle famiglie*, Quaderni di Economia del Lavoro n. 59, Franco Angeli, Milano, pp. 9-34.
- U. TRIVELLATO (a cura di), (2001), *Servizi per l'impiego e ricerche sul lavoro. L'esperienza del Veneto* (a cura di), Quaderni di Economia del Lavoro n. 72, Franco Angeli, Milano.
- C. ZANZOTTERA, (2003), *La posizione delle donne in Lombardia fra partecipazione e inattività*, in M. SAMEK LODOVICI, R. SEMENZA (a cura di), *Partecipazione femminile e ruolo dell'età. La posizione lavorativa delle donne adulte in Lombardia*, Istituto per la Ricerca Sociale, Milano, pp. 49-76.

RIASSUNTO

La mobilità dei lavoratori da fonti amministrative e da 'surveys' sulle famiglie: un'analisi comparata

In Italia gli studi sulla mobilità dei lavoratori utilizzano prevalentemente gli archivi dell'Inps. Si esamina la possibilità di utilizzare per questo scopo la Rilevazione Trimestrale delle Forze di Lavoro (RTFL). L'interesse nell'utilizzo della RTFL è duplice: le analisi pos-

sono essere riferite a pressoché tutta l'occupazione (invece che ai soli lavoratori dipendenti nel settore privato extra-agricolo); inoltre possono essere svolte tempestivamente (invece che con un *lag* dell'ordine di 4-5 anni). All'utilizzo della RTFL si frappone, tuttavia, un serio ostacolo: il piano di osservazione risultante dal questionario documenta in modo assai parziale le storie lavorative. A fronte di ciò, in primo luogo si identificano e si controllano le principali ragioni di divergenza fra le misure di mobilità fornite dalle due fonti – diversa popolazione di riferimento e diverso piano di osservazione –. In secondo luogo, si propone un metodo detto 'dell'abbinamento a tre occasioni', che utilizza dati abbinati di tre rilevazioni della RTFL nell'arco dell'anno, e se ne valutano le *performances*. L'analisi empirica è condotta intorno all'anno 1995. L'evidenza segnala la capacità del metodo, e quindi dei corrispondenti *data set* longitudinali della RTFL, di cogliere larga parte della mobilità dei lavoratori documentata dall'Inps: pressoché completamente per quanto attiene a misure di riallocazione; con una sottostima contenuta per quanto riguarda misure di *gross worker turnover*.

SUMMARY

Worker mobility from social security registers and household surveys: a comparative analysis

In Italy most of the research on worker and job mobility is carried out on social security (INPS) administrative databases. We investigate the possibility of using the Italian Quarterly Labour Force Survey (RTFL) as an additional source for currently measuring worker mobility. In principle, the motivation for using the RTFL is twofold: it covers (almost) the whole employment, while INPS data are restricted to dependent employment in the private non-agricultural sector; it is timely, while INPS data are available with a lag of 4-5 years. The RTFL suffers, however, from a serious hindrance: the observation plan of the questionnaire provides fragmentary information on work histories. First, we identify and, to a large extent, control for the main reasons of discrepancy of worker mobility measures from the two sources, *i.e.*, differences in the reference population and in the observation plan. Second, we propose a method capable of largely reconciling the evidence from the two sources. The 'three-waves linkage' method – as we call it – exploits the rotating panel design of the survey and uses linked data over three waves of the RTFL within a calendar year. Its salient features are presented and some main results are discussed. The empirical analysis is carried out for the year 1995. The evidence suggests that the three-waves linkage method, and thus the corresponding longitudinal datasets from RTFL, captures large part of the worker mobility documented by INPS databases: entirely as worker reallocation measures are concerned; with a significant, but moderate underestimation with respect to gross worker turnover.