

Lavoro: partecipazione, dinamica e valutazione di politiche

Misura, metodi, modelli

Progetto anno 2005

Giove-VWH: Esperienza di *linkage* tra data base amministrativi

D. Maurizio^{*}, *E. Trevisan*^{**}

* *Veneto lavoro*

** *Università di Venezia & Veneto lavoro*

Working Paper n. 78, marzo 2009

La valutazione dell'impatto di interventi pubblici: metodi e studi di caso.

Cofinanziamento MIUR, anno 2005

Unità locali del progetto:

Dip. di Economia "S. Cagnetti De Martiis", Univ. di Torino

(coord. B. Contini)

Dip. di Statistica "G.Parenti", Univ. Di Firenze

(coord. F. Mealli)

Dip. di Scienze Statistiche, Univ. di Padova

(coord. E. Rettore)

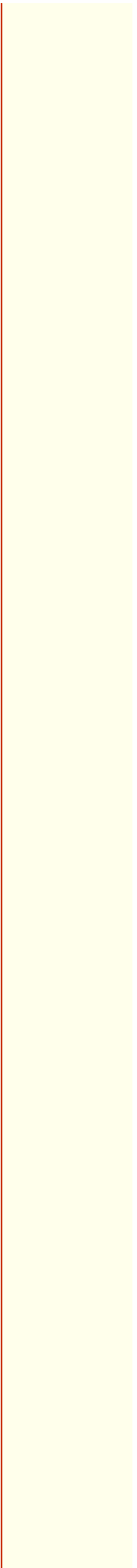
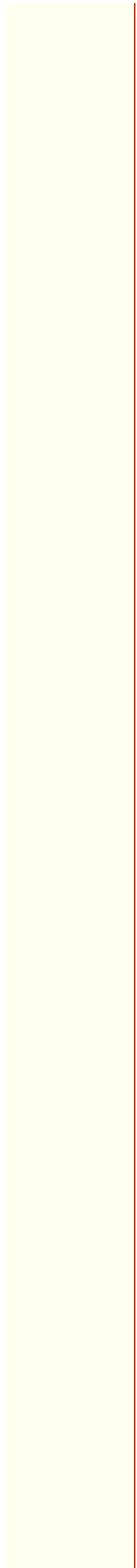
Dip. di Scienze Economiche, Univ. di Salerno

(coord. S. Destefanis)

Dip. di Politiche Pubbliche e Scelte Collettive, Univ. del Piemonte Orientale

(coord. D. Bondonio)

Dipartimento di Scienze Statistiche
via C. Battisti 241-243, 35121 Padova
<http://valutazione2003.stat.unipd.it>



1. Introduzione¹

Le potenzialità di ricerca offerte dai dati amministrativi e le disponibilità di due dataset – Giove e VWH (Veneto Workers Histories) - contenenti informazioni, seppure diverse, complementari sull'intera popolazione di lavoratori e imprese in Veneto, sono state alla base dell'avvio di questo progetto di linkage.

In particolare, l'utilizzo dei due *dataset* è stato possibile nell'ambito dell'attività di ricerca svolta da Elisabetta Trevisan presso il Dipartimento di Scienze Economiche dell'Università di Venezia, sotto la supervisione del prof. Giuseppe Tattara. Quest'ultimo, infatti ha ottenuto, previo apposita convenzione, gli archivi amministrativi Inps per il Veneto. Il Dipartimento, sulla base degli archivi amministrativi VWH riferiti al Veneto, ha elaborato il data base VWH (Veneto Worker Histories) ed ha, altresì, aderito al protocollo di ricerca per l'utilizzazione del *Public Use File* Giove. Veneto Lavoro, invece, nella persona di Danilo Maurizio ha fornito il supporto tecnico all'utilizzazione di Giove.

L'idea sottostante questo progetto di *linkage* tra Giove e VWH fa riferimento alle potenzialità offerte dalla possibilità di arricchire il contenuto informativo dei due archivi, sia dal punto di vista della quantità di informazioni disponibili per imprese e lavoratori, sia dal punto di vista dell'allungamento dell'orizzonte temporale coperto dal *data set* congiunto.

Il desiderio di accrescere il contenuto informativo e la qualità dei dati esistenti ben si colloca in un contesto di ricerca dove, nell'ultimo decennio, l'interesse e il conseguente utilizzo dei dati amministrativi – come forma aggiuntiva e più completa rispetto alle fonti tradizionali, come indagini campionarie e *survey* – è aumentato notevolmente.

La crescente diffusione e utilizzo dei dati amministrativi di varia origine ha rappresentato una risposta ad alcuni problemi che si sono posti i produttori di statistiche ufficiali, da un lato, e i ricercatori dall'altro. Dal lato della ricerca, infatti, vi è stato un incremento e una diversificazione della domanda di informazione statistica, con particolare riferimento alla crescente necessità di dati adatti ad analisi più dettagliate, come ad esempio quelle relative a particolari categorie di soggetti, a microaree e nuovi fenomeni. Dal lato dei fornitori, invece, si è delineata l'impossibilità di far fronte a questa nuova domanda incrementando il numero di indagini statistiche tradizionali. La crescente domanda si scontrava, infatti, con la necessità di ridurre i tempi e di contenere i costi di produzione (in termini di risorse sia umane che finanziarie), di accrescere il "carico statistico" gravante sui rispondenti, in particolare gli agenti economici (Costanzo, 2008).

In generale, i dati amministrativi offrono una base informativa a basso costo. Tuttavia, bisogna tener conto del fatto che la loro raccolta è progettata ed eseguita per fini diversi da quelli statistici.

Per questo motivo, l'approccio ad una fonte amministrativa impone, in primo luogo, una verifica preliminare al fine di valutare le potenziali pregiudiziali della qualità delle informazioni rilevabili. Un primo elemento riguarda il fatto che la popolazione di interesse potrebbe non coincidere con quella rilevata dalla fonte amministrativa (cosiddetti *errori di copertura*). Secondariamente, i concetti, le definizioni e le classificazioni adottate dalla fonte amministrativa possono non coincidere con quelle della ricerca statistica. Infine, la qualità dei dati può essere insufficiente, soprattutto, per quelle variabili ritenute non fondamentali per le procedure amministrative, ma invece fondamentali ai fini statistici. Le difficoltà, quindi, derivano principalmente dal non-controllo del processo di produzione, che comporta un

¹ Il progetto di linkage dei due archivi è stato sviluppato, e parzialmente finanziato, nell'ambito del PRIN 'La valutazione dell'impatto di interventi pubblici: metodi e studi di caso', Anno 2005 - prot. 2005131989.

diverso approccio al problema della qualità, rispetto alle fonti tradizionali, con particolare riferimento a tre aspetti²: *rilevanza*³, *accuratezza e completezza*⁴.

Tenendo conto di queste limitazioni, i dati amministrativi presentano tuttavia una serie di indubbi vantaggi, che si misurano soprattutto in termini di costi, copertura e tempestività⁵:

- la riduzione dei costi deriva dal fatto che la raccolta avviene per altri scopi, con mezzi e risorse che non impegnano il produttore di statistiche ufficiali;
- la copertura è generalmente totale e quindi consente di produrre dati territoriali dettagliati diversamente da quanto avviene con le indagini campionarie;
- se la rilevazione è continua, come spesso avviene, è possibile accedere in maniera tempestiva ai dati.

Tenendo conto di tutti questi aspetti, è dunque possibile individuare alcuni ambiti in cui la diffusione dell'uso dei dati amministrativi può accrescere e migliorare l'informazione statistica disponibile, riducendone i costi. In particolare, l'utilizzo dei dati amministrativi consentirebbe, ad esempio, la produzione diretta di informazione statistica, l'integrazione dei sistemi informativi, la sostituzione di indagini statistiche (campionarie) con dati amministrativi (totali), l'elaborazione di stime dirette che utilizzino congiuntamente dati statistici e amministrativi, la produzione di archivi di estrazione per le indagini campionarie, la validazione dei risultati di indagini statistiche (Costanzo, 2008).

In quest'ottica, questo *paper* ha lo scopo di descrivere l'esperienza di *linkage* tra due data base amministrativi – Giove e VWH – cercando di evidenziare, al di là degli aspetti tecnici, le criticità e le potenzialità di utilizzo di questi dati.

Nel paragrafo successivo vengono descritte la struttura e le caratteristiche dei due archivi. Il paragrafo 3 descrive la procedura di *linkage* tra i dati, sia per quanto attiene agli archivi delle imprese che a quelli dei lavoratori e i risultati ottenuti. Nel paragrafo 4 sono descritti brevemente vantaggi e cautele nell'utilizzo del *link Giove-VWH*. Il paragrafo 5 offre un'analisi descrittiva delle caratteristiche di imprese e lavoratori che rimangono esclusi dal raccordo finale tra le due banche dati, sia con riferimento all'intero periodo che ad un anno specifico: il 2001. Infine, nel paragrafo 6 sono contenute alcune brevi considerazioni conclusive.

2. Le banche dati utilizzate

2.1. VWH⁶

Il database VWH è un archivio amministrativo contenente informazioni riguardanti le imprese attive⁷ in Veneto nel periodo dal 1975 al 2001 e i lavoratori delle stesse imprese, indipendentemente dalla provincia di residenza e dalla durata del rapporto di lavoro. I dati

² Cfr. Costanzo (2008).

³ È infatti difficile (se non impossibile) definire a priori un sistema di metadati coerente con gli obiettivi delle indagini.

⁴ È spesso necessario integrare fonti diverse, sia per l'individuazione e la correzione degli errori, sia per ottenere una descrizione completa del fenomeno di interesse.

⁵ Cfr. Costanzo (2008).

