

Research Institute for the Evaluation of Public Policies



Un'esperienza pilota di integrazione di dati amministrativi e di survey per l'analisi dell'evoluzione delle storie lavorative

Martina Bazzoli

Sonia Marzadro

Antonio Schizzerotto

Ugo Trivellato

FEBRUARY 2018

FBK-IRVAPP WORKING PAPER No. 2018-01

Un'esperienza pilota di integrazione di dati amministrativi e di survey per l'analisi dell'evoluzione delle storie lavorative dei giovani

Martina Bazzoli

FBK-IRVAPP

Sonia Marzadro

FBK-IRVAPP

Antonio Schizzerotto

FBK-IRVAPP & University of Trento

Ugo Trivellato

FBK-IRVAPP, Università di Padova, ZEW e CES-Ifo

FBK-IRVAPP Working Paper No. 2018-01

February 2018



Research Institute for the Evaluation of Public Policies

Bruno Kessler Foundation

Vicolo dalla Piccola 2, 38122 Trento (Italy)

Phone: (+39) 0461.314209

Fax: (+39) 0461.314240

E-mail: info@irvapp.it

Website: <http://irvapp.fbk.eu>

The purpose of the IRVAPP Working Papers series is to promote the circulation of working papers prepared within the Institute or presented in IRVAPP seminars by outside researchers with the aim of stimulating comments and suggestions. Updated review of the papers are available in the Reprint Series, if published, or directly at the IRVAPP.

The views expressed in the articles are those of the authors and do not involve the responsibility of the Institute.

Un'esperienza pilota di integrazione di dati amministrativi e di survey per l'analisi dell'evoluzione delle storie lavorative dei giovani §

Martina Bazzoli^a, Sonia Marzadro^a, Antonio Schizzerotto^b e Ugo Trivellato^c

^a FBK-IRVAPP

^b FBK-IRVAPP, Università di Trento

^c FBK-IRVAPP, Università di Padova, ZEW e CES-Ifo

Abstract

Nell'ultimo decennio si è venuta affermando la pratica di abbinare a indagini longitudinali dati amministrativi permettendo di superare, tramite l'integrazione, i limiti che usualmente affliggono queste due fonti di dati. L'articolo si colloca entro questa prospettiva; esso descrive un'esperienza pilota di *matching* deterministico e di integrazione tra due fonti di dati – l'indagine panel Condizioni di vita delle famiglie trentine e gli archivi amministrativi dell'Inps regionale – messa a punto al fine di ricostruire l'evoluzione dei tratti salienti della fase iniziale delle storie lavorative di tre coorti di giovani trentini in un arco temporale che si estende, complessivamente, dal 1974 al 2014. Pur essendo uno studio condotto su piccola scala – la provincia di Trento – esso presenta l'indubbio interesse di essere una delle rare esperienze italiane di integrazione di basi di dati amministrativi e di survey.

Keywords: Integrazioni di fonti, archivi amministrativi, dati di survey, mercato del lavoro

§ Ricerca svolta nell'ambito del Protocollo d'intesa per attività di ricerca tra FBK-Irvapp, Direzione Regionale Trentino-Alto Adige dell'Inps e Ispat (n. 157/176704109). Una versione preliminare è stata presentata al Convegno nazionale "L'integrazione tra archivi amministrativi e indagini campionarie come strumento per le analisi socio-economiche e la valutazione di politiche pubbliche" (FBK-Irvapp, Trento, 17 ottobre 2016). Siamo grati a Marco Zanotelli e a Giovanna Fambri per il supporto fornito rispettivamente da Inps e Ispat nell'acquisizione e integrazione dei dati, e a Bruno Anastasia, Cinzia Giannelli, Gianluca Mazzarella e Michele Mosca per commenti e suggerimenti.

Un'esperienza pilota di integrazione di dati amministrativi e di survey per l'analisi dell'evoluzione delle storie lavorative dei giovani

1. Le ragioni di un'esperienza pilota

I giovani di oggi stanno peggio di quelli di ieri? L'interrogativo si pone ormai spesso nella stampa, nel dibattito pubblico, nella ricerca economica e sociale. Rispondervi in maniera argomentata è, peraltro, difficile. Lo "stare peggio" – o meglio o allo stesso modo – comporta infatti confronti fra generazioni. Tali confronti richiedono di affrontare complesse questioni di metodo ma, ancor prima, necessitano di dati longitudinali, che seguano più coorti di individui – i giovani di ieri e quelli di oggi – per l'intero corso di vita o comunque per un suo lungo segmento.

Tipicamente, in parecchi paesi avanzati tali dati sono rilevati tramite *multi-purpose household panel surveys*, di elevata numerosità – che consentano quindi adeguate disaggregazioni, innanzitutto per coorte – e condotte con attenzione a un appropriato disegno dell'indagine. Esempio, in proposito, è *Understanding Society*, lo studio longitudinale del Regno Unito avviato nel 2009 con onde annuali sugli individui di circa 40.000 famiglie – inclusi quelli del campione di circa 8.000 famiglie della precedente *British Household Panel Survey*, iniziata nel 1991¹.

In paesi nei quali la produzione dell'informazione statistica poggia in larga parte su registri di origine amministrativa – tipicamente quelli scandinavi –, l'alternativa è costituita da basi di dati longitudinali per aree tematiche approntate a partire dai registri. Esse presentano notevoli vantaggi, tra i quali il fatto di riferirsi all'intera popolazione di interesse e di essere ad alta frequenza, quindi con informazioni dettagliate sul succedersi di eventi ed episodi. L'esperienza danese in tema di lavoro è paradigmatica: *Statistics Denmark* ha realizzato, e aggiorna annualmente, *IDA (Integrated Database for Labour Market Research)*, un base di dati longitudinali sulla popolazione di lavoratori e imprese a partire dal 1980.

Nell'ultimo decennio, poi, soprattutto in paesi di lingua inglese si è venuta affermando la pratica di abbinare a indagini longitudinali dati amministrativi, così come la ricerca sulle questioni di metodo che ne sono coinvolte (Calderwood e Lessof, 2009). L'esempio è ancora fornito da *Understanding Society*, nella quale, previo consenso dei rispondenti adulti, ai dati dell'indagine sono collegati i pertinenti dati amministrativi su salute e istruzione.

Sfortunatamente in Italia basi di dati longitudinali che seguano per un tempo sufficientemente lungo appropriati insiemi di individui, alimentate da *household panel surveys*² o realizzate a partire da fonti amministrative³, non sono correntemente disponibili alla scala nazionale. Né tornano utili

¹ Vedi <https://www.understandingsociety.ac.uk/>. Tra altre indagini longitudinali sulle famiglie, meritano di essere segnalate il *German Socio-Economic Panel (SOEP)*, condotto dal 1994 (https://www.diw.de/en/diw_02.c.221178.en/about_soep.html), e lo statunitense *Panel Study of Income Dynamics (PSID)*, iniziato nel 1968 (<http://psidonline.isr.umich.edu/>).

² Esistono sì indagini nazionali con una componente longitudinale: tra queste due dell'Istat, *IT-SILC* (il segmento italiano della *EU-Survey on Income and Living Conditions*) e l'Indagine sulle forze di lavoro, e una della Banca d'Italia, l'Indagine sui bilanci delle famiglie italiane; la componente longitudinale ha però un ristretto periodo di osservazione. Due *multi-purpose household panel surveys* sono state condotte nel decennio a cavallo del 2000: la *European Community Household Panel* (8 onde annuali, tra il 1994 e il 2001; vedi <http://ec.europa.eu/eurostat/web/microdata/european-community-household-panel>), che ha coinvolto anche l'Italia, e l'Indagine longitudinale sulle famiglie italiane (5 onde biennali, dal 1997 al 2005; vedi <http://www.soc.unitn.it/ilfi/>).

