

LUOGHI E SEGNALI LAVORATIVI DA ARCHIVI AMMINISTRATIVI PER LA
PREVISIONE DEL TRASFERIMENTO DI RESIDENZA: UNA SPERIMENTAZIONE SUI
DATI DELL'EMILIA-ROMAGNA TRA IL 2011 E IL 2012

Simona Toti¹, Romina Filippini²

SOMMARIO

Il trasferimento di residenza può essere legato a motivazioni diverse: familiari, fiscali, lavorative e soggettive. L'obiettivo del presente lavoro è di valutare l'utilizzo dell'informazione d'archivio sull'attività lavorativa al fine di isolare l'effetto del luogo di lavoro nel prevedere il trasferimento di residenza. La popolazione di riferimento è quella degli individui con almeno un contratto di lavoro dipendente presente negli archivi amministrativi nel 2012 in Emilia Romagna. L'attività lavorativa viene riferita all'anno attraverso una sequenza di 12 indicatori dello stato di attività del soggetto (attivo/non attivo) in ciascun mese, in un dato comune. Le analisi dell'informazione lavorativa (sequenza e luoghi di lavoro e di residenza) permettono di identificare gli eventi potenzialmente determinanti il trasferimento di residenza: l'inizio di un lavoro in un comune diverso dalla residenza o il cambio del comune di lavoro. Sul sottoinsieme di individui che sperimentano questi eventi viene condotta un'analisi di regressione logistica che stima la probabilità del trasferimento in relazione alle variabili di soggetto, profilo e comune precedentemente descritte, identificando anche il potenziale comune di destinazione.

¹ Istat, via Cesare Balbo 16, 00184, Roma, toti@istat.it.

² Istat, via Cesare Balbo 16, 00184, Roma, filippini@istat.it (corresponding author).

1. Introduzione

Data la crescente disponibilità di informazioni provenienti da archivi amministrativi, l'Istat ha sviluppato un Sistema Integrato di Microdati (SIM) a partire dal quale è possibile costruire database tematici per far fronte a specifiche esigenze di ricerca (Ambroselli 2014). In particolare è stato costruito un database per lo studio della popolazione abitualmente dimorante (AIDA) che ha come popolazione di riferimento gli individui presenti in Italia che abbiano almeno un segnale mensile di lavoro o studio in un dato comune nell'anno 2012 (Chieppa *et al.* 2018, Borrelli *et al.* 2018)

Le variabili che si vogliono studiare in relazione al cambio di residenza tra il 2011 e il 2012 sono quelle relative al profilo lavorativo (sequenza di attività e tipologia lavorativa), alle caratteristiche dei comuni di lavoro e del comune di residenza (dimensione, sistema locale del lavoro e distanza tra lavoro e residenza) e alle caratteristiche socio demografiche degli individui (sesso, età e cittadinanza).

Per quanto riguarda l'attività lavorativa, l'informazione è presente con 12 variabili dicotomiche 0/1 che indicano lo stato di attività del soggetto in ciascun mese dell'anno in un dato comune. Per poter cogliere l'informazione complessiva legata alla sequenza lavorativa, nel paragrafo 3 vengono individuate quattro variabili numeriche capaci di sintetizzare le 12 variabili mensili e di descrivere l'intensità e la frequenza dell'attività nell'anno. Nel paragrafo 4 si studiano le informazioni sui luoghi di lavoro mettendole in relazione con la residenza al fine di individuarne l'associazione con il fenomeno del trasferimento. Le analisi sull'informazione lavorativa (sequenze e luoghi) permettono di identificare l'evento potenzialmente determinante il trasferimento e di selezionare un sottoinsieme di individui maggiormente esposti a sperimentare il cambio di residenza (paragrafo 5). Su questa sottopopolazione viene condotta l'analisi di regressione logistica che stima la probabilità del trasferimento in relazione alle variabili di soggetto, profilo e comune precedentemente descritte (paragrafo 6).