⁶ Cfr. Tattara e Valentini (2006).

⁷ Nell'archivio sono registrate le singole unità locali di ogni impresa.

sulle imprese provengono dagli archivi relativi ai modelli mensili DM10⁸, mentre i dati sui lavoratori sono ricavati dai modelli OM1⁹.

L'universo coperto dagli archivi è quello delle imprese private con lavoratori alle dipendenze. Sono escluse le imprese operanti nel settore agricolo, nella Pubblica Amministrazione e in altri servizi pubblici che fanno riferimento, per la gestione delle pratiche contributive, ad Istituti diversi da Inps (ad esempio Sanità e Ferrovie dello Stato). Sono escluse altresì tutte le imprese prive di lavoratori dipendenti¹⁰.

Più specificatamente, la banca dati VWH è strutturata in tre archivi separati: archivio anagrafico, archivio contributivo e archivio aziendale.

L'archivio anagrafico contiene *record* relativi a tutti i lavoratori di imprese venete¹¹, per i quali siano stati versati contributi Inps nel periodo 1975-2001¹², nonché quelli relativi a lavoratori impiegati in aziende (unità locali) non venete ma per i quali la sede Inps di riferimento è quella di residenza. I dati sono individuali e anonimi: i lavoratori sono identificati attraverso un codice progressivo. Per ogni lavoratore sono presenti informazioni riguardanti il genere, la data di nascita, il comune (o lo stato estero) di nascita e il comune di residenza. Mancano, invece, informazioni riguardanti il titolo di studio del lavoratore, lo stato civile e altre caratteristiche riguardanti gli altri componenti del nucleo familiare di appartenenza.

L'archivio contributivo, che rappresenta la parte più consistente – sia in termini di informazioni che di dimensione – del data base VWH, è costruito sulla base dei modelli OM1 relativi ai lavoratori dipendenti presentati dalle aziende dal 1975 al 2001. Ogni *record* è identificato da un codice progressivo¹³ e dalla matricola aziendale relativa all'azienda in cui, in quell'anno specifico, il lavoratore è assunto. Le informazioni contenute nell'archivio riguardano la qualifica, la retribuzione totale annua, giorni-settimane-mesi retribuiti, il tipo di contratto, tipo di rapporto, livello di inquadramento e data di cessazione per ogni rapporto di lavoro verificatosi all'interno dell'anno.

L'archivio aziendale, infine, è composto da *record* relativi a tutte le imprese (unità locali) che hanno operato in Veneto dal 1975 al 2001¹⁴. Ad ogni azienda è associata una matricola aziendale e alcune altre informazioni, come la ragione sociale, il codice fiscale e la partita IVA, che permettono di identificare univocamente l'azienda. Per ogni impresa, nell'archivio sono contenute informazioni relative a indirizzo, data di costituzione-sospensione-cessazione, attività economica, dimensione¹⁵.

2.2. Giove¹⁶

Il database statistico Giove è il risultato di un'attività di correzione, integrazione e editing dei dati amministrativi estratti dalle banche dati in gestione presso i Cpi del Veneto. I dati raccolti sono per la maggior parte generati dal flusso di comunicazioni a cui le imprese sono tenute

⁸ I modelli DM10 sono quelli relativi alla denuncia del personale alle dipendenze a livello aziendale.

⁹ I modelli OM1, relativi ai singoli lavoratori, erano quelli utilizzati per il calcolo dei diritti maturati individualmente ai fini pensionistici. Dal 2005 sono stati sostituiti dal sistema Emens.

¹⁰ Imprese in cui lavorano solamente il titolare, i soci e/o i familiari, come succede per le micro-imprese artigianali e commerciali.

¹¹ Per imprese Venete si intendono tutte le imprese con sede in Veneto, unità locali localizzate in Veneto, anche se facenti parte di imprese non Venete.

¹² Indipendentemente dalla durata dell'episodio lavorativo.

¹³ Il codice progressivo è lo stesso che identifica i lavoratori nell'archivio anagrafico.

¹⁴ Nell'archivio è presente la totalità delle imprese, sia quelle attive che quelle cessate.

¹⁵ La dimensione aziendale è stata ricostruita sulla base del numero di lavoratori presenti al 31 ottobre di ogni anno e si trova in un file separato.

¹⁶ Cfr. Anastasia e altri (2006).

per obbligo di legge¹⁷ e, in misura minore, dal flusso di informazioni raccolte quando il lavoratore in cerca di occupazione spontaneamente si presenta agli sportelli dei servizi per l'impiego¹⁸.

L'universo coperto dal database è dato da tutti i lavoratori e le imprese operanti nel settore privato in Veneto dal 1998. Più precisamente, sono presenti tutte le unità locali localizzate nel territorio di pertinenza di ogni Cpi e quelle 'esterne'¹⁹. Dal 2007, esso è stato esteso anche ai lavoratori parasubordinati e a lavoratori e aziende operanti nella pubblica amministrazione.

Il database è strutturato su quattro tabelle principali: anagrafica lavoratori, anagrafica aziende, anagrafica agenzie e rapporti di lavoro; a queste si aggiungono una serie di tabelle di appoggio.

L'anagrafica dei lavoratori contiene dati relativi a tutti i lavoratori oggetto di almeno una comunicazione obbligatoria da parte delle aziende venete presso le quali lavorano o hanno lavorato a partire dal 1998. La tabella contiene informazioni relative al codice fiscale, Cpi di riferimento, genere, cittadinanza, data di nascita, comune di nascita, residenza e domicilio, titolo di studio.

L'anagrafica delle aziende contiene informazioni derivanti dalle comunicazioni obbligatorie che le imprese devono rendere ai Centri per l'Impiego. La tabella contiene informazioni relative al ragione sociale, codice fiscale, partita IVA, indirizzo e attività economica.

In modo del tutto analogo all'anagrafica aziende è stata costruita l'anagrafica delle agenzie di lavoro interinale. Nella tabella sono riportate le stesse informazioni dell'anagrafica delle aziende, ad eccezione del codice di attività economica.

La tabella dei rapporti di lavoro contiene le storie lavorative dei soggetti. In essa sono contenute informazioni relative a data di avviamento e di cessazione del rapporto di lavoro, tipologia contrattuale, orario di lavoro, qualifica e livello.

Le tabelle di appoggio, infine, contengono codifiche e informazioni relative ai campi valorizzati nelle 4 tabelle principali.

3. Il link: struttura e procedura

I due database, pur essendo entrambi basati su dati amministrativi e fornendo un set di informazioni simili e complementari, non sono completamente omogenei. Le diversità sono molteplici e si riferiscono al metodo di raccolta dei dati e ai suoi fini, al livello di aggiornamento e all'orizzonte temporale coperto, nonché alla struttura stessa dei due archivi, con particolare riferimento alla ricostruzione dei percorsi lavorativi dei soggetti.

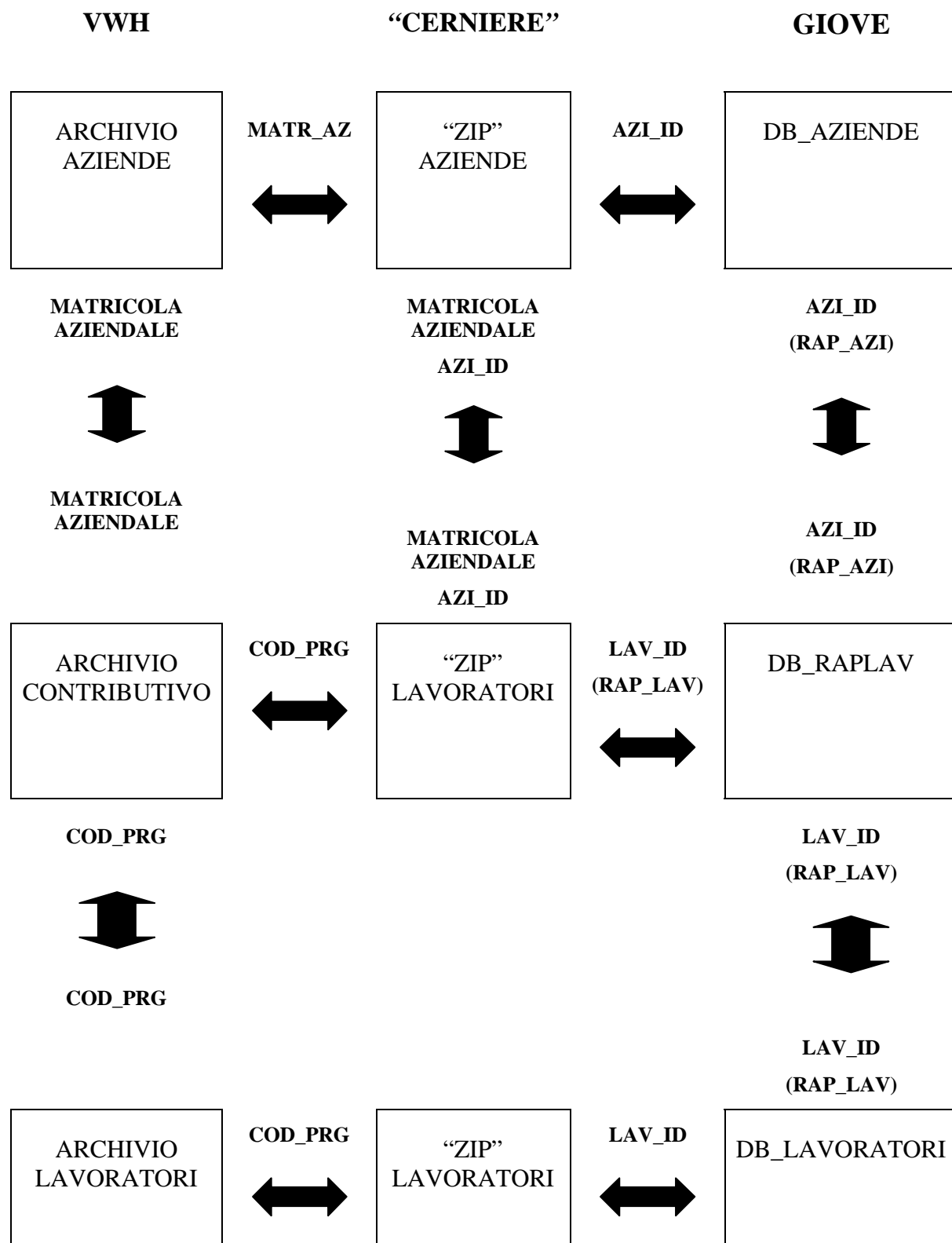
Come si evince dalla descrizione delle due banche dati nel paragrafo precedente, i dati VWH sono costruiti sulla base di informazioni raccolte ai fini del calcolo dei *benefit* pensionistici, mentre i dati contenuti in Giove derivano dalle comunicazioni obbligatorie effettuate dalle imprese ai Centri per l'Impiego. Questa differenza determina, per quanto riguarda i dati VWH, una maggiore ricchezza di informazioni rispetto ad ogni singolo rapporto di lavoro, soprattutto in termini di salario; vi sono anche maggiori informazioni sulle dinamiche delle imprese, sia in termini dimensionali che di mortalità; più scarse le informazioni anagrafiche sui lavoratori. In Giove, le informazioni aziendali sono prevalentemente di tipo anagrafico e non ci sono informazioni relative alla dimensione d'impresa, alle date di creazione e cessazione; rispetto, ad VWH, le informazioni relative ai lavoratori sono più ricche offrendo, tra l'altro, indicazioni relative al titolo di studio e un maggior dettaglio rispetto alle diverse tipologie contrattuali.