³ Le basi di dati di origine amministrativa, infatti, sono ristrette a segmenti della popolazione di interesse. Le esperienze più significative in tema di lavoro riguardano basi di dati longitudinali costruite a partire dagli

i notevoli avanzamenti dell'Istat sul fronte dell'integrazione di microdati da fonti amministrative e da surveys; essi si incentrano, infatti, sull'area delle imprese e per il tema affrontato sono circoscritti all'integrazione dei dati di IT-Silc sui redditi da lavoro con fonti fiscali e dell'Inps (Sestito e Trivellato, 2011; Istat, 2009)⁴. In generale, poi, sulle possibilità di procedere con abbinamenti e integrazioni⁵ pesa la normativa immotivatamente restrittiva sulla tutela della privacy per i trattamenti di dati personali per scopi statistici o di ricerca scientifica, in particolare per basi di microdati di origine amministrativa (Trivellato 2017).

A fronte di questo quadro, FBK-IRVAPP con la cruciale collaborazione della Direzione Regionale Trentino-Alto Adige dell'Inps e dell'Istituto di Statistica della provincia di Trento (Ispat), ha quindi avviato un progetto sulle storie lavorative di un campione di giovani trentini, che poggia sull'integrazione di microdati tratti da due fonti, l'indagine panel *Condizioni di vita delle famiglie trentine* (nel seguito *Panel*) e gli archivi amministrativi dell'Inps regionale (nel seguito *Inps*). Esso è servito a documentare tratti salienti dell'evoluzione delle storie lavorative di tre coorti di giovani trentini in un arco temporale che si estende, complessivamente, dal 1974 al 2014⁶.

Lo studio è stato condotto, dunque, su piccola scala – la provincia di Trento – e si è misurato con questioni di ricerca empirica che non contano su una solida letteratura. A nostro avviso, lo studio presenta l'indubbio interesse di essere una delle rare esperienze italiane di integrazione di basi di dati amministrativi e di survey.

Il lavoro prosegue come segue. Il paragrafo 2 dà conto delle due fonti utilizzate, il paragrafo 3 delinea i passi compiuti per la loro integrazione mentre il paragrafo 4 offre un primo vaglio di massima della ragionevolezza della base di dati integrata confrontando le stime sezionali per il 2012 con quelle derivate dalla rilevazione Istat sulle forze di lavoro. Al paragrafo 5 sono dedicate le conclusioni.

2. La descrizione delle fonti

2.1 Le caratteristiche dell'indagine Panel

Il campione di riferimento per l'integrazione dei dati consta di circa 5.730 individui maggiorenni intervistati nel corso dell'onda 2012 del Panel⁷.

Il Panel è una indagine longitudinale multi-purpose, con prima onda retrospettiva, iniziata nel 2005/06 e articolata in rilevazioni biennali condotte su un campione rappresentativo delle convivenze domestiche presenti in provincia di Trento⁸. In ciascuna ondata sono intervistati tutti i componenti maggiorenni delle famiglie campionate, per quanto qui di interesse in merito agli episodi (di lavoro e non lavoro) che ne hanno scandito la storia lavorativa dall'ingresso nelle forze di lavoro alla data dell'intervista. Per ciascuno di essi il Panel rileva l'anno e il mese di inizio e di fine, nonché le sue caratteristiche principali. Nelle onde che, per ciascun intervistato, sono

archivi gestionali dell'Inps. Tra queste spicca WHIP (*Work Histories Italian Panel*), la cui sezione riguardante il lavoro dipendente extra-agricolo è un *linked employer-employee longitudinal database*. WHIP è stato ampiamente utilizzato per studi sulla mobilità del mercato del lavoro italiano. Il suo aggiornamento, peraltro, è fermo al 2004 (vedi <http://www.laboratoriorevelli.it/it/whip>).

⁴ Un abbinamento di microdati sulla partecipazione al lavoro di IT-SILC, rilevazione del 2005, con microdati longitudinali da archivi dell'Inps è poi alla base del modello di microsimulazione dinamico T-DYMM del Ministero dell'Economia e delle Finanze (Caretta *et al.*, 2012).

⁵ Importa evidenziare la differenza fra "abbinamento" e "integrazione" di microdati da fonti diverse. L'abbinamento attiene al solo *linkage*, in via di principio realizzato tramite *matching* esatto utilizzando come chiave un "numero di identificazione personale" – in Italia il codice fiscale. L'integrazione si estende all'armonizzazione di definizioni e classificazioni e alla combinazione dei dati dalle diverse fonti – tramite criteri di priorità e/o procedure di riconciliazione –.

⁶ I risultati sostantivi sono rinvenibili in Bazzoli *et al.* (2018).

⁷ Rimangono esclusi i minori e i casi (23) con codice fiscale mancante.

⁸ L'indagine è condotta dall'Ispat, con la collaborazione di FBK-Irvapp. Vedi Fambri e Schizzerotto (2008).

successive alla prima, la rilevazione della storia lavorativa avviene mediante *dependent interviewing* (cioè cominciando col chiedere se la condizione – di lavoro o non lavoro – sia rimasta inalterata rispetto a quella dichiarata nell'onda precedente). In presenza di più lavori svolti contemporaneamente l'intervistato era, inoltre, invitato a riportare le caratteristiche dell'occupazione ritenuta principale.

La storia lavorativa di un intervistato è rappresentata da una sequenza di episodi non sovrapposti – di lavoro, disoccupazione e inattività – di durata almeno mensile nell'arco di tempo compreso fra la data del primo ingresso nel mercato del lavoro e la data dell'intervista.

È opportuno precisare che il lavoro rilevato riguarda ogni attività extra-domestica remunerata che rappresenti, in quanto fonte di reddito, uno strumento importante per ottenere le risorse attraverso le quali l'intervistato assicura – o contribuisce ad assicurare – il mantenimento personale o della sua famiglia. Il concetto di lavoro qui utilizzato non rinvia, quindi, solo ad attività e occupazioni continuative e regolari, ma comprende anche attività più o meno occasionali e discontinue, sotto il profilo temporale, e più o meno informali e irregolari, sotto il profilo della stabilità del posto e della relazione di impiego, purché, appunto, da esse dipendano, in misura apprezzabile, le condizioni di vita dell'intervistato e della sua eventuale famiglia. Non sono invece considerate piccole attività ("lavoretti") svolte dagli intervistati, le cui normali possibilità di vita sono assicurate da uno o più familiari, per il solo scopo di soddisfare propri modesti bisogni voluttuari.

2.2 La descrizione della base di dati Inps

Per il campione di riferimento tratto dal Panel si sono poi acquisite le informazioni individuali da Inps, in particolare dal Cassetto previdenziale del cittadino, che contiene l'elenco delle contribuzioni registrate a favore del lavoratore dall'apertura della sua posizione assicurativa⁹.

Gli episodi contributivi contenuti nell'archivio afferiscono alle seguenti gestioni: 1) cassa regolare; 2) fondo clero; 3) fondo sport e spettacolo; 4) gestione separata. Ne restano esclusi gli episodi di lavoro come dipendente pubblico, come professionista iscritto ad ordini con cassa previdenziale propria e, ovviamente, come occupato nell'economia informale ("lavoro nero").

Per ciascun episodio contributivo Inps rileva la data di inizio e di fine e, nella gestione regolare, il tipo di contribuzione e la retribuzione percepita a fini previdenziali¹⁰. Si precisa che per tutti gli episodi appartenenti alla gestione separata l'inizio e la fine sono convenzionalmente pari rispettivamente al 1° gennaio e al 31 dicembre.