2. La base dati

Per ciascun individuo sono indicati i comuni cui sono riferiti i segnali di studio/lavoro nell'anno e per ogni comune è riportato il segnale mensile di presenza di almeno un contratto attivo, indicato da 1, o nessuno contratto attivo, indicato da 0. L'iscrizione a scuola o all'università in un certo anno scolastico implica la valorizzazione a 1 della stessa variabile nei mesi che vanno da settembre ad agosto. Tale variabile si assume indicativa della presenza dell'individuo nel comune.

In questa fase sperimentale è stata acquisita la sola base relativa alla regione Emilia Romagna, costituita da 2.797.587 individui. Il numero di individui con informazioni complete sui luoghi di residenza e sui luoghi di studio/lavoro scende a 2.494.132.

Partendo dall'osservazione che in taluni casi il segnale mensile è fittizio, in quanto alcune fonti amministrative indicano solo la presenza annuale dell'individuo, si è scelto di considerare solo quegli individui per cui il segnale di presenza nelle fonti amministrative fosse effettivamente un segnale mensile. Ciò equivale a selezionare gli individui con contratto di lavoro dipendente poiché per i lavoratori autonomi è disponibile solo l'informazione relativa all'apertura o meno della partita iva e non l'effettiva attività lavorativa, e per gli studenti è noto solo l'anno scolastico di iscrizione ma non l'effettiva frequenza. Tale operazione riduce ulteriormente la base dati utilizzabile che passa così a 1.740.164 individui con contratto di lavoro dipendente (escludendo autonomi e studenti).

Le analisi successive si concentrano sui dati relativi al 2012: gli individui selezionati (con almeno un segnale lavorativo nel 2012) sono pari a 1.656.516.

3. L'informazione sull'attività lavorativa mensile

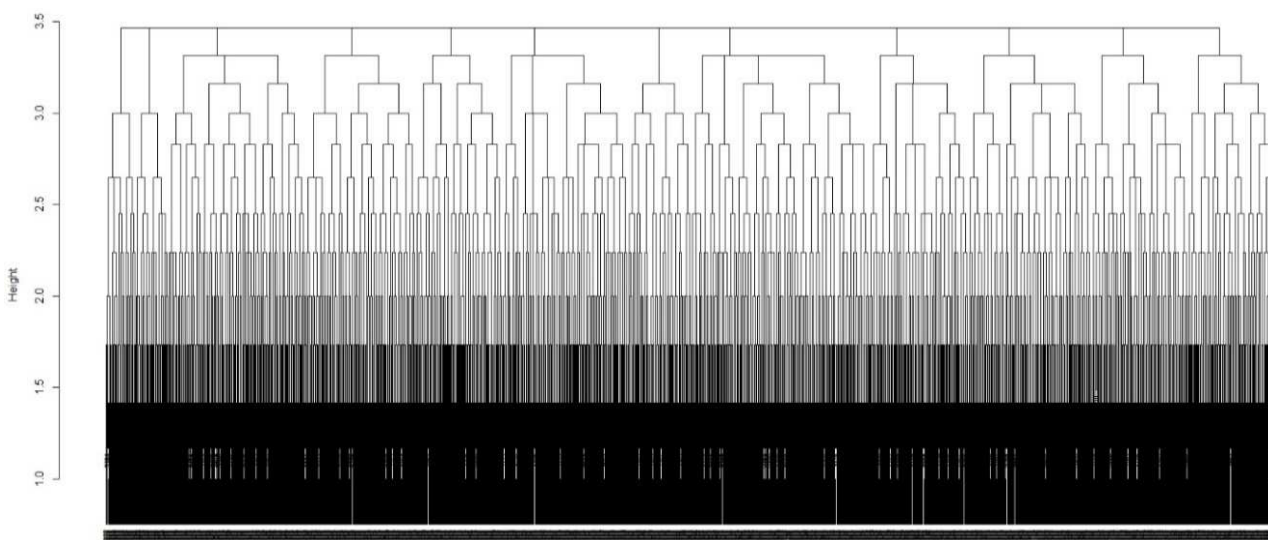
Per valutare la forza e tipologia del segnale lavorativo dell'individuo si analizzano sia la sequenza di 12 valori 0/1, realizzazione delle 12 variabili indicatrici di presenza nel comune di un contratto attivo o meno per ognuno dei 12 mesi del 2012, sia la lista dei luoghi di lavoro dell'anno ai quali tali sequenze sono associate.