¹⁷ Assunzioni, cessazioni e trasformazioni di rapporti di lavoro.

¹⁸ Cfr. Maurizio (2006).

¹⁹ Purché, pur essendo afferenti ad un altro Cpi, abbiano movimentato lavoratori afferenti al Cpi in questione.

Fig. 1: Struttura del link tra Giove e VWH.



Per quanto riguarda l'orizzonte temporale coperto, i dati contenuti nella banca dati VWH riguardano il periodo dal 1975 al 2001, mentre i dati contenuti in Giove coprono il periodo tra il 1998 e il 2006²⁰. In questo modo, al di là del periodo comune 1998-2001, VWH permette una ricostruzione retrospettiva dei percorsi lavorativi dei soggetti, mentre Giove permette di aggiornare continuamente la situazione lavorativa degli individui.

Sulla base di queste disomogeneità e data la notevole dimensione delle due banche dati, che ne renderebbe difficoltoso l'utilizzo una volta unite, si è deciso di creare una sorta di piattaforma comune tra i due dataset. In altre parole, sono stati creati due file '*cerniera*' – uno per le imprese e uno per i lavoratori – contenenti i *record* identificativi, che permettono di agganciarsi ai due database originari per ottenere le informazioni di interesse. La **fig. 1** descrive la struttura ottenuta e le variabili identificative che permettono di legare i vari dataset.

Per facilitare la comprensione dell'utilizzo delle *cerniere* facciamo un esempio. Supponiamo di avere il codice fiscale di un lavoratore del quale vogliamo conoscere le caratteristiche e la storia lavorativa. Date le informazioni in nostro possesso (i.e. codice fiscale) la ricostruzione del profilo del lavoratore partirà da Giove e il percorso è il seguente:

- Con il codice fiscale si accede a *dbb_lavoratori*, ottenendo le informazioni anagrafiche del lavoratore e il codice identificativo (i.e. *lav_id*);
- Con il codice identificativo, utilizzando '*zip*' *lavoratori* si ottiene il codice identificativo del lavoratore in VWH. In questo modo, si può facilmente accedere all'archivio anagrafico e ottenere ulteriori informazioni.
- Sempre attraverso i due codici identificativi si può accedere rispettivamente a *dbb_raplav_4* (archivio dei rapporti di lavoro di Giove) e all'archivio contributivo di VWH. In questo modo, si ricostruisce l'intera storia lavorativa del soggetto;
- Attraverso la ricostruzione della storia lavorativa si ottengono i codici identificativi delle aziende per le quali il soggetto ha lavorato (*rap_azi* e *matr_az*);
- I codici identificativi delle aziende permettono di ottenere informazioni relative alle aziende da entrambi gli archivi.

Per completare la descrizione della nuova struttura ottenuta, vediamo in dettaglio le procedure di *linkaggio* tra i due dataset.

3.1. '*zip*' aziende: procedura

La procedura di link tra i due dataset è partita dalla creazione della '*cerniera*' tra gli archivi aziendali. Tale scelta è stata determinata dal fatto che solo per le unità locali, in entrambi gli archivi, è presente il codice fiscale/partita IVA (ossia, una chiave di identificazione univoca). Nell'anagrafica aziendale in Giove sono registrate 483.493 unità locali, mentre in quella di VWH sono registrate 1.126.568 unità locali.

La procedura di link tra i due dataset può essere riassunta in tre fasi principali:

- *Fase 1*: utilizzo di codice fiscale e partita IVA per individuare per ogni unità locale in Giove la corrispondente unità locale in VWH;

²⁰ Il link è stato effettuato su Giove2006, contenenti informazioni sino al 2006; la versione aggiornata di Giove contiene dati sino al 2007.

- *Fase 2*: utilizzo di ragione sociale, indirizzo, comune e CAP per individuare ‘coppie’ di aziende Giove-VWH per le quali il codice fiscale e partita IVA non sono riportati o non sono corretti²¹;
- *Fase 3*: le informazioni utilizzate nella seconda fase sono state usate per controllare ulteriormente il *linkage* ottenuto con la *Fase 1*.

La procedura descritta ha portato, dunque, alla creazione di una tabella di raccordo – ‘*zip*’ aziende – contenente per ogni unità locale presente in Giove (caratterizzata dal codice identificativo *azi_id*) il codice identificativo – *matr_az* – della corrispondente unità locale VWH.

La tabella di raccordo contiene 341.778 record, corrispondenti a circa il 70% delle unità locali presenti nell’anagrafica aziendale di Giove.

La tabella ‘*zip*’ aziende, oltre ad essere la tabella di raccordo tra le due anagrafiche aziendali, gioca, come accennato in precedenza, un ruolo fondamentale anche nella creazione del link tra i lavoratori, data la mancanza di una chiave di identificazione univoca dei lavoratori in VWH (i.e. codice fiscale).

3.2. ‘*zip*’ lavoratori: procedura

L’anagrafica dei lavoratori di Giove contiene 2.421.761 record e l’archivio anagrafico VWH contiene 3.718.858 lavoratori.

Come già anticipato nel paragrafo precedente, la procedura di link tra i due dataset è in questo caso più complicata a causa della mancanza di una chiave univoca di identificazione per i lavoratori in VWH.

La procedura si è quindi articolata nel modo seguente:

- *Fase 1*: utilizzando la tabella ‘*zip*’ aziende sono stati individuati i corrispondenti lavoratori in Giove e VWH;
- *Fase 2*: informazioni relative a genere, data di nascita e comune di nascita²² vengono utilizzate per fare il link tra i lavoratori di Giove e VWH. In caso di corrispondenza delle tre informazioni, la coppia di lavoratori viene inserita nella tabella di raccordo (*flusso 1*), mentre se nessuna delle informazioni corrisponde la coppia non viene considerata;
- *Fase 3*: per i casi nei quali una delle 3 informazioni utilizzate nella fase due non corrisponda o sia mancante sono stati confrontati gli episodi lavorativi (data di assunzione e data di cessazione del rapporto di lavoro). Qualora lo spell lavorativo corrisponda la coppia viene inserita nella tabella di raccordo (*flusso 2*);
- *Fase 4*: individuazione di eventuali match ‘1 a molti’ in entrambe le direzioni (da Giove a VWH, da VWH a Giove);
- *Fase 5*: diagnostica su quanto individuato nella fase 4. La procedura di diagnostica si è articolata nel seguente modo: controllo del flusso di provenienza ed eliminazione delle coppie derivanti dal *flusso 2* perché meno affidabile; su i rimanenti è stato effettuato un controllo utilizzando il comune di residenza: se dal controllo risulta un link ‘1-1’ la coppia viene inserita nella tabella di raccordo, se non viene individuato e se permane il link ‘1 a molti’ si è proceduto con un controllo degli spell lavorativi.

²¹ L’assenza di codice fiscale/partita IVA o la loro inesattezza sono stati riscontrati prevalentemente nell’archivio aziendale di Inps.

²² Queste informazioni corrispondono all’ultima parte del codice fiscale.

Anche in questo caso, la procedura descritta ha portato alla creazione di una tabella di raccordo – ‘zip’ lavoratori – contenente i codici identificativi dei lavoratori presenti in Giove (*lav_id*) e dei corrispondenti lavoratori VWH (*cod_prg*). La tabella di raccordo contiene 1.396.674 record, corrispondenti a circa il 58% del totale dei lavoratori presenti in Giove e il 69% di quelli occupati presso le aziende presenti in ‘zip’ aziende.

4. Vantaggi e cautele

Come già accennato nell’introduzione, l’utilizzo dei dati amministrativi nell’attività di ricerca presenta una serie di indiscussi vantaggi, riconducibili principalmente alla possibilità di ottenere analisi più dettagliate e puntuali su tutta la popolazione di interesse.