Per la gran parte dei dipendenti, ma solo a partire dal 1974, è presente anche la matricola Inps identificativa dell'impresa presso la quale lavorano. Essa ha permesso di unire, a ciascun episodio contributivo il codice fiscale dell'impresa, il suo codice di attività economica, il suo indirizzo, la sua data di costituzione e di (eventuale) cessazione¹¹.

L'archivio contiene inoltre le contribuzioni dette figurative¹²: la cassa integrazione guadagni, i contratti di solidarietà, le indennità di mobilità o di disoccupazione.

2.3 Il trattamento della base di dati Inps

⁹ Per ciascun soggetto contribuente il Cassetto previdenziale contiene le informazioni alle quali lo stesso può accedere collegandosi alla propria pagina personale nella sezione Servizi Online sul sito dell'Inps. Oltre alle contribuzioni versate dai datori di lavoro nel caso di lavoratori dipendenti o dai soggetti interessati nel caso di lavoratori autonomi, vi sono anche le contribuzioni figurative, associate alla cassa integrazione guadagni, ai contratti di solidarietà e alle indennità di disoccupazione e di mobilità.

¹⁰ Nell'archivio amministrativo sono, tra l'altro, disponibili il numero di settimane utilizzabili ai fini della determinazione della consistenza della prestazione previdenziale (settimane utili al calcolo) e dell'anzianità contributiva (settimane utili al diritto).

¹¹ Questi dati provengono dai flussi *E-Mens* relativi alle imprese.

¹² Sono, inoltre, presenti altri tipi di contribuzione, quali il riscatto degli anni di università o del servizio militare. Di essi non si è però tenuto conto.

Per la ricostruzione delle storie lavorative si è proceduto nel seguente modo. Il tipo di contribuzione è servito per individuare gli episodi lavorativi e, in particolare, per classificarli nelle seguenti condizioni occupazionali: “Dipendenti del settore privato extra-agricolo”, “Autonomi dell’industria e del terziario (artigiani, commercianti, imprenditori)”, “Liberi professionisti”, “Collaboratori”, “Autonomi e dipendenti dell’agricoltura”. Oltre ai periodi trascorsi in occupazione si è poi tenuto conto dei periodi di non lavoro. In particolare, entro i periodi di contribuzione figurativa sono stati individuati i segmenti temporali di disoccupazione sussidiata in senso stretto (indennità di disoccupazione ordinaria, indennità di mobilità, ASpl e MiniASpl)¹³.

Per disporre di dati statistici idonei a rappresentare le storie lavorative, si è poi resa necessaria una minuziosa attività di affinamento dell’informazione amministrativa¹⁴ (Figura 1) che ha interessato:

- i) gli episodi privi di settimane contributive;
- ii) gli episodi multipli;
- iii) gli episodi segmentati;
- iv) i casi di mobilità spuria.

La prima scelta è stata quella di ignorare gli episodi, sia di lavoro sia di non lavoro, a cui non era associata alcuna settimana contributiva.

Successivamente sono stati identificati i *record multipli*: tale operazione ha riguardato gli episodi di lavoro dipendente extra-agricolo recanti l’identificativo dell’impresa¹⁵, le occupazioni autonome (commercianti, artigiani, imprenditori), quelle dell’agricoltura (autonomi e dipendenti) e gli episodi di disoccupazione sussidiata in senso stretto. Nel caso di coincidenza tra date di inizio e fine e, per gli episodi di lavoro, tipo di contribuzione e identificativo dell’impresa, si è mantenuto un solo episodio.

Il terzo tipo di intervento ha riguardato gli *episodi di lavoro segmentati*, cioè a dire i casi in cui un episodio lavorativo intrattenuto con la medesima impresa fosse registrato come se si trattasse di una pluralità di rapporti distinti con date di inizio e fine diverse ma consecutive¹⁶. In questo caso, i record sono stati riuniti in unico episodio lavorativo.

Infine, i dati relativi alle imprese sono stati utilizzati per identificare eventuali casi di *mobilità spuria*, ovvero situazioni nelle quali ad un soggetto corrispondono più episodi di lavoro alle dipendenze di datori di lavoro solo formalmente diversi¹⁷. È questo il caso di rapporti di impiego avuti con un’azienda che è stata interessata da trasformazioni societarie (e, dunque, ha visto modificato il numero di matricola), senza che vi siano stati reali discontinuità nei beni o servizi da essa prodotti e nella sua organizzazione complessiva¹⁸. Se si ignorasse la peculiarità di questi flussi, dal lato

¹³ La disoccupazione è composta dalla disoccupazione sussidiata in senso stretto (indennità di disoccupazione ordinaria, indennità di mobilità, ASpl e MiniASpl) e da “altra non occupazione sussidiata” (disoccupazione a requisiti ridotti, agricola e edile). Per queste ultime indennità sono riportate convenzionalmente periodi annuali, con inizio nel mese di gennaio e fine nel mese di dicembre. Ne viene che la loro durata, ottenuta come differenza tra la data di fine e la data di inizio dell’episodio, sovrastima, per costruzione, l’eventuale “vera” durata della disoccupazione. Per di più il godimento del trasferimento monetario è compatibile anche con la condizione di inattivo. Per tali ragioni si è scelto di non includere nell’archivio integrato queste altre condizioni di non occupazione sussidiata.

¹⁴ Per norme di legge, uno stesso episodio, sia di lavoro sia di disoccupazione sussidiata, può comparire negli archivi in più di una registrazione.

¹⁵ Nessun intervento è stato effettuato in vista della costruzione dell’archivio integrato *Inps-Panel* per gli episodi di lavoro dipendente extra-agricolo privi di matricola aziendale.

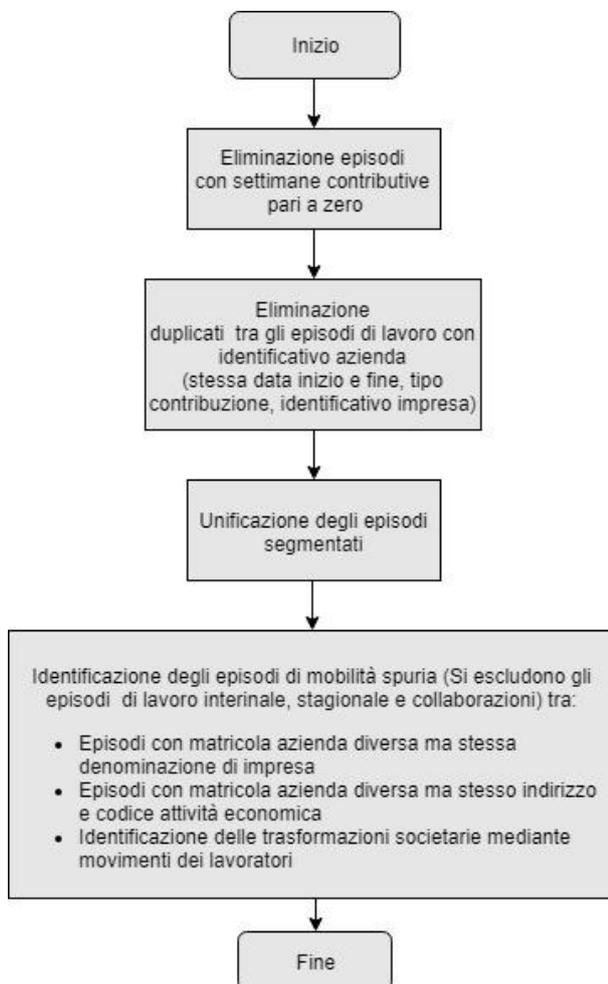
¹⁶ Non necessariamente da gennaio a dicembre.