La prima analisi del segnale lavorativo di sequenza parte dall'osservazione che alla stessa tipologia lavorativa possono corrispondere sequenze simili ma non uguali. Si vogliono dunque individuare i gruppi di sequenze simili che possono essere associate ad una stessa tipologia lavorativa.

Per individuare delle strutture di vicinanza tra le 3.398 sequenze mensili distinte, osservate nei dati, si è utilizzata la matrice delle distanze di Hamming per un'analisi dei gruppi di tipo agglomerativo gerarchico con legame completo. La distanza di Hamming è la più semplice distanza calcolabile tra due stringhe della stessa lunghezza ed è data dal numero di posizioni con valori diversi, equivalente alla distanza euclidea per dati dicotomici.

La figura 1 mette in evidenza come l'analisi dei gruppi non riesca a caratterizzare l'insieme delle sequenze. Infatti, non essendo presenti salti evidenti nella distanza tra i gruppi che passo a passo si formano, non è possibile individuare un'altezza ottimale a cui tagliare il dendrogramma, fissando così il numero di gruppi presenti.

Figura 1. Dendrogramma del clustering sulle sequenze mensili



Fonte: nostre elaborazioni su dati Istat

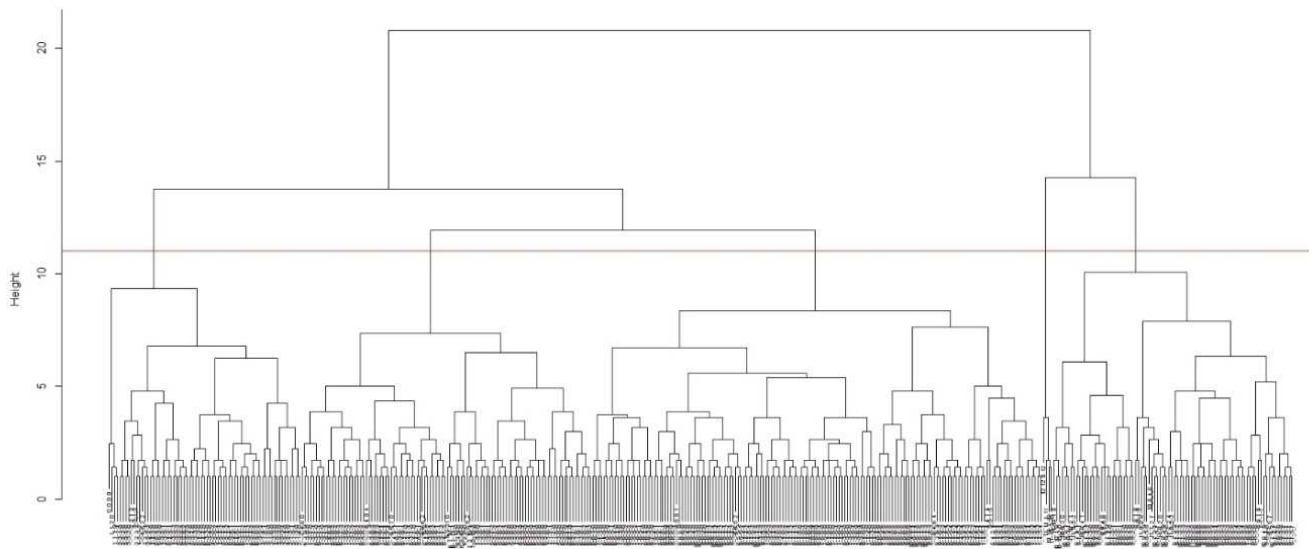
Si sono definite allora 4 variabili di sintesi delle sequenze dei 12 valori 0/1, cui è associata una interpretazione in termini di segnale lavorativo:

- Somma degli 1 nella sequenza (Num1): numero di mesi per cui è presente un segnale nelle fonti di lavoro nell'anno 2012;
- Numero massimo di 1 consecutivi (Num1_max): numero di mesi del periodo più lungo di attività;
- Numero di interruzioni di segnale (Num_int): numero di volte in cui l'individuo passa dall'essere attivo all'essere non attivo e viceversa;
- Numero di 1 consecutivi che comprendono la posizione 12 (Num12): numero di mesi dell'eventuale periodo di attività dell'individuo che comprende il mese di dicembre.

Sulla base di queste variabili le combinazioni osservate si riducono a 424, e su queste viene ripetuta l'analisi dei gruppi.

In questo caso il risultato dell'analisi è più chiaro ed in corrispondenza del primo salto della distanza è possibile individuare 5 gruppi (figura 2, tabella 1).

Figura 2. Dendrogramma del clustering sulle 4 variabili



Fonte: nostre elaborazioni su dati Istat

Tabella 1. Numero di sequenze e di individui nei 5 gruppi

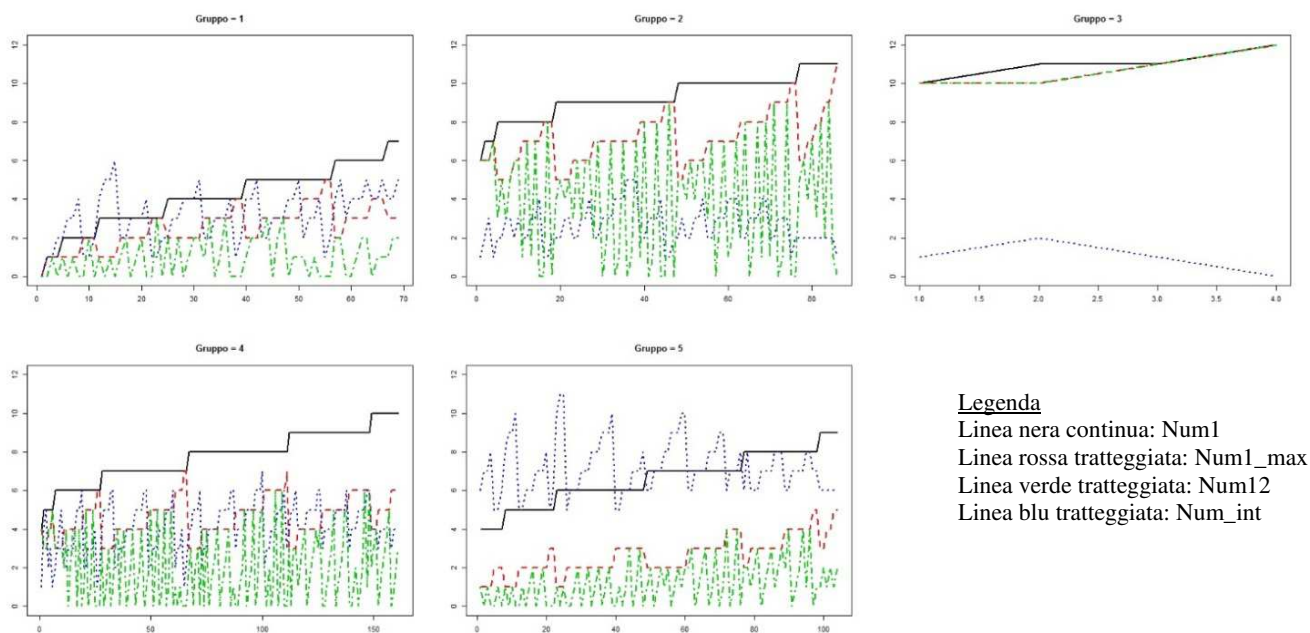
Gruppo	Numero di sequenze distinte	Numero di individui
Gruppo 1	796	368.747
Gruppo 2	142	196.358
Gruppo 3	4	1.271.407
Gruppo 4	1.023	119.283
Gruppo 5	1.433	5.014