Tuttavia, l’utilizzo degli archivi amministrativi spesso implica costi maggiori in termini di tempi di computazione ed elaborazione, rispetto all’utilizzo dei dati di *survey*. L’elevata numerosità che caratterizza i dataset costruiti sulla base di dati amministrativi, inoltre, è spesso accompagnata da un numero ridotto di informazioni (i.e. variabili) che possono essere utilizzate nell’analisi statistica. Questa contrapposizione può creare, nell’attività di ricerca, non pochi problemi di tipo metodologico e interpretativo.

Il link tra i dati Giove e VWH, così come è stato concepito e strutturato, ovvia – se non totalmente almeno in parte – a questi limiti.

L’idea di creare due *cerniere* – una per i lavoratori e una per le imprese – tra i due archivi amministrativi consente una maggiore snellezza nell’utilizzo dei dati. La struttura del link permette di accedere direttamente alle informazioni di interesse, escludendo tutto il resto. In questo modo, la massa critica di dati da elaborare si riduce notevolmente, abbassando di conseguenza i tempi di elaborazione statistica.

La complementarità dei due archivi, dall’altro lato, consente invece di ovviare al problema dello scarso numero di informazioni. Il link consente, in primo luogo, un allungamento dell’orizzonte temporale coperto dai dati. I dati VWH coprono un periodo che va dal 1975 al 2001, mentre Giove copre il periodo dal 1998 al 2006; di conseguenza il link consente di mettere assieme i due periodi: 31 anni di storie lavorative e aziendali.

Per quanto riguarda, invece, le variabili disponibili, il link consente di ampliarne la quantità. In particolare, alle informazioni già presenti in Giove, con particolare riferimento a quelle relative alle imprese, si aggiungono quelle relative alla data di creazione e cessazione dell’attività di impresa e la dimensione aziendale. Rispetto ai lavoratori, invece, si può accedere ad informazioni relative ai salari e ai periodi contributivi.

A fronte della maggiore ricchezza informativa venutasi a creare dall’unione dei due archivi amministrativi è necessario avere qualche cautela nell’utilizzo dei dati. Principalmente, è necessario fare attenzione all’utilizzo dei dati relativi ai lavoratori stranieri²³. Il problema principale consiste nel fatto che, in molti casi, questi lavoratori hanno la stessa data (convenzionale) di nascita. La procedura di link così come è stata concepita, considerando che in VWH non sono presenti codici fiscali dei lavoratori, ha portato, per questi soggetti, ad un match spurio²⁴.

5. Chi sono gli esclusi?

La procedura di *linkage* tra i due archivi amministrativi ha portato, oltre che alla creazione di un database di raccordo tra i due, anche all’individuazione di un gruppo di imprese e lavoratori per i quali non è stato possibile effettuare il match.

²³ I lavoratori stranieri sono facilmente identificabili in Giove grazie al codice fiscale. In particolare, per i lavoratori stranieri il codice relativo al comune di nascita inizia con Z (ad esempioZ210.).

²⁴ Per questi soggetti sarebbe necessario un ulteriore controllo utilizzando informazioni sul comune di residenza e gli spell lavorativi.

Al fine di avere una piena visione e comprensione del database creato e della sua correttezza, è opportuno analizzare brevemente le caratteristiche di questo gruppo di lavoratori e imprese, in modo da poter individuare le cause di esclusione dalla procedura di link e la sua attendibilità. In particolare, tale analisi è stata svolta, prima sull'intero periodo e, successivamente solo per imprese e lavoratori presenti al 2001.

5.1. Dal 1975 al 2006

La prima parte dell'analisi degli 'esclusi' considera l'intero periodo coperto dai due data set. Alcune ragioni sottostanti l'esclusione dal link di lavoratori e aziende dipendono fortemente dal diverso periodo coperto dai due archivi. Infatti, è ragionevole pensare che non troveranno corrispondenze in VWH: aziende presenti in Giove create dopo il 2001 e i lavoratori presenti in Giove entrati dopo il 2001. Nella stessa maniera, non troveranno corrispondenza in Giove le aziende presenti in VWH cessate prima del 1998 e i lavoratori fuoriusciti prima del 1998. In questa sezione, verrà analizzato il peso di questi soggetti nel gruppo degli esclusi e si cercherà di individuare altre possibili motivi di esclusione.

5.1.1. Le imprese

Come già accennato in precedenza, 'zip' aziende è composto da 341.778 aziende che corrispondono al 70% delle aziende presenti in Giove e il 30% circa delle aziende presenti in VWH.

In primo luogo, analizziamo le caratteristiche delle 143.204 aziende che rimangono escluse dal link. Dato il tipo di informazioni fornite da Giove, tali aziende possono essere classificate principalmente ed esclusivamente sulla base di due elementi: collocazione geografica e tipologia di attività economica.

Tab. 1: Aziende (Giove) 'escluse' dal link per provincia e peso % sul totale delle imprese.

Province	N. aziende mancanti	N. aziende totale	Peso %
Belluno	5.315	20.103	26,44%
Padova	24.653	77.022	32,01%
Rovigo	8.166	22.243	36,71%
Treviso	22.274	74.399	29,94%
Venezia	20.394	74.902	27,23%
Verona	22.537	71.079	31,71%
Vicenza	17.793	68.881	25,83%
No Veneto	22.049	74.738	29,50%
Dato mancante	23	126	18,25%
Totale	143.204	483.493	29,62%

Fonte: ns. elaborazioni su Giove-Sirlv.

La distribuzione per provincia delle aziende escluse dal link (**tab. 1**) non evidenzia una particolare caratterizzazione geografica. In media, per ogni provincia è rimasto escluso il 30% delle unità locali, con percentuali leggermente superiori a Padova (32%), Rovigo (37%) e Verona (32%).

La distribuzione delle aziende per attività economica (**tab. 2**) è, rispetto a quella per provincia, più caratterizzante. Appare evidente che per quanto riguarda le attività agricole, di pesca e estrattive (65%), le attività 'Altri servizi' (62%) e, seppur in misura minore, per il settore costruzioni (31%) si registrano le percentuali più alte di imprese escluse dal link. Questi risultati sono spiegabili dal fatto che nell'archivio aziendale VWH sono parzialmente

escluse le imprese che operano nel settore agricolo e in altri servizi afferenti alla pubblica amministrazione, o più in generale le imprese che, per quanto riguarda le pratiche contributive, fanno riferimento a casse di previdenza diverse da VWH.

Tab. 2: Aziende (Giove) 'escluse' dal link per attività economica e peso % sul totale delle imprese.

Attività	N. aziende mancanti	N. aziende totali	Peso %
Agric., pesca estrattive	14.446	22.338	64,67%
Carta, poligrafica	848	4.934	17,19%
Chimica, gomma	1.040	6.196	16,79%
Ind. alimentare	1.626	7.582	21,45%
Ind. metalmeccanica	9.181	50.262	18,27%
Ind. mezzi di trasporto	422	1.854	22,76%
Legno mobilio	2.288	13.481	16,97%
Settore moda (tessile - abbigl. - calzature)	6.226	26.128	23,83%
Min. non metall.	1.064	6.074	17,52%
Altre manifatturiere	634	3.483	18,20%
Costruzioni	17.804	57.052	31,21%
Gas, acqua, energia elett.	165	787	20,97%
Commercio	18.578	80.653	23,03%
Alberghi, ristorazione	15.096	52.775	28,60%
Credito e assicurazione	1.529	9.190	16,64%
Trasporti e comunicazione	4.233	18.917	22,38%
Servizi alle imprese	12.335	47.901	25,75%
Altri servizi	30.994	50.209	61,73%
Pubblica amministrazione	4.065	20.093	20,23%
Dato mancante	630	3.584	17,58%
Totale	143.204	483.493	29,62%

Fonte: ns. elaborazioni su Giove-Sirly.

Le informazioni fornite da Giove in merito alle aziende non consentono, tuttavia, di individuare caratteristiche ed elementi precisi che abbiano determinato l'impossibilità di trovare per alcune aziende la loro corrispondente in VWH. Si può intuire che – nonostante l'impossibilità di una verifica precisa – parte di queste aziende sia stata creata dopo il 2001: ultimo anno utile per VWH. In altre parole, è probabile che la data di creazione successiva al 2001 e l'appartenenza ad alcuni settori economici non contemplati nell'archivio VWH possano essere le principali ragioni alla base dell'esclusione di queste 143.204 aziende.

Guardando al 70% (897.401) di imprese presenti in VWH che non trovano una loro corrispondente in Giove, si può procedere in primo luogo all'analisi della data di cessazione dell'attività d'impresa.

Partendo da questa caratteristica, infatti, si è proceduto alla verifica della percentuale di imprese che hanno cessato la loro attività prima del 1998 (anno di inizio dell'archivio amministrativo Giove) e che, verosimilmente non possono trovare una loro azienda corrispondente nei dati dei Centri per l'Impiego.

La distribuzione delle aziende per anno di cessazione dell'attività (**tab. 3**) mostra con chiarezza come gran parte delle aziende mancanti (76%) abbia cessato l'attività prima del 1998. Per il restante 25% è opportuna una verifica della distribuzione geografica e del tipo di attività economica.

Come già fatto per Giove, quindi, viene analizzata la distribuzione geografica delle unità locali di VWH (**tab. 4**).

La quasi totalità (84%) delle imprese escluse dal link è collocata fuori dal Veneto. Questo probabilmente è dovuto al fatto che in VWH, per lungo tempo, l'iscrizione avveniva sulla base della sede VWH del lavoratore e non dell'impresa. In questo modo, è possibile seguire i lavoratori veneti anche quando vanno a lavorare al di fuori della regione, mentre in Giove, essendo il Cpi l'unità di riferimento, tale possibilità è più limitata.