¹⁷ Va da sé che l’individuazione dei rapporti di mobilità spuria si è potuta effettuare solo per gli episodi di lavoro dipendente dotati di identificativo dell’impresa. Rimangono, quindi, esclusi oltre agli episodi antecedenti il 1975, anche tutti gli episodi che rientrano nella gestione separata, in quella dello sport e spettacolo e in quella del clero (oltre che, banalmente, quelli di lavoro autonomo). Data la loro peculiarità, sono stati inoltre esclusi gli episodi di lavoro interinale di lavoro stagionale abituale.

¹⁸ Come noto, le imprese possono subire dei cambiamenti nel corso del tempo: nei loro caratteri identificativi (la ragione sociale o l’indirizzo), nelle loro caratteristiche strutturali (la dimensione o l’attività prevalente), o nei casi di fusioni, scorpori o cessazioni. In molti casi, queste modificazioni non derivano da cambiamenti

delle imprese si osserverebbero distruzioni e creazioni fittizie di posti di lavoro, mentre dal lato dei lavoratori cambi di lavoro che tali non sono. Per lo studio della mobilità occupazionale, diventa, dunque, rilevante discriminare fra movimenti derivanti da scelte dei lavoratori e movimenti spuri indotti, cioè, da cambiamenti che riguardano le imprese.

Figura 1 Diagramma di flusso contenente le procedure di affinamento della base di dati Inps



L'individuazione di questi *falsi* flussi è un'operazione piuttosto complessa. Al fine di identificare queste situazioni sono stati considerati i passaggi diretti da un'azienda ad un'altra (i cosiddetti *job-to-job*) avvenuti nell'arco di 30 giorni¹⁹. Successivamente, per stabilire quali di essi fossero mobilità spuria sono stati utilizzati due criteri così definiti²⁰:

reali ma dall'adequarsi dei soggetti legali a leggi e a regole amministrative. Ciononostante, tutti questi eventi, anche quelli che, di fatto, non generano delle reali discontinuità nella vita dell'impresa, sono registrati negli archivi amministrativi definendo flussi cosiddetti *spuri*.

¹⁹ Si è quindi giunti a costruire una base dati *ad hoc* del tipo *employee/employers* contenente, per ciascun individuo con almeno un *job-to-job* entro i 30 giorni (5.801), le caratteristiche dell'azienda di partenza e quelle dell'azienda di arrivo.

²⁰ Si fa riferimento a procedure elaborate per la costruzione delle basi dati WHIP: vedi Contini B. e U. Trivellato (2005), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, Bologna, Il Mulino. Si ringrazia, inoltre, il dott. Michele Mosca per il supporto in questa fase del progetto.

- (i) la stabilità delle caratteristiche dell'impresa (si stabilisce l'identità tra due aziende con matricola diversa sulla base dell'identità della loro denominazione²¹ e/o dell'ubicazione della loro sede²²);
- (ii) e (ii) l'identificazione delle trasformazioni societarie formali mediante il riferimento al flusso di dipendenti della nuova impresa e quelli della precedente azienda.

I casi identificati tramite questi criteri sono poi stati controllati e validati manualmente uno ad uno²³ (per dettagli si rimanda all'Appendice).

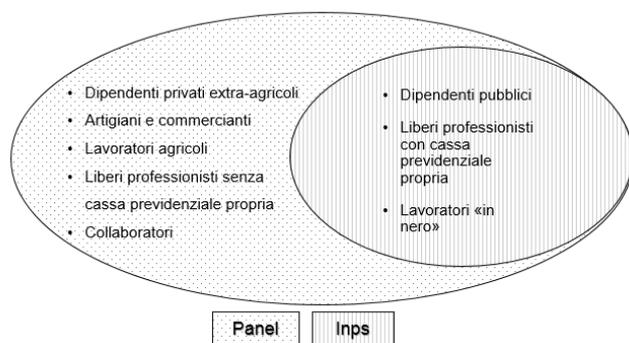
3. La procedura di integrazione delle fonti

È opportuno ribadire che tra le due fonti vi sono alcune importanti differenze quanto a copertura dell'*insieme delle condizioni occupazionali e di disoccupazione* (detto con precisione, degli *episodi* di occupazione e di disoccupazione).

La prima discrepanza riguarda la *condizione occupazionale* registrata. Il *Panel*, infatti, si propone di rilevare tutte le condizioni occupazionali, mentre in *Inps* mancano le condizioni di dipendente pubblico, di libero professionista con cassa previdenziale propria e di "lavoratore in nero"²⁴ (Figura 2).

La seconda discrepanza riguarda la *condizione di disoccupazione*. Come già detto, dell'insieme degli episodi di disoccupazione amministrativa registrati in *Inps* ai nostri fini abbiamo incluso solo gli episodi di disoccupazione sussidiata in senso stretto (vedi la nota 13). In *Panel*, invece, sono classificati come disoccupati tutti coloro che si dichiarano tali, rispondendo a una sintetica domanda (Figura 3). L'indagine rileva quindi sia coloro che sono alla ricerca del primo impiego sia quanti cercano un lavoro pur non beneficiando di alcuna indennità.

Figura 2 - Occupazioni presenti nelle due fonti



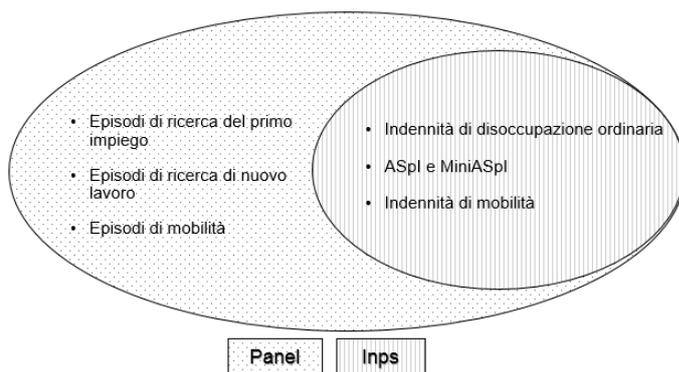
²¹ È questo il caso di imprese con più unità operative.

²² Si è fatto ricorso alla geocodifica degli indirizzi via google map.

²³ L'identificazione dei passaggi aziendali spuri ha consentito di correggere 951 episodi lavorativi, corrispondenti al 3,8% del totale degli episodi di lavoro dipendente.

²⁴ È possibile che ci sia anche una sottostima di co.co.co e co.co.occ dal momento che per i primi fino al gennaio 1996 non vi era obbligo di iscrizione e, per i secondi dipende dall'autodichiarazione del lavoratore.

Figura 3 - Condizioni di non lavoro presenti nelle due fonti



Il terzo fattore di dissimilarità tra le fonti riguarda la *cadenza temporale delle osservazioni*: in *Inps* gli episodi sono documentati con unità di tempo il giorno (salvo quanto detto per gli iscritti alla gestione separata), mentre in *Panel* si registrano i segmenti di storie lavorative dichiarati dagli intervistati avendo come unità di tempo il mese.

Quarta ed ultima fonte di distorsione è legata al fatto che, mentre in *Panel* si registra la condizione prevalente, escludendo quindi la possibilità di sovrapposizione tra episodi, in *Inps* è registrato anche l'eventuale contemporaneo svolgimento di più lavori.

La procedura di integrazione delle due fonti di dati è avvenuta dapprima con un abbinamento esatto via codice fiscale. In altre parole, sono state recuperate le (eventuali) registrazioni nell'archivio *Inps* degli episodi di occupazione e/o di disoccupazione sussidiata che hanno caratterizzato la storia lavorativa di ciascun individuo appartenente all'onda 2012 del *Panel*²⁵.

Al fine di ovviare alle difformità sopra descritte, nell'integrare le informazioni tratte dalle due fonti si è seguito un insieme di criteri che privilegiano la condizione di occupazione e, nel complesso danno la prevalenza alla fonte amministrativa²⁶ (Figura 4).