Fonte: nostre elaborazioni su dati Istat

Notiamo come al gruppo 3 corrisponda il maggior numero di individui ma il numero di sequenze minore, mentre al gruppo 5 corrisponda il più alto numero di sequenze distinte ma il più basso numero di soggetti. Ne consegue che, mentre il gruppo 3 è caratterizzato da una grande omogeneità degli individui dal punto di vista del profilo lavorativo, all'opposto il gruppo 5 risulta composto da individui con profili lavorativi molto diversi. I restanti 3 gruppi rappresentano delle situazioni intermedie.

La figura 3 permette di caratterizzare i gruppi sulla base del comportamento delle 4 variabili.

Figura 3. Valore delle variabili Num1, Num1_max, Num_int e Num12 per ciascuna delle sequenze presenti nei 5 gruppi

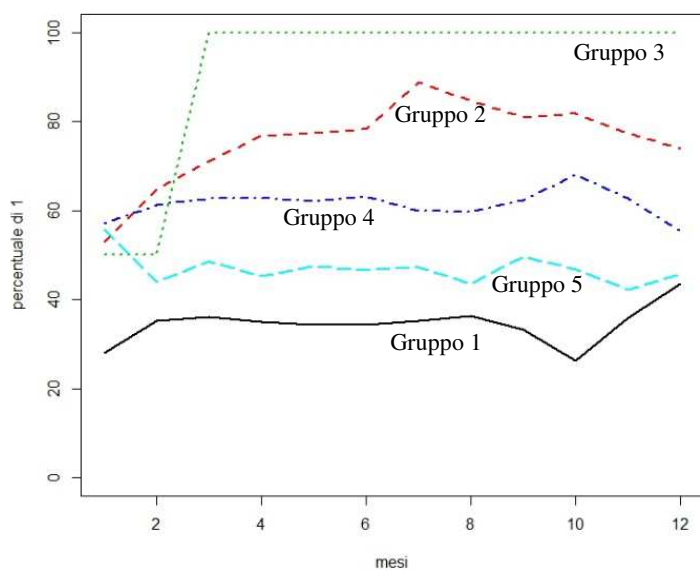


Fonte: nostre elaborazioni su dati Istat

Il gruppo 1 si caratterizza principalmente per un basso numero di mesi di attività lavorativa (linea nera continua). Il gruppo 2 al contrario presenta sequenze con un numero alto di mesi di attività consecutivi (linea rossa tratteggiata). Il gruppo 3 è sostanzialmente caratterizzato da “profili pieni”, ossia dipendenti che hanno lavorato per tutto l’anno (o quasi). Il gruppo 4 rappresenta soggetti con un segnale intermedio fra il gruppo 1 e 2, con un numero di mesi di attività che va da 5 a 10. Il gruppo 5 è caratterizzato da un alto numero di interruzioni (linea blu tratteggiata).

L’analisi del “profilo medio” di gruppo (derivante dal calcolo della percentuale delle sequenze con valore pari a 1, entro ciascun gruppo, per ogni mese) conferma le osservazioni precedenti. La figura 4 mostra in maniera evidente quanto già detto per i primi 4 gruppi, mentre il gruppo 5 sembra esprimere un profilo apparentemente stabile intorno al 50% su ogni mese, dovuto in effetti alla grande eterogeneità delle sequenze che lo compongono, imputabile alla natura residuale del gruppo.