Tab. 3: Aziende (VWH) 'escluse' dal link per anno di cessazione attività. Valori assoluti e percentuali.

Anno cessazione	N. aziende mancanti	Valori %
< 1998	685.046	76,34%
1975 e precedenti	27.420	3,06%
1976	24.579	2,74%
1977	23.661	2,64%
1978	23.112	2,58%
1979	24.369	2,72%
1980	29.862	3,33%
1981	29.716	3,31%
1982	29.596	3,30%
1983	28.075	3,13%
1984	26.610	2,97%
1985	30.174	3,36%
1986	27.103	3,02%
1987	28.600	3,19%
1988	31.754	3,54%
1989	34.173	3,81%
1990	36.934	4,12%
1991	36.738	4,09%
1992	37.910	4,22%
1993	33.650	3,75%
1994	29.712	3,31%
1995	30.614	3,41%
1996	31.283	3,49%
1997	29.401	3,28%
1998	28.597	3,19%
1999	30.654	3,42%
2000	36.479	4,06%
2001 e successivi	116.625	13,00%
Totale	897.401	100,00%

Fonte: ns. elaborazioni su dati VWH.

Per concludere l'analisi sulle imprese, nella **tab. 5** viene presentata la distribuzione delle aziende mancanti per settore di attività economica.

Differentemente rispetto alle aziende mancanti in Giove, nel caso di VWH non c'è una specifica caratterizzazione settoriale; percentuali superiori alla media si riscontrano solamente per il settore costruzioni (65%), gas-elettricità-acqua (62%) e per la pubblica amministrazione (62%).

Più in generale, bisogna tener conto del fatto che Giove, come accennato nei paragrafi precedenti, non copre l'intero universo delle imprese (unità locali). In particolare, se un'azienda non opera attività di attivazione e/o cessazione di rapporti di lavoro non viene registrata nell'archivio aziendale²⁵.

²⁵ Cfr. Maurizio (2006).

Tab. 4: Aziende (VWH) 'escluse' dal link cessate dopo il 1998 per provincia e peso % sul totale delle imprese cessate dopo il 1998.

Province	N. aziende mancanti	N. aziende totale	Peso %
Belluno	330	7.535	4,38%
Padova	2.689	31.200	8,62%
Rovigo	680	8.109	8,39%
Treviso	1.426	29.307	4,87%
Venezia	2.545	29.566	8,61%
Verona	2.313	30.368	7,62%
Vicenza	1.579	29.943	5,27%
No Veneto	200.787	238.595	84,15%
Dato mancante	6	13	46,15%
Totale	212.355	404.636	52,48%

Fonte: ns. elaborazioni su dati VWH.

Tab. 5: Aziende (VWH) 'escluse' dal link cessate dopo il 1998 per attività economica e peso % sul totale delle imprese cessate dopo il 1998.

Attività	N. aziende mancanti	N. aziende totali	Peso %
Agric., pesca estrattive	14.384	26.466	54,35%
Carta, poligrafica	2.570	5.112	50,27%
Chimica, gomma	3.531	6.587	53,61%
Ind. alimentare	6.162	11.458	53,78%
Ind. metalmeccanica	18.240	37.747	48,32%
Ind. mezzi di trasporto	1.158	2.040	56,76%
Legno mobilio	4.255	11.430	37,23%
Settore moda (tessile - abbigl. - calzature)	9.799	21.602	45,36%
Min. non metall.	2.792	5.600	49,86%
Altre manifatturiere	12.520	26.396	47,43%
Costruzioni	33.724	51.886	65,00%
Gas, acqua, energia elett.	235	379	62,01%
Commercio	28.601	60.915	46,95%
Alberghi, ristorazione	20.631	35.612	57,93%
Credito e assicurazione	5.451	12.052	45,23%
Trasporti e comunicazione	9.315	16.034	58,10%
Servizi alle imprese	17.890	35.822	49,94%
Altri servizi	7.823	15.994	48,91%
Pubblica amministrazione	13.249	21.371	62,00%
Dato mancante	25	133	18,80%
Totale	212.355	404.636	52,48%

Fonte: ns. elaborazioni su dati VWH.

5.1.2. I lavoratori

La tabella 'zip' lavoratori contiene 1.396.674 lavoratori che corrispondono a circa il 58% degli individui registrati in Giove (69% di quelli occupati nelle aziende linkate) e al 38% circa di quelli presenti in VWH.

Al fine di determinare le cause sottostanti l'esclusione di alcuni lavoratori dal link, va considerato anche in questo caso il diverso orizzonte temporale dei due data set. Più specificatamente, è plausibile pensare che i lavoratori entrati per la prima volta in Giove successivamente al 2001 non siano rintracciabili nel database VWH. Allo stesso modo, i

lavoratori fuoriusciti “definitivamente” da VWH prima del 1998 non sono rintracciabili in Giove.

Tab. 6: *Lavoratori (VWH) ‘esclusi’ dal link per ultimo anno di rilevazione. Valori assoluti e percentuali.*

Anno uscita	N. lavoratori mancanti	Valori %
<1998	1.467.591	66,24%
1975	56.138	2,53%
1976	65.299	2,95%
1977	67.547	3,05%
1978	60.634	2,74%
1979	53.622	2,42%
1980	62.458	2,82%
1981	68.720	3,10%
1982	74.004	3,34%
1983	70.317	3,17%
1984	65.487	2,96%
1985	59.674	2,69%
1986	72.336	3,26%
1987	81.279	3,67%
1988	63.968	2,89%
1989	62.347	2,81%
1990	63.678	2,87%
1991	71.969	3,25%
1992	72.916	3,29%
1993	58.861	2,66%
1994	63.128	2,85%
1995	49.952	2,25%
1996	47.530	2,15%
1997	55.727	2,52%
1998	92.368	4,17%
1999	43.854	1,98%
2000	49.766	2,25%
2001	562.070	25,37%
Totale	2.215.649	100,00%

Fonte: ns. elaborazioni su dati VWH.

La **tab. 6** mostra che per il 66% circa dei lavoratori VWH l’esclusione dal link è attribuibile al fatto di essere usciti definitivamente dal database prima del 1998.

Nella stessa maniera, circa il 67% dei lavoratori registrati in Giove che non hanno trovato un loro corrispondente in VWH, sono apparsi la prima volta nel database dopo il 2001 (**tab.7**).

Per il rimanente 33-34% dei lavoratori di Giove e VWH, la cui esclusione non trova motivazione nell’anno di entrata e/o uscita dal rispettivo database, è opportuno dare uno sguardo ad alcune caratteristiche anagrafiche: provincia di nascita e di residenza²⁶.

²⁶ Sono state scelte queste due caratteristiche perché plausibilmente legate a differenze di iscrizione nei database; cosa che non si può ragionevolmente presumere per altre caratteristiche come età e sesso.

Tab. 7: *Lavoratori (Giove) 'esclusi' dal link per anno di ingresso. Valori assoluti e percentuali.*

Anno entrata	N. lavoratori mancanti	Valori %
>2001	493.551	67,46%
<=1975	17.061	2,33%
1976	567	0,08%
1977	458	0,06%
1978	467	0,06%
1979	604	0,08%
1980	11.974	1,64%
1981	801	0,11%
1982	554	0,08%
1983	527	0,07%
1984	526	0,07%
1985	853	0,12%
1986	796	0,11%
1987	1.043	0,14%
1988	1.435	0,20%
1989	2.050	0,28%
1990	5.209	0,71%
1991	8.643	1,18%
1992	9.681	1,32%
1993	7.904	1,08%
1994	9.293	1,27%
1995	12.149	1,66%
1996	13.575	1,86%
1997	16.337	2,23%
1998	20.181	2,76%
1999	24.100	3,29%
2000	31.468	4,30%
2001	39.733	5,43%
2002	139.028	19,00%
2003	123.305	16,85%
2004	109.977	15,03%
2005	87.278	11,93%
2006	33.963	4,64%
Dato mancante	67	0,01%
Totale	731.607	100,00%

Fonte: ns. elaborazioni su Giove-Sirlv.

Guardando ai lavoratori Giove, il cui primo accesso è precedente al 2001, e alla loro distribuzione per provincia di nascita (**tab. 8**), si può notare che la maggior parte di essi è nata al di fuori del Veneto (44%); il 20% circa, nasce rispettivamente nelle province di Venezia (11%) e Verona (10%). È importante ricordare che nella quota di lavoratori nati al di fuori del Veneto rientrano anche i lavoratori stranieri, per i quali, come già accennato nel paragrafo 4,

il link è ancora spurio ed è probabile che alcuni lavoratori siano stati esclusi pur avendo un loro corrispondente in VWH. Per quanto riguarda le province venete, le percentuali più alte riscontrate a Venezia e Verona, e in misura minore a Treviso e Padova, sembrano rispecchiare la distribuzione territoriale dei lavoratori più che dare un'indicazione di qualche problematica di rilevazione in VWH.

Tab. 8: *Lavoratori (Giove) 'esclusi' dal link con anno di entrata precedente il 2001 per provincia di nascita e residenza. Valori assoluti, % e peso % sul totale dei lavoratori entrati prima del 2001.*

Province	N. lavoratori mancanti	Peso %
<i>Provincia di nascita</i>		
Belluno	8.078	3,39%
Padova	21.650	9,09%
Rovigo	13.412	5,63%
Treviso	22.512	9,46%
Venezia	25.359	10,65%
Verona	24.208	10,17%
Vicenza	18.518	7,78%
No Veneto	104.176	43,76%
Dato mancante	138	0,06%
Totale	238.051	100,00%
<i>Provincia di residenza</i>		
Belluno	9.523	4,00%
Padova	21.691	9,11%
Rovigo	13.944	5,86%
Treviso	25.081	10,54%
Venezia	25.286	10,62%
Verona	26.878	11,29%
Vicenza	16.305	6,85%
No Veneto	8.857	3,72%
Dato mancante	90.486	38,01%
Totale	238.051	100,00%

Fonte: ns. elaborazioni su Giove-Sirlv.