Per quanto riguarda le storie lavorative, la fonte *Inps* è stata trattata come fonte:

- (a) esclusiva nel caso dei dipendenti privati extra-agricoli²⁷;
- (b) iniziale, integrata da *Panel*, nel caso di soggetti operanti nell'agricoltura (autonomi e dipendenti), artigiani e commercianti e imprenditori (ipotizzando che la loro assenza fosse dovuta al mancato versamento dei contributi).

Nei restanti casi, si è fatto ricorso:

- (c) al dato *Panel* come fonte esclusiva per gli episodi di lavoro non rinvenibili, per definizione, in *Inps* (lavoro dipendente nelle amministrazioni pubbliche e lavoro "nero");

²⁵ Una volta ultimato l'abbinamento, i codici fiscali sono stati sostituirli con identificativi generati da un apposito *tool* di anonimizzazione, in ragione delle disposizioni in materia di protezione dei dati personali. Le successive operazioni, fino alla realizzazione del file di microdati integrato, si sono basate sull'archivio anonimizzato prodotto da *Ispat* e sono state effettuate da *FBK-IRVAPP* presso il Laboratorio *ADELE* dell'Istituto

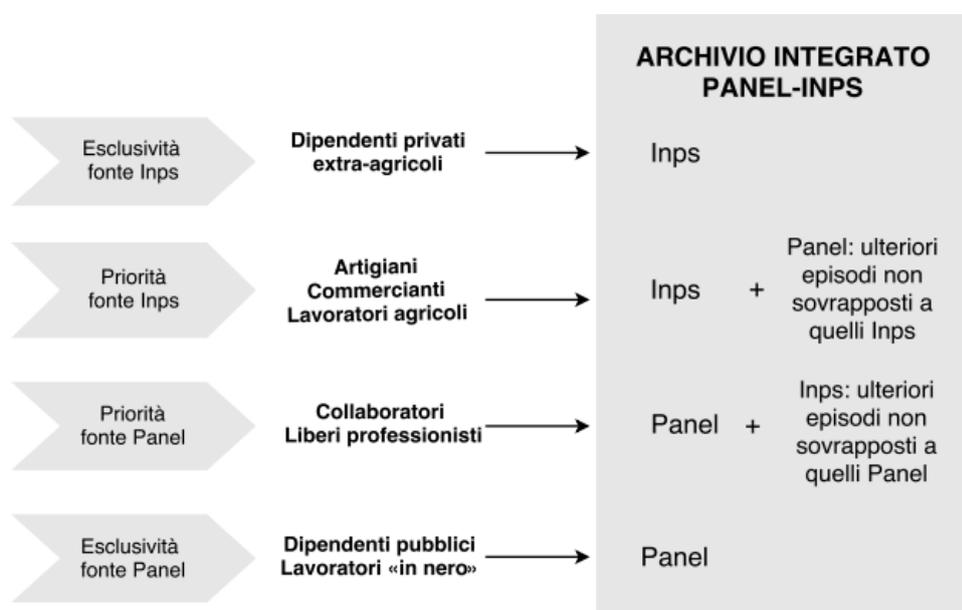
²⁶ Questo per limitare i possibili errori di risposta connessi ad una rilevazione tramite intervista.

²⁷ Si fa notare che per 71 persone dichiaratisi dipendenti privati extra-agricoli in *Panel* non sono stati riscontrati episodi di quel tipo in *Inps*. In particolare, 23 casi non compaiono nemmeno nell'archivio amministrativo perciò i loro episodi di lavoro dipendente sono stati sì mantenuti ma etichettati come "nero". Nei restanti casi si è scelto di eliminare solo quegli episodi di lavoro dipendente privato extra-agricolo rilevati tramite intervista qualora in *Inps*, nello stesso anno, fosse presente un episodio di lavoro in altra categoria (ipotizzando in tal modo l'esistenza di errori di classificazione in *Panel*).

- (d) prioritariamente al *Panel* per gli episodi di attività di libero professionista e di collaboratore: (d1) ai primi perché, se con cassa propria, presenti soltanto in *Panel*; (d2) ad entrambi perché rinvenibili in *Inps* con significativi limiti informativi sulle date di inizio e di fine²⁸.

Per quanto attiene, invece, gli episodi di disoccupazione, si è utilizzato il criterio di priorità a *Inps*, dando la priorità agli episodi di disoccupazione sussidiata in senso stretto (a meno di brevi episodi di occupazione dipendente, consentiti entro l'episodio di disoccupazione). Si è poi ricorso all'integrazione con dati *Panel* soltanto se – o comunque per la parte in cui – l'episodio non si sovrapponeva ai periodi di lavoro così come sopra descritti e ai periodi di disoccupazione sussidiata²⁹.

Figura 4 – Sintesi delle procedure di integrazione delle due fonti per la realizzazione dell'archivio integrato degli episodi di lavoro



In complesso, l'archivio integrato *Inps-Panel* comprende 5.489 individui con almeno un episodio di lavoro³⁰. Gli episodi lavorativi sono 48.580, dei quali 42.408 di fonte *Inps*, e 6.172 provenienti da *Panel* (Tabella 1). Gli episodi di disoccupazione sono, complessivamente, 6.804, rispettivamente 5.021 di fonte *Inps* e i rimanenti da *Panel*.

²⁸ Si ricorda che tutti gli episodi appartenenti alla gestione separata l'inizio e la fine sono convenzionalmente pari rispettivamente al 1° gennaio e al 31 dicembre.

²⁹ Si considera la disoccupazione *Inps* integrandola con episodi di disoccupazione *panel* solo se questi coprono nel file integrato *Panel-Inps* periodi di inattività (ovvero né lavoro né disoccupazione). In particolare, sono state individuate due situazioni. Nei casi (1.764) in cui la parte *panel* ricada in un periodo in cui in *Inps* non si trovano né episodi di lavoro né di non lavoro, si è aggiunto un nuovo episodio di disoccupazione. Nei casi in cui la parte *panel* è adiacente a *Inps* (193), gli episodi di disoccupazione sussidiata (*Inps*) vengono allungati con la componente non sussidiata (*Panel*)

³⁰ Vi sono poi 245 persone senza alcun episodio di lavoro.

Tabella 1 - Episodi di lavoro e di disoccupazione presenti nell'archivio integrato prima e dopo la risoluzione delle sovrapposizioni

| Tipo di episodi | Con sovrapposizioni | | | Senza sovrapposizioni | | |
|--|---------------------|----------------|---------------|-----------------------|----------------|---------------|
| | Inps | Panel | Totale | Inps | Panel | Totale |
| Dipendenti privati extra agricoli | 34.959 | 0 ^a | 34.959 | 29.648 | 0 ^a | 29.648 |
| Dipendenti Pubblica Amministrazione | n.d ^b | 3.367 | 3.367 | n.d ^b | 3.172 | 3.172 |
| Artigiani, Commercianti e Imprenditori | 1.465 | 204 | 1.669 | 1.379 | 188 | 1.567 |
| Liberi professionisti | 324 | 127 | 451 | 231 | 122 | 353 |
| Collaboratori | 3.205 | 263 | 3.468 | 1.743 | 249 | 1.992 |
| Agricoli (autonomi e dipendenti) | 2.455 | 592 | 3.047 | 2.210 | 507 | 2.717 |
| Nero | n.d ^b | 1.619 | 1.619 | n.d ^b | 1.553 | 1.553 |
| Totale occupazione | 42.408 | 6.172 | 48.580 | 35.211 | 5.791 | 41.002 |
| Totale disoccupazione | 5.021 | 1.783 | 6.804 | 5.000 | 1.783 | 6.783 |

^a Per gli episodi di lavoro dipendente extra- agricolo si è scelto un criterio di esclusività del dato amministrativo ignorando gli episodi di quel tipo dichiarati nell'indagine.