Figura 4. Valore delle variabili Num1, Num1_max, Num_int e Num12 per ciascuna delle sequenze presenti nei 5 gruppi



Fonte: nostre elaborazioni su dati Istat

La capacità delle 4 variabili individuate di caratterizzare i gruppi di sequenze in termini di tipologie lavorative specifiche, è confermata dalla scarsa sovrapposizione dei profili medi. Inoltre, i profili sembrano contrassegnati anche da particolari comportamenti stagionali: in corrispondenza di ottobre si osservano picchi positivi per i gruppi 2 e 4 e un picco negativo per il gruppo 1, mentre il gruppo 2 si caratterizza anche per un picco in corrispondenza del mese di luglio.

Sulla base di questi risultati, si può affermare che le sequenze cui corrispondono valori vicini delle quattro variabili possono essere considerate simili ed associate a tipologie lavorative analoghe. Nel prosieguo, dunque, il segnale lavorativo dell'individuo nel comune sarà analizzato a partire da questi 4 indicatori invece che dalle 12 variabili mensili.

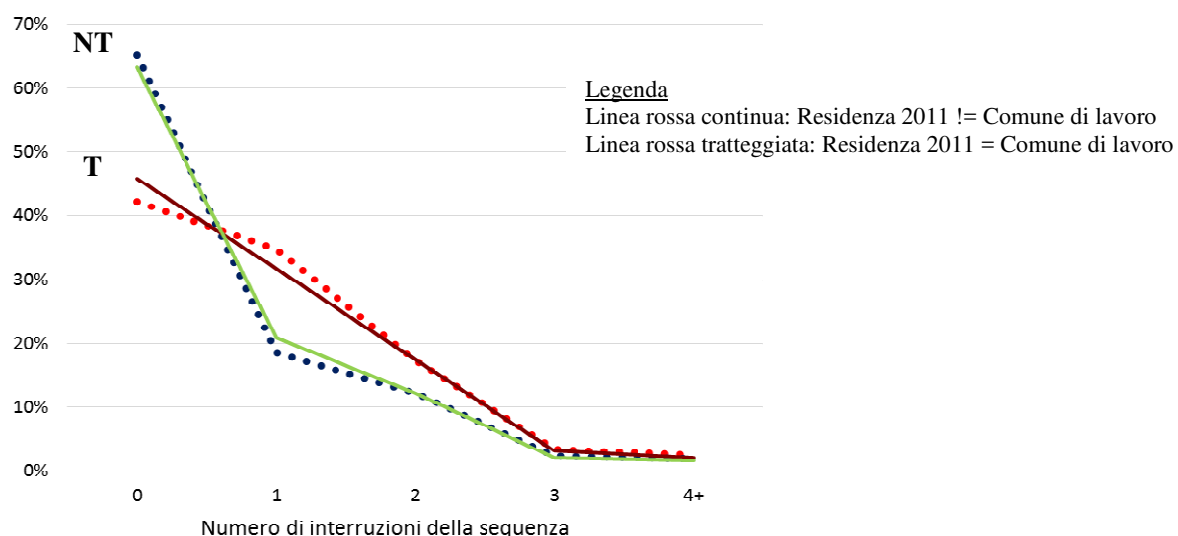
4. L'informazione sui luoghi di lavoro

Ogni sequenza mensile di segnali di attività è associata al comune dove tale attività si svolge (o meglio dove viene registrata) ed è quindi interpretabile, in una certa misura, come segnale di presenza in quel comune. La lista dei comuni lavorativi associabili ad un individuo è essa stessa un'informazione da analizzare in relazione al fenomeno oggetto di studio. Il trasferimento di residenza è sintetizzato con la variabile "Trasferimento" che assume valore 1 quando la residenza al 31 dicembre 2012 è diversa dalla residenza alla data del censimento 2011.