Per quanto riguarda il comune di residenza, si può notare che per il 38% dei lavoratori esclusi dal link tale dato è mancante. In questo caso, la componente extra-veneto è assolutamente esigua, mentre la distribuzione tra le province venete è simile a quella emersa per i comuni di nascita.

Per quanto riguarda invece i lavoratori VWH esclusi e fuoriusciti successivamente al 1998 (**tab. 9**), si può notare una netta concentrazione su lavoratori nati e/o residenti fuori dal Veneto (rispettivamente 52% e 48%). Tale dato può essere spiegato sia per i già menzionati problemi relativi al match spurio dei lavoratori stranieri, sia per la struttura di rilevazione di VWH che permette di seguire i lavoratori anche se occupati al di fuori del Veneto.

Anche in questo caso, la distribuzione per provincia – sia di nascita che di residenza – non presenta particolari caratterizzazioni.

Tab. 9: *Lavoratori (VWH) 'esclusi' dal link fuoriusciti dopo il 1998 per provincia di nascita e residenza. Valori assoluti, % e peso % sul totale dei lavoratori fuoriusciti dopo il 1998.*

Province	N. lavoratori mancanti	Peso %
<i>Provincia di nascita</i>		
Belluno	11.117	1,49%
Padova	77.809	10,40%
Rovigo	21.342	2,85%
Treviso	44.970	6,01%
Venezia	70.657	9,45%
Verona	67.564	9,03%
Vicenza	50.758	6,79%
No Veneto	387.877	51,85%
Dato mancante	15.964	2,13%
Totale	748.058	100,00%
<i>Provincia di residenza</i>		
Belluno	11.732	1,57%
Padova	88.806	11,87%
Rovigo	17.904	2,39%
Treviso	53.576	7,16%
Venezia	74.163	9,91%
Verona	78.716	10,52%
Vicenza	57.375	7,67%
No Veneto	358.840	47,97%
Dato mancante	6.946	0,93%
Totale	748.058	100,00%

Fonte: ns. elaborazioni su dati VWH.

5.2. Il 2001

L'analisi del 2001 – anno di intersezione tra i due data set – permette di valutare, al netto delle evidenti ragioni di esclusione sottolineate in partenza, la qualità del match ottenuto e di individuare ancora una volta possibili specificità caratterizzanti le aziende e i lavoratori esclusi dal match.

5.2.1. Le imprese

Per il 2001, in Giove sono presenti 186.171 aziende e quelle escluse dal link sono 16.489 (8,9% circa). In VWH, invece, sono presenti 270.441 e 116.625 sono quelle 'mancanti', corrispondenti a circa il 43%.

Seguendo la linea di analisi proposta nel paragrafo precedente, le aziende escluse dal link sono state analizzate sulla base della distribuzione per provincia e per attività economica.

La distribuzione per provincia relativa alle aziende VWH presenti al 2001 e escluse dal link (**tab. 10**) non evidenzia una particolare caratterizzazione territoriale con riferimento alle 7 province venete. Tuttavia, appare evidente che la grande maggioranza delle imprese che non trovano una loro corrispondente in Giove siano localizzate fuori dal Veneto (94%).

Spostando lo sguardo su Giove (**tab. 11**), invece, la componente extra-Veneto assume una rilevanza minore, mentre si registrano percentuali maggiori di esclusione per le province di Verona (20%) , Padova (18%) e, se si considera il peso percentuale sul totale delle aziende presenti al 2001, Rovigo (15%).

Tab. 10: Aziende (VWH) 'escluse' dal link presenti al 2001 per provincia, % e peso % sul totale delle aziende presenti al 2001.

Provincie	Aziende mancanti	%	Aziende totali	Peso % su tot
Belluno	181	0,16%	5.923	3,06%
Padova	1.522	1,31%	24.860	6,12%
Rovigo	425	0,36%	6.264	6,78%
Treviso	847	0,73%	23.634	3,58%
Venezia	1.603	1,37%	23.361	6,86%
Verona	1.317	1,13%	24.226	5,44%
Vicenza	1.020	0,87%	24.462	4,17%
Mancanti	5	0,00%	11	45,45%
No Veneto	109.705	94,07%	137.700	79,67%
Totale	116.625	100,00%	270.441	43,12%

Fonte: ns. elaborazioni su dati VWH.

Tab. 11: Aziende (Giove) 'escluse' dal link presenti al 2001 per provincia, % e peso % sul totale delle aziende presenti al 2001.

Provincie	Aziende mancanti	%	Aziende totali	Peso % su tot
Belluno	484	2,94%	7.468	6,48%
Padova	3.022	18,33%	33.431	9,04%
Rovigo	1.236	7,50%	8.088	15,28%
Treviso	2.360	14,31%	29.899	7,89%
Venezia	1.917	11,63%	29.084	6,59%
Verona	3.330	20,20%	30.524	10,91%
Vicenza	1.640	9,95%	29.136	5,63%
Mancanti	0	0,00%	13	0,00%
No Veneto	2.500	15,16%	18.528	13,49%
Totale	16.489	100,00%	186.171	8,86%

Fonte: ns. elaborazioni su Giove-Sirlv.

L'analisi per settori delle aziende 'escluse' di VWH evidenzia alcune particolarità (**tab. 12**). Guardando alla distribuzione percentuale per settore, circa il 40% delle aziende si concentra nei settori metalmeccanico (15%), delle costruzioni (15%) e del commercio.

Per quanto riguarda invece, il peso percentuale sul totale delle imprese, non si evidenziano particolari concentrazioni rispetto a specifici settori, tranne per il settore edile (55%), Gas-acqua-energia elettrica (59%), trasporti e comunicazione (51%) e pubblica amministrazione (52%).

La distribuzione per attività economica delle imprese 'escluse' di Giove presenta invece una caratterizzazione più netta (**tab. 13**).

Sia guardando alla distribuzione percentuale che al peso sul totale delle imprese, emerge con chiarezza una concentrazione di imprese nel settore agricolo (27% and 56% rispettivamente) e 5n 'Altri servizi' (15% e 22% rispettivamente). La forte prevalenza del settore agricolo può essere ricondotta, come già descritto in precedenza, al fatto che in VWH questo tipo di aziende non vengono (parzialmente) rilevate perché, in termini pensionistici, fanno riferimento ad altri Istituti.

Tab. 12: Aziende (VWH) 'escluse' dal link presenti al 2001 per attività economica, % e peso % sul totale delle aziende presenti al 2001.

Attività	N. aziende mancanti	%	N. aziende totali	Peso %
Agric., pesca estrattive	7.881	6,76%	17.441	45,19%
Carta, poligrafica	1.483	1,27%	3.611	41,07%
Chimica, gomma	2.146	1,84%	4.734	45,33%
Ind. alimentare	3.125	2,68%	7.371	42,40%
Ind. metalmeccanica	17.467	14,98%	44.847	38,95%
Ind. mezzi di trasporto	677	0,58%	1.375	49,24%
Legno mobilio	2.283	1,96%	8.228	27,75%
Settore moda (tessile - abbigl. - calzature)	5.008	4,29%	13.919	35,98%
Min. non metall.	1.638	1,40%	3.969	41,27%
Altre manifatturiere	53	0,05%	135	39,26%
Costruzioni	16.866	14,46%	30.628	55,07%
Gas, acqua, energia elett.	183	0,16%	312	58,65%
Commercio	16.127	13,83%	42.001	38,40%
Alberghi, ristorazione	10.203	8,75%	21.313	47,87%
Credito e assicurazione	3.288	2,82%	8.596	38,25%
Trasporti e comunicazione	5.614	4,81%	11.004	51,02%
Servizi alle imprese	11.026	9,45%	25.781	42,77%
Altri servizi	4.093	3,51%	10.706	38,23%
Pubblica amministrazione	7.446	6,38%	14.360	51,85%
Dato mancante	18	0,02%	110	16,36%
Totale	116.625	100,00%	270.441	43,12%

Fonte: ns. elaborazioni su dati VWH.

Tab. 12: Aziende (Giove) 'escluse' dal link presenti al 2001 per attività economica, % e peso % sul totale delle aziende presenti al 2001.

Attività	N. aziende mancanti	%	N. aziende totali	Peso %
Agric., pesca estrattive	4.565	27,69%	8.030	56,85%
Carta, poligrafica	108	0,65%	2.424	4,46%
Chimica, gomma	125	0,76%	2.858	4,37%
Ind. alimentare	191	1,16%	3.501	5,46%
Ind. metalmeccanica	1.048	6,36%	22.962	4,56%
Ind. mezzi di trasporto	44	0,27%	782	5,63%
Legno mobilio	257	1,56%	6.643	3,87%
Settore moda (tessile - abbigl. - calzature)	739	4,48%	10.748	6,88%
Min. non metall.	135	0,82%	2.966	4,55%
Altre manifatturiere	55	0,33%	1.764	3,12%
Costruzioni	1.403	8,51%	19.836	7,07%
Gas, acqua, energia elett.	35	0,21%	371	9,43%
Commercio	1.653	10,02%	33.016	5,01%
Alberghi, ristorazione	991	6,01%	17.445	5,68%
Credito e assicurazione	317	1,92%	4.289	7,39%
Trasporti e comunicazione	397	2,41%	7.804	5,09%
Servizi alle imprese	1.059	6,42%	18.872	5,61%
Altri servizi	2.548	15,45%	11.440	22,27%
Pubblica amministrazione	663	4,02%	9.024	7,35%
Dato mancante	156	0,95%	1.396	11,17%
Totale	16.489	100,00%	186.171	8,86%

Fonte: ns. elaborazioni su Giove-Sirlv.