^b Gli episodi di lavoro dipendente pubblico e di lavoro "nero" non sono rinvenibili in Inps.

La tabella riporta la numerosità degli episodi anche al netto delle eventuali sovrapposizioni tra episodi di occupazione (doppi lavori) e tra occupazione e disoccupazione. I criteri seguiti sono stati i seguenti:

- i) Nel caso in cui un episodio di lavoro fosse completamente contenuto in un altro più esteso il primo è stato eliminato
- ii) Infine, nei casi di parziale sovrapposizione tra episodi di lavoro si è, dapprima stabilito un ordine gerarchico tra i lavori - lavoro dipendente privato (compresi gli agricoli), lavoro dipendente PA, lavoro autonomo (compresi gli agricoli), collaborazioni e lavoro in nero – e, in base a questo, si è modificata la data di inizio o fine dell'episodio gerarchicamente inferiore.
- iii) Nel caso di episodi di lavoro (non part-time) sovrapposti a episodi di disoccupazione sussidiata si è fatto prevalere la disoccupazione³¹.
- iv) Nel caso di sovrapposizione tra lavoro part-time e disoccupazione sussidiata o mobilità si è fatto prevalere il lavoro³².

Attraverso la risoluzione dei casi di sovrapposizione tra episodi il numero di episodi è stato ridotto di circa 7600 casi.

Oltre ai dipendenti privati extra agricoli per i quali il dato amministrativo è stato considerato come fonte esclusiva, l'apporto del dato Inps nel database integrato finale è risultato dominante anche nei casi in cui il dato panel è stato considerato prioritario, e segnatamente nel caso delle collaborazioni e delle libere professioni (Tabella 1).

Le ragioni possono essere diverse. Innanzitutto, come ricordato in precedenza, il Panel registra la condizione lavorativa prevalente, perciò, nel caso di doppi lavori (per esempio, un lavoro alle dipendenze e una collaborazione o il caso, peraltro frequente, di un insegnante con partita i.v.a.) gli intervistati tenderanno a dichiarare quello più importante o quello maggiormente stabile nel tempo. La seconda ragione è intrinseca al fatto che al momento della prima intervista del Panel vengono raccolte informazioni retrospettive a partire dall'ingresso nel mercato del lavoro, e che quanto più lontano è l'evento da ricordare tanto più difficile è per l'intervistato identificare eventi di breve periodo o verificatisi occasionalmente. L'integrazione con il dato amministrativo ha dunque l'indubbio vantaggio di correggere, almeno per alcuni tipi di episodi (Cfr. par. 2) gli errori di collocazione temporale, completare le omissioni di eventi e, in generale, arricchire la base

³¹ L'episodio di lavoro è stato tolto solo se totalmente incluso in quello di disoccupazione. Qualora, invece, la disoccupazione fosse collocata all'interno di un episodio di lavoro, quest'ultimo è stato spezzato in due episodi distinti. Infine, in caso di parziale sovrapposizione solo a destra o a sinistra dell'episodio di lavoro, la data di fine o di inizio del lavoro è stata modificata.

³² La logica seguita ricalca quella descritta nella nota precedente con l'unica differenza che, in questo caso, ciò che si va a modificare è l'episodio di disoccupazione.

informativa per lo studio della dinamica delle storie lavorative di un campione di popolazione trentina.

4. Validazione dell'archivio integrato

Definito l'archivio, diventa di indubbio interesse fornire un primo vaglio di concordanza delle stime dei tassi di partecipazione al mercato del lavoro con i dati ufficiali Istat basati sul campione provinciale della Rilevazione Continua delle Forze di Lavoro (Rcfl).

Nel caso di sovrapposizioni tra episodi di lavoro e di non lavoro è stata utilizzata la gerarchia usata nella Rcfl che dà la priorità alla condizione di occupato e, a seguire, a quella di disoccupato. Nell'archivio integrato si è fatto riferimento alla ricostruzione sezionale della popolazione rispetto al lavoro a maggio 2012³³, trattando le quattro osservazioni per ciascun individuo come misure ripetute. Per la rilevazione Istat si è, invece, considerato come riferimento temporale il II trimestre dello stesso anno. Tenendo conto che l'intervista panel è fatta a maggiorenni, la classe di età considerata è quella 18-64 anni.

Tabella 2 - Tassi di occupazione, disoccupazione e attività per la popolazione 18-64 anni, risultanti dall'archivio integrato Panel-Inps per il mese di maggio 2012 e dalla rilevazione sulle forze di lavoro per il II trimestre 2012

| Tipo di occupazione | Panel-Inps, maggio 2012 ^a | | Istat, II, trimestre 2012 ^b | |
|-------------------------|--------------------------------------|-------------------|--|------|
| | % | e.s. ^c | % | e.s. |
| Tasso di occupazione | 61,6 | 0,03 | 68,5 | 0,18 |
| Tasso di disoccupazione | 6,1 | 0,10 | 5,9 | 0,30 |
| Tasso di attività | 65,6 | 0,05 | 72,8 | 0,43 |

^a Media delle quattro settimane di maggio 2012.

^b Fonte: ISPAT.

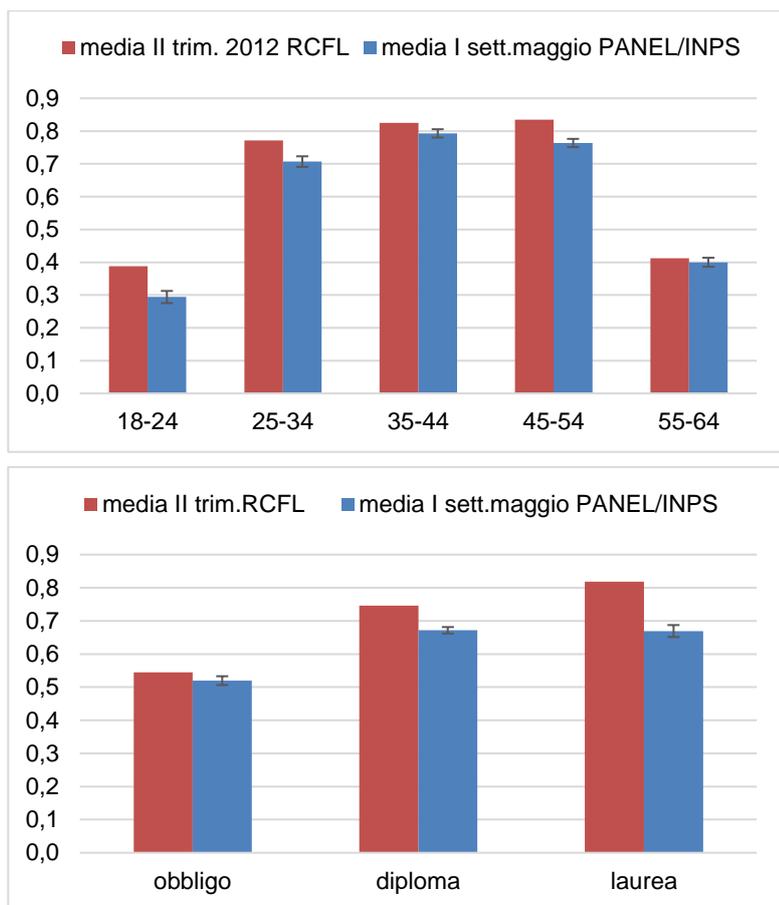
^c Stime degli errori standard *robust-clustered*.

Il confronto (Tabella 2) mostra una sorprendente coincidenza del tasso di disoccupazione. C'è, invece una significativa, ma non abnorme, sottostima dell'occupazione verosimilmente per fenomeni di trasferimento di residenza – entro e ancor più fuori della PaT – e di pendolarismo per motivi di lavoro. Va infatti ricordato che le regole di “inseguimento” dei membri del campione originario – e del *refreshment* sample introdotto nel 2010, non sono state stringenti, per vincoli organizzativi e di costo.