Si introduce quindi la variabile indicatrice "Residenza 2011 != Comune di lavoro" che prende valore 1 se il comune di residenza al 2011 è diverso da tutti i comuni di lavoro dell'individuo e 0 altrimenti, con l'obiettivo di valutare la relazione tra i luoghi di lavoro e il trasferimento. La percentuale di trasferiti tra coloro che risiedono in uno dei comuni di lavoro è pari al 2,3%, mentre i trasferiti che non risiedevano in uno dei comuni di lavoro sono il 6,8% suggerendo come la variabile sui luoghi di lavoro risulti associata alla variabile trasferimento.

L'informazione relativa alla sequenza e quella relativa ai luoghi risultano indipendenti. Come si evidenzia nella figura 5, le sequenze dei trasferiti hanno una distribuzione percentuale, relativamente al numero di interruzioni della sequenza, diversa da quelle dei non trasferiti, sia che l'individuo risieda in uno dei comuni di lavoro oppure no. In generale, confrontando analoghe distribuzioni per ognuna delle restanti 3 variabili di sintesi delle sequenze, i non trasferiti tendono ad avere sequenze più stabili, con un numero più elevato di 1 e un numero minore di interruzioni di segnale ma sempre indipendenti dalla coincidenza o meno della residenza con uno dei luoghi di lavoro.

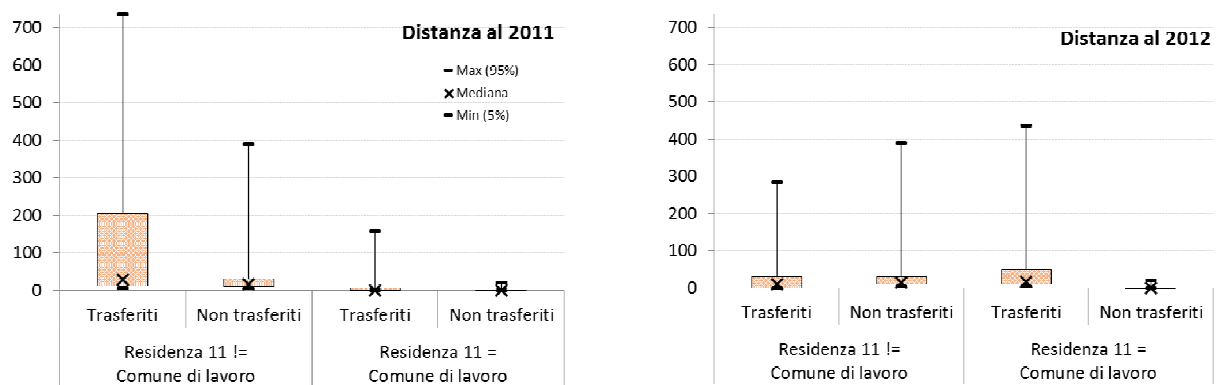
Figura 5. Distribuzione del numero di interruzioni della sequenza per trasferiti (T) e non trasferiti (NT) e per residenza al 2011 in uno dei luoghi di lavoro del 2012



Fonte: nostre elaborazioni su dati Istat

Aggiungendo a ciascun luogo le informazioni relative ad ampiezza demografica e distanza tra comune di lavoro e comune di residenza al 2011 e al 2012, è possibile valutare la capacità attrattiva del comune di lavoro. In particolare, si sono calcolate le distanze medie percorse tra la residenza al 2011 e al 2012 e i luoghi di lavoro del 2012. La figura 6 mostra come le distanze percorse da coloro che si trasferiscono da una residenza 2011 che non coincide con un luogo di lavoro diminuiscono tra il 2011 e il 2012 coerentemente con l'ipotesi che si trasferiscano per avvicinarsi al luogo di lavoro. Al contrario coloro che si trasferiscono pur risiedendo già in un luogo di lavoro, dopo il trasferimento aumentano la distanza media percorsa per raggiungere i luoghi di lavoro. Questo suggerisce che, in questo caso, l'evento che determina il trasferimento non sia direttamente associabile alla storia lavorativa.

Figura 6. Distribuzione della distanza media percorsa dagli individui per raggiungere il comune di lavoro a partire dal comune di