5.2.2. I lavoratori

Al 2001, in Giove sono presenti 916.672 lavoratori e quelli esclusi dal link sono 102.834 (11,22%). In VWH sono presenti, invece 1.648.069 lavoratori, dei quali gli esclusi sono 562.070: circa il 34%.

Seguendo l'impostazione di analisi adottata in precedenza, i lavoratori sono stati classificati principalmente in base alla provincia di nascita e di residenza.

Per quanto riguarda i lavoratori VWH 'esclusi' dal link appare evidente, guardando sia alla provincia di nascita che a quella di residenza, nonché sia in termini di distribuzione percentuale che peso sul totale delle imprese, che la maggioranza è composta prevalentemente da soggetti nati o residenti fuori dal Veneto (**tab. 14**): rispettivamente il 54 e il 50%.

Tab. 14: *Lavoratori (VWH) 'esclusi' dal link presenti al 2001 per provincia di nascita e residenza. Valori assoluti, percentuali e peso % sul totale dei lavoratori presenti al 2001.*

Caratteristiche	Lavoratori mancanti	%	Lavoratori totali	Peso %
Totale	562.070	100,00%	1.648.069	34,10%
<i>Provincia nascita</i>				
Belluno	7.173	1,28%	49.745	14,42%
Padova	56.231	10,00%	193.781	29,02%
Rovigo	15.453	2,75%	63.191	24,45%
Treviso	34.980	6,22%	179.760	19,46%
Venezia	51.689	9,20%	201.416	25,66%
Verona	48.511	8,63%	180.208	26,92%
Vicenza	39.324	7,00%	195.890	20,07%
No Veneto	303.551	54,01%	576.511	52,65%
Mancante	5.158	0,92%	7.567	68,16%
<i>Provincia di residenza</i>				
Belluno	7.058	1,26%	59.345	11,89%
Padova	61.894	11,01%	221.278	27,97%
Rovigo	12.384	2,20%	57.972	21,36%
Treviso	40.062	7,13%	228.706	17,52%
Venezia	51.805	9,22%	218.319	23,73%
Verona	54.013	9,61%	218.332	24,74%
Vicenza	43.207	7,69%	238.727	18,10%
No Veneto	285.224	50,75%	395.491	72,12%
Mancante	6.423	1,14%	9.899	64,89%

Fonte: ns. elaborazioni su dati VWH.

Per quanto riguarda i lavoratori Giove 'esclusi' dal link (**tab. 15**) la componente di lavoratori nati al di fuori del Veneto continua ad avere una certa rilevanza (46% in termini di distribuzione percentuale e 19% in termini di peso sul totale dei lavoratori). È invece meno significativa rispetto al comune di residenza (solo rispetto al peso sul totale dei lavoratori residenti fuori dal Veneto assume una percentuale del 28% circa).

Rispetto alle province venete, la caratterizzazione è meno evidente, come già visto per i lavoratori esclusi da VWH. In particolare, rispetto alla provincia di nascita si registra una piccola concentrazione per la provincia di Rovigo (14% in termini di peso % sul totale). Per quanto attiene, invece, alla provincia di residenza, in termini di distribuzione totale emergono Padova (13%), Treviso (11%) e Verona (14%). Considerando invece il peso degli esclusi sul totale dei lavoratori per provincia, si nota nuovamente una maggiore incidenza per la provincia di Rovigo (15%).

Tab. 15: *Lavoratori (Giove) 'esclusi' dal link presenti al 2001 per provincia di nascita e residenza. Valori assoluti, percentuali e peso % sul totale dei lavoratori presenti al 2001.*

Caratteristiche	Lavoratori mancanti	%	Lavoratori totali	Peso %
Totale	102.834	100,00%	916.672	11,22%
<i>Provincia nascita</i>				
Belluno	3.018	2,93%	36.162	8,35%
Padova	11.264	10,95%	120.881	9,32%
Rovigo	5.321	5,17%	37.712	14,11%
Treviso	8.745	8,50%	118.349	7,39%
Venezia	8.619	8,38%	116.199	7,42%
Verona	11.216	10,91%	113.184	9,91%
Vicenza	7.205	7,01%	122.543	5,88%
No Veneto	47.380	46,07%	251.452	18,84%
Mancante	66	0,06%	190	34,74%
<i>Provincia di residenza</i>				
Belluno	4.000	3,89%	45.683	8,76%
Padova	13.344	12,98%	126.045	10,59%
Rovigo	5.842	5,68%	38.753	15,07%
Treviso	11.651	11,33%	143.115	8,14%
Venezia	9.730	9,46%	128.599	7,57%
Verona	14.090	13,70%	126.345	11,15%
Vicenza	7.264	7,06%	129.039	5,63%
No Veneto	6.134	5,96%	21.282	28,82%
Mancante	30.779	29,93%	157.811	19,50%

Fonte: ns. elaborazioni su Giove-Sirlv.

6. Conclusioni

Dalla necessità di arricchire e aggiornare le fonti amministrative a nostra disposizione, mantenendo tuttavia la fruibilità e la facilità di utilizzo dei dati, è nato il link tra gli archivi VWH e Giove.

L'esperienza di *linkage* ha portato alla creazione di due data set '*cerniera*' – '*zip*' aziende e '*zip*' lavoratori – che contengono rispettivamente gli identificativi delle aziende e dei lavoratori risultanti dall'operazione di linkage. Tale struttura permette di accedere facilmente ai due archivi amministrativi e ottenere le informazioni di interesse.

In termini numerici, il risultato ottenuto è più che soddisfacente: la copertura per quanto riguarda le aziende è del 70% e del 58% (69% se si considerano i lavoratori occupati nelle aziende di Giove che hanno una corrispondente in VWH) per quanto riguarda i lavoratori.

Le analisi dei lavoratori e delle imprese esclusi dal link – sia sull'intero periodo che per il 2001, anno di sovrapposizione dei due data set – hanno evidenziato che la mancata corrispondenza è dovuta principalmente a cause imputabili alla diversa struttura dei due archivi, come il diverso orizzonte temporale o particolarità di rilevazione. Questi elementi rappresentano un'ulteriore conferma che la procedura attuata ha un buon margine di correttezza.

Rimangono tuttavia ancora alcuni elementi spuri, come ad esempio il match dei lavoratori stranieri, per i quali è necessaria un'ulteriore attività di pulizia e controllo.

Bibliografia

Anastasia B., Disarò M., Gambizza M., Maurizio D., Rasera M. (2006), “ ‘Giove’: il public use database prodotto da Veneto Lavoro a partire dai dati dei centri per l’impiego”, presentazione per *Convegno Istat-Cisis “I dati amministrativi per le statistiche sui mercati del lavoro locali”*, Roma.

Costanzo, L. (2008), “Gli archivi amministrativi delle province. Una risorsa per la statistica ufficiale”, presentazione per “*Progetto ‘Censimento degli archivi provinciali’ – Giornata di formazione*”, Roma.

Maurizio M. (2006), “Giove: un database statistico sul mercato del lavoro veneto. Costruzioni e indicazioni per l’utilizzo”, *Veneto Lavoro Serie I Tartufi*, No. 22/2006.

Tattara G., Valentini M. (2006), “Gli archivi anagrafici VWH relativi alle imprese ed ai lavoratori dipendenti”, poster session *Final workshop ‘The Evaluation of Labour Market, Welfare and Firms Incentives Programmes’*, Venezia.

Working Papers

78. D. Maurizio, E. Trevisan, Giove-VWH: Esperienza di *linkage* tra data base amministrativi, Marzo 2009
77. F. Bassi, A. Padoan, U. Trivellato, Inconsistencies in Reported Employment Characteristics among Employed Stayers: a Case-study with Two-wave Panels from the Italian Labour Force Survey, 1993-2003, Luglio 2008
76. D. Bondonio, Evaluating the Effects of Business Incentives Policies: A Discussion Note on Identification Strategies, Luglio 2008
75. B. Anastasia, D. Maurizio, Salari e stipendi in Veneto 2001-2002, Febbraio 2007
74. G. Tattara, M. Valentini, Labour market segmentation, flexibility and precariousness in the Italian North East, Aprile 2006
73. B. Anastasia, D. Maurizio, I lavoratori dipendenti in Veneto 1998-2003: profili e percorsi. Statistiche sistematiche da "Giove 2005", Dicembre 2005
72. A. Fossaluzza, Evaluation of the Effects of Two Classes of Active Labour Market Policies for Welfare Recipients: a Danish Study Case, Dicembre 2005
71. P. Cipollone, A. Guelfi, Financial support to permanent jobs. The Italian case, Novembre 2005
70. F. Bassi, U. Trivellato, The latent class approach to estimating gross flows affected by correlated classification errors, with application to data from the French Labour Force Survey, Novembre 2005
69. M. Fort, Education and the timing of births: evidence from a natural experiment in Italy, Novembre 2005
68. D. Contini, N. Negri, Would declining exit rates from welfare provide evidence of welfare dependence in homogeneous environments?, Ottobre 2005
67. E. Battistin, E. Rettore, U. Trivellato, Choosing among alternative classification criteria to measure the labour force state, Ottobre 2005
66. G. Tattara, M. Valentini, Evaluating the Italian training on the job contract (CFL), Novembre 2005
65. A. Paggiaro, E. Rettore, U. Trivellato, The impact of the Italian "Mobility Lists" on employment chances: new evidence from linked administrative archives, Ottobre 2005
64. D. Bondonio, R.T. Greenbaum, Do Tax Incentives Affect Local Economic Growth? What Mean Impacts Miss in the Analysis of Enterprise Zone Policies, Ottobre 2005
63. D. Bondonio, The employment impact of business incentive programs in declining areas. Mean impacts versus impacts by degrees of economic distress, Ottobre 2005
62. U. Trivellato, F. Bassi, A. Discenza, A. Giraldo, Transizioni e mobilità nel mercato del lavoro italiano, 1979-2003, Giugno 2005
61. A. De Angelini, A. Giraldo, La mobilità dei lavoratori nel Veneto. Confronto fra misure su dati RTFL e su dati NETLABOR, Settembre 2003
60. F. Bassi, E. Salvan, Dinamiche di ricollocamento per lavoratori che perdono un'occupazione stabile, Settembre 2003
59. M. Romano, Abbinamento tra gli archivi INPS e Netlabor: il caso degli iscritti alle liste di mobilità, Settembre 2003