La sottostima degli occupati e degli attivi interessa sia la popolazione maschile (70,1% vs 76,2% Istat), sia quella femminile (53,6% vs 60,7%), ma è più evidente per i giovani con meno di 35 anni e per la fascia d'età 45-54 anni (Figura xx). Inoltre, è più accentuata tra i laureati (Figura xx).

³³ Si tratta del mese di avvio dell'ondata 2012 del panel.

Figura 5 – Tasso di occupazione secondo la fascia d'età e il titolo di studio



5. Conclusioni

Per effetto congiunto di interpretazioni restrittive di una normativa già di per sé restrittiva e, ancor più, dell'assenza di una adeguata cultura scientifica e politica, il tesoro di informazioni contenuto negli archivi della pubblica amministrazione italiana è rimasto a lungo improduttivamente sepolto nelle sue sedi e solo di recente ha iniziato a essere utilizzato, sia pure in modi sporadici.

Si deve, però, anche sottolineare che gli archivi amministrativi da soli non sono sempre in grado di rispondere agli interrogativi degli studiosi e alle esigenze di disegno e valutazione delle politiche pubbliche. Del tutto comprensibilmente, questi archivi contengono informazioni sui singoli individui che sono segmentate, filtrate, cioè, alla luce delle esigenze istituzionali. Essi mancano, così, di altri significativi dati che, per molti versi, risultano cruciali al fine di capire la configurazione stessa di quanto contenuto nell'archivio amministrativo e, soprattutto, per riuscire ad utilizzare quei dati in vista di effettuare analisi conoscitive e di mettere a punto e valutare le politiche pubbliche.

Una via – nota da tempo e da tempo praticata nei già ricordati paesi scandinavi, ma di difficile attuazione in Italia –, per ovviare al problema della segmentazione delle informazioni consiste nell'unire archivi amministrativi tra loro complementari. Tuttavia, neppure questa procedura, ammesso che sia davvero attuabile, si rivela sempre adeguata per rispondere alle domande di conoscenza e di valutazione che emergono dalla ricerca scientifica e dalla sfera politica. Un'innovativa e promettente procedura per superare queste possibili ulteriori inadeguatezze consiste nell'integrare, grazie alla strumentazione informatica, gli archivi amministrativi con indagini campionarie su larga scala.

In questo contesto ha preso il via il progetto sulle storie lavorative oggetto del presente paper. L'integrazione tra l'archivio amministrativo Inps e l'indagine campionarie Panel come strumento

per le analisi socio-economiche ci ha messo di fronte alle molteplici fonti di potenziale distorsione nei dati delle due fonti, alle difficoltà di identificarle, conseguentemente alla miriade di ostacoli che si incontrano nell'integrazione alla scala micro. Sovente sono state adottate soluzioni semplificate (ma plausibili per l'analisi dei cambiamenti, tanto più quanto più è ragionevole ritenere che la distorsione sia invariante – comunque presenti bassa variabilità – nel tempo). Ci siamo comunque inoltrati in un terreno ancora poco esplorato e, non appena sarà battuto da altri ricercatori, non potrà che migliorare.

Riferimenti

Bazzoli, M., Marzadro, S., Schizzerotto, A., Trivellato, U. (2018), *Come sono cambiate le storie lavorative dei giovani negli ultimi quarant'anni? Evidenze da uno studio pilota*, Mimeo.

Calderwood, L., Lessof, C. (2009). "Enhancing longitudinal surveys by linking to administrative data", in P. Lynn (Ed.), *Methodology of longitudinal surveys*, Chichester: Wiley, Ch. 4.

Caretta, A., Flisi, S., Frale, C. e Tedeschi, S. (2012). Il modello di microsimulazione T-DYMM: caratteristiche e potenzialità, in "Economia & lavoro, Rivista di politica sindacale, sociologia e relazioni industriali", 3: 61-78.

Fambri, G., Schizzerotto, A. (2008), a cura di. *Le condizioni di vita delle famiglie trentine. Rapporto di ricerca*, Quaderni della Programmazione n. 21, Provincia Autonoma di Trento e Università degli Studi di Trento, Trento: Edizioni 31.

Istat (2009). Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa, Metodi e Norme n. 38, Roma.

Sestito, P., Trivellato, U. (2011). "Indagini dirette e fonti amministrative: dall'alternativa all'ancora incompiuta integrazione", in *Rivista di Politica Economica*, luglio-settembre 2010-11, pp. 183-227.

Trivellato, U. (2017). Microdata for social sciences and policy evaluation as a public good, in Crato N. and Paruolo P. (Eds), FBK-IRVAPP Working Paper No 2017-06.

Appendice

La mobilità dei lavoratori da fonti amministrative: analisi dei flussi spuri

In complesso, l'analisi dei flussi spuri è stata svolta su un totale di 28.732 episodi di lavoro dipendente appartenenti alla gestione regolare. Partendo dai segmenti di carriera dei lavoratori sono stati dapprima identificati i passaggi diretti da un'azienda ad un'altra (i cosiddetti *job-to-job*) e, successivamente, sono stati considerati solo quelli avvenuti nell'arco di 30 giorni. Si è quindi giunti a costruire una base dati *ad hoc* del tipo *employee/employers* contenente, per ciascun individuo con almeno un *job-to-job* entro i 30 giorni (5.801), le caratteristiche dell'azienda di partenza e quelle dell'azienda di arrivo.

Per depurare gli episodi di mobilità selezionati dei passaggi spuri sono stati individuati due criteri:

- i) *La somiglianza degli attributi*: questa tecnica stabilisce un legame tra due aziende con matricola diversa sulla base della somiglianza nei valori di alcuni caratteri posseduti dalle imprese collegate e, segnatamente, la denominazione e l'indirizzo
- ii) *Identificazione delle trasformazioni societarie mediante movimenti dei lavoratori*: questa tecnica deduce l'esistenza di trasformazioni societarie tramite il flusso di dipendenti

Entrando nel dettaglio, per quanto attiene il primo criterio sono stati inizialmente candidati ad essere considerati movimenti *spuri* quelli intervenuti tra aziende con uguale denominazione. È questo il caso di imprese con più unità operative (eg. Unicredit). In complesso, si identificano 216 *job-to-job*.

Oltre alla denominazione, un secondo carattere di cui si è tenuto conto per stabilire un legame tra due imprese è stato l'indirizzo. Tuttavia, tenendo conto del fatto che uno stesso indirizzo può essere scritto in modi diversi (Via S. Croce 77, Trento vs via Santa Croce 77, Trento), si è preferito ricorrere alla geocodifica via google map. In base alla coerenza tra le coordinate geografiche, il codice di attività economica e la differenza tra la data di cessazione dell'azienda di provenienza e la data di costituzione dell'azienda di arrivo (*delta*) sono stati definiti i seguenti livelli di qualità dell'abbinamento:

- livello alto: aziende con stesso indirizzo, codice attività economica uguale (al 3° digit) e *delta* inferiore ai 30gg
- livello medio-alto: come il livello 5 senza vincolo dell'attività economica
- livello medio: come livello 5 ma con una discrepanza massima nelle coordinate geografiche delle due aziende di 0,0001 punti (corrispondenti a circa mezzo isolato: ne è un esempio un fabbricato con più numeri civici).
- livello medio-basso: come livello 3 ma senza vincolo dell'attività economica
- livello basso: discrepanza massima nelle coordinate geografiche delle due aziende di 0,001 punti (circa un isolato), codice attività economica al 2° digit, *delta* < 30gg

La medesima graduatoria è stata assegnata anche nei casi in cui la data di cessazione dell'azienda di provenienza e/o la data di costituzione dell'azienda di arrivo fosse mancante. La tabella riassume il numero di *job-to-job* identificati con la geolocalizzazione (190) secondo la qualità dell'abbinamento.

Tabella A1 – Numero di job-to-job secondo il livello di precisione della geolocalizzazione e la completezza delle informazioni disponibili

| Qualità della geolocalizzazione | Casi con informazioni complete | Casi con informazioni incomplete ^a |
|---------------------------------|--------------------------------|---|
| Alta | 45 | 75 |
| Medio-alta | 7 | 40 |
| Media | 2 | 1 |
| Medio-bassa | 0 | 3 |
| Bassa | 9 | 8 |
| Totale | | 190 |

^a Casi in cui manca la data di cessazione dell'azienda di provenienza e/o la data di costituzione dell'azienda di arrivo.

Passando al secondo criterio, per individuare l'esistenza di trasformazioni societarie si è, dapprima, considerato il numero di individui che hanno effettuato uno stesso passaggio d'azienda. Tenuto conto che si sta osservando un campione di circa 5.000 lavoratori, qualora vi siano almeno tre individui con questa caratteristica, sembra sensato ipotizzare che si tratti di mobilità spuria. Tutti questi casi (101) sono dunque passati direttamente al controllo manuale.

Nei casi in cui si siano verificati dei job-to-job nello stesso mese (133 casi) o su mesi adiacenti (3 casi) si è, invece, seguita una procedura diversa basata sulle variazioni nell'ampiezza delle aziende e, più precisamente, sulle serie mensili dei dipendenti. L'algoritmo utilizzato a tale scopo tiene conto della variazione del numero di dipendenti nel momento del passaggio d'azienda: il delta dell'azienda di provenienza ($\Delta \text{azienda_da}$) corrisponde alla differenza nel numero di dipendenti tra il tempo t (in cui si verifica il job-to-job) e il tempo $t-1$; parimenti, il delta dell'azienda di arrivo ($\Delta \text{azienda_a}$) corrisponde alla differenza nel numero di dipendenti tra il tempo t e il tempo $t+1$. La logica sottostante è che quanto più simili sono i Δ tanto più probabile sarà la presenza di mobilità spuria dovuta a trasformazioni societarie (Spin Off, Acquisizioni Azienda, Acquisizioni Ramo Azienda). I casi identificati (104) variano a seconda del numero di dipendenti dell'azienda di arrivo:

- $|\Delta \text{azienda_da} - \Delta \text{azienda_a}| \leq 2$ se $N_{\text{dip_a}} < 20$
- $|\Delta \text{azienda_da} - \Delta \text{azienda_a}| \leq 5$ se $N_{\text{dip_a}}$ compreso tra 20 e 49
- $|\Delta \text{azienda_da} - \Delta \text{azienda_a}| / N_{\text{dip_a}} \leq 0.1$ se $N_{\text{dip_a}} \geq 50$

Successivamente sono stati isolati i casi di chiusura dell'azienda di provenienza qualora il numero dipendenti di questa passasse da N a 0 (8 casi):

- $\Delta \text{azienda_da} = n_{\text{dip_da}}$

Validazione della categorizzazione in mobilità spuria

In conclusione, la somiglianza degli attributi ha consentito di individuare 406 casi di probabile mobilità spuria, rispettivamente 216 sulla base della denominazione e 190 via geolocalizzazione. Dall'analisi dei flussi dei lavoratori ne sono stati, invece, identificati 236. Tenuto conto che alcuni casi sono stati individuati da più di un criterio (vedi tabella sottostante), si è proceduto al controllo manuale di 525 passaggi aziendali attraverso la consultazione dei siti web delle unità coinvolte. Complessivamente il 90,3% dei controlli ha avuto esito positivo cioè ha confermato l'ipotesi di mobilità spuria.

Tabella A2 – Numero di probabili passaggi aziendali spuri secondo il criterio utilizzato per identificarli e l'esito del controllo manuale

| Criterio | Controllo manuale | | N |
|---|-------------------|----------|-----|
| | negativo | positivo | |
| Geolocalizzazione (tutti i livelli) | 30 | 112 | 142 |
| Denominazione | 0 | 147 | 147 |
| N dipendenti | 21 | 124 | 145 |
| Geolocalizzazione+denominazione | 0 | 23 | 23 |
| Geolocalizzazione+N dipendenti | 0 | 22 | 22 |
| Denominazione + N dipendenti | 0 | 43 | 43 |
| Geolocalizzazione+denominazione+ n dipendenti | 0 | 3 | 3 |
| Totale | 51 | 474 | 525 |

Tutti i casi individuati via somiglianza nella denominazione sono risultati validati. Il 100% di successo si riscontra anche per i job-to-job identificati da più di un criterio. Per quanto riguarda i passaggi aziendali identificati solo sulla base della comunanza degli indirizzi la probabilità di successo è del 79% se si considerano tutti i livelli di qualità dell'abbinamento, ma sale all'88% se si guarda solo al livello più alto di questa. Anche la presenza, nel campione, di almeno 3 individui con lo stesso passaggio di azienda è un buon indicatore di mobilità spuria: infatti, il 90% di questi casi è stato confermato dal controllo manuale. Per quanto riguarda, infine, le informazioni derivate dalle serie mensili dei dipendenti, gli algoritmi utilizzati hanno avuto un riscontro positivo in circa nove casi su dieci.

Tabella A3 – Esito del controllo manuale sui passaggi aziendali identificati tramite geolocalizzazione

| Qualità della geolocalizzazione | Esito controllo manuale | | N |
|--------------------------------------|-------------------------|----------|-----|
| | negativo | positivo | |
| Informazioni complete | | | |
| Alta | 5 | 40 | 45 |
| Medio-alta | 0 | 7 | 7 |
| Media | 0 | 0 | 0 |
| Medio-bassa | 0 | 2 | 2 |
| Bassa | 3 | 6 | 9 |
| Informazioni incomplete ^a | | | |
| Alta | 9 | 66 | 75 |
| Medio-alta | 7 | 33 | 40 |
| Media | 0 | 1 | 1 |
| Medio-bassa | 2 | 1 | 3 |
| Bassa | 4 | 4 | 8 |
| Totale | 30 | 160 | 190 |

^a Casi in cui manca la data di cessazione dell'azienda di provenienza e/o la data di costituzione dell'azienda di arrivo.

Tabella A4 – Esito del controllo manuale sui passaggi aziendali identificati utilizzando la loro diffusione nel campione

| Numero di job-to-job nel campione | Esito controllo manuale | | N |
|-----------------------------------|-------------------------|----------|-----|
| | negativo | positivo | |
| 3 | 5 | 41 | 46 |
| 4 | 2 | 14 | 16 |
| 5 | 0 | 13 | 13 |
| 6 | 0 | 6 | 6 |
| 7 | 0 | 5 | 5 |
| 8 | 0 | 1 | 1 |
| 9 | 1 | 3 | 4 |
| 10 o più ^a | 0 | 10 | 10 |
| Totale | 21 | 192 | 213 |

^a Nel campione sono stati individuati fino a 34 job-to-job

Tabella A5 – Esito del controllo manuale sui passaggi aziendali identificati via serie mensile dei dipendenti

| Serie mensili dipendenti per job-to-job=2 | Esito controllo manuale | | N |
|---|-------------------------|------------|------------|
| | negativo | positivo | |
| Coerenza nei delta | 12 | 92 | 104 |
| Chiusure aziendali senza coerenza nei delta | 1 | 7 | 8 |
| Totale | 21 | 192 | 213 |

L'identificazione di 474 passaggi aziendali spuri ha consentito di correggere 1.081 episodi lavorativi (corrispondenti al 3,8% del totale dei rapporti di lavoro dipendente).