58. P. Baretta, U. Trivellato, La mobilità dei lavoratori da fonti amministrative e da surveys sulle famiglie: un'analisi comparata, Settembre 2003
57. D. Bondonio, R.T. Greebaum, A comparative evaluation of spatially targeted economic revitalization programs in the European Union and the United States, Settembre 2003
56. F. Berton, L. Pacelli, G. Segre, Tra lavoro dipendente e lavoro parasubordinato: chi sono, da dove vengono e dove vanno i lavoratori parasubordinati, Settembre 2003
55. G. Betti, V. Verma, M. Natilli, F. Ballini, Statistical imputation in conjunction with micro-simulation of income data, Settembre 2003
54. V. Verma, G. Betti, F. Ballini, M. Natilli, S. Galgani, Personal income in the gross and net forms: applications of the Siena micro-simulation model (SM2), Settembre 2003
53. R. Canu, G. Tattara, Quando le farfalle mettono le ali. Osservazioni sull'ingresso delle donne nel lavoro dipendente, Settembre 2003
52. D. Favaro, S. Magrini, Gender wage differentials among young workers: methodological aspects and empirical results, Settembre 2003
51. F. Devicienti, Downward nominal wage rigidity in Italy: evidence and consequences, Novembre 2002
50. V. Verma, G. Betti, Longitudinal measures of income poverty and life-style deprivation, Novembre 2002.
49. A. Borgarello, F. Devicienti, Trends in the Italian earnings distribution, 1985-1996, Novembre 2002
48. S. Bragato, F. Occari, M. Valentini, I problemi di contabilizzazione dei lavoratori extracomunitari. Una verifica nelle province di Treviso e Vicenza, Novembre 2002
47. B. Anastasia, D. Maurizio, Misure dell'occupazione temporanea: consistenza, dinamica e caratteristiche di uno stock eterogeneo, Novembre 2002
46. E. Rettore, U. Trivellato, A. Martini, La valutazione delle politiche del lavoro in presenza di selezione: migliorare la teoria, i metodi o i dati?, Novembre 2002
45. U. Trivellato, A. Giraldo, Assessing the 'choosiness' of the job seekers. An exploratory approach and evidence for Italy, Novembre 2002
44. E. Battistin, E. Rettore, Another look at the regression discontinuity design, Novembre 2002
43. B. Contini, F. Cornaglia, C. Malpede, E. Rettore, Measuring the impact of the Italian CFL programme on the job opportunities for the youths, Novembre 2002
42. F. Devicienti, Estimating poverty persistence in Britain, Novembre 2001
41. A. Giraldo, E. Rettore, U. Trivellato, Attrition bias in the bank of Italy's survey on household income and wealth, Novembre 2001
40. A. Giraldo, E. Rettore, U. Trivellato, The persistence of poverty: true state dependence or unobserved heterogeneity? Some evidence from the Italian survey on household income and wealth, Novembre 2001
39. G. Barbieri, P. Gennari, P. Sestito, Do public employment services help people in finding a job? An evaluation of the Italian case, Novembre 2001
38. N. Torelli, A. Paggiaro, Estimating transition models with misclassification, Novembre 2001
37. B. Anastasia, M. Gambuzza, M. Rasera, Le sorti dei flussi: dimensioni della domanda di lavoro, modalità di ingresso e rischio disoccupazione dei lavoratori extracomunitari in Veneto, Novembre 2001

36. L. Pacelli, Fixed term contracts, social security rebates and labour demand in Italy, Novembre 2001
35. A. Borgarello, F. Deficienti, C. Villosio, Mobilità retributiva in Italia 1985-1996, Novembre 2001
34. B. Contini, C. Villosio, Job changes and wage dynamics, Novembre 2001
33. B. Contini, Earnings mobility and labor market segmentation in Europe and USA: preliminary explorations, Novembre 2001
32. A. Borgarello, F. Devicienti, Trend nella distribuzione dei salari italiani 1985-1996, Novembre 2001
31. F. Bassi, U. Trivellato, Gross flows from the French labour force survey: a reanalysis, Novembre 2001
30. F. Bassi, M. Gambuzza, M. Rasera, Imprese e contratti di assunzione: prime analisi da Netlabor, Novembre 2001
29. N. Rosati, Further results on inequality in Italy in the 1980s and the 1990s, Aprile 2001
28. A. Ichino, M. Polo, E. Rettore, Are judges biased by labor market conditions?, Novembre 2000
27. E. Battistin, E. Rettore, Testing for the presence of a programme effect in a regression discontinuity design with non compliance, Novembre 2000
26. E. Battistin, E. Rettore, U. Trivellato, Measuring participation at work in the presence of fallible indicators of labour force state, Giugno 2000
25. A. D'Agostino, G. Ghellini, L. Neri, Percorsi di ingresso dei giovani nel mercato del lavoro, Giugno 2000
24. F. Bassi, M. Gambuzza, M. Rasera, E. Rettore, L'ingresso dei giovani nel mercato del lavoro: prime esplorazioni dall'archivio Netlabor, Giugno 2000
23. A. Paggiaro, U. Trivellato, Assessing the effects of the "Mobility List" programme in an Italian region: do (slightly) better data and more flexible models matter?, Marzo 2000
22. A. D'Agostino, G. Ghellini, L. Neri, Modelli statistici per l'analisi dei comportamenti di transizione scuola lavoro, Marzo 2000
21. G. Betti, B. Cheli, A. Lemmi, Analisi delle dinamiche di povertà e disoccupazione su uno pseudo panel italiano, Marzo 2000
20. N. Rosati, Permanent and Temporary Inequality in Italy in the 1980s and 1990s, Marzo 2000
19. F. Bassi, M. Gambuzza, M. Rasera, Struttura e qualità delle informazioni del sistema NETLABOR. Una verifica sui dati delle Scica delle province di Belluno e Treviso, Marzo 2000
18. B. Anastasia, M. Gambuzza, M. Rasera, La durata dei rapporti di lavoro: evidenze da alcuni mercati locali del lavoro veneti, Marzo 2000
17. G. Betti, B. Cheli, A. Lemmi, Occupazione e condizioni di vita su uno pseudo panel italiano: primi risultati, avanzamenti e proposte metodologiche, Ottobre 1999
16. A. D'Agostino, G. Ghellini, L. Neri, A Multiple Imputation Method for School to Work Panel Data, Ottobre 1999
15. A. Paggiaro, N. Torelli, Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro, Ottobre 1999
14. S. Campostrini, A. Giraldo, N. Parise, U. Trivellato, La misura della partecipazione al lavoro in Italia: presupposti e problemi metodologici di un approccio "time use", Ottobre 1999

13. T. Di Fonzo, P. Gennari, Le serie storiche delle forze di lavoro per il periodo 1984.1-92.3: prospettive e problemi di ricostruzione, Giugno 1999
12. A. Paggiaro, Un modello di mistura per l'analisi della disoccupazione di lunga durata, Maggio 1999
11. M. Lalla, F. Pattarin, Alcuni modelli per l'analisi delle durate complete e incomplete della disoccupazione: il caso Emilia Romagna, Maggio 1999
10. F. Bassi, M. Gambuzza, M. Rasera, Il sistema informatizzato NETLABOR. Caratteristiche di una nuova fonte sul mercato del lavoro, Maggio 1999
9. B. Contini, M. Filippi, L. Pacelli, C. Villosio, Working careers of skilled vs. unskilled workers, Gennaio 1999
8. B. Contini, L. Pacelli, C. Villosio, Short employment spell in Italy, Germany and Great Britain: testing the "Port-of-entry" hypothesis, Gennaio 1999
7. B. Contini, Wage structures in Europe and in the USA: are they rigid, are they flexible?, Gennaio 1999
6. M. Forni, S. Paba, Industrial districts, social environment and local growth. Evidence from Italy, Novembre 1998
5. Ginzburg, M. Scaltriti, G. Solinas, R. Zoboli, Un nuovo autunno caldo nel Mezzogiorno? Note in margine al dibattito sui differenziali salariali territoriali, Ottobre 1998
4. F. Bassi, Un modello per la stima di flussi nel mercato del lavoro affetti da errori di classificazione in rilevazioni retrospettive, Ottobre 1998
3. U. Trivellato, Il monitoraggio della povertà e della sua dinamica: questioni di misura e evidenze empiriche, Settembre 1998
2. N. Rosati, E. Rettore, G. Masarotto, A lower bound on asymptotic variance of repeated cross-sections estimators in fixed-effects models, Agosto 1998
1. E. Battistin, A. Gavosto, E. Rettore, Why do subsidized firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy, Agosto 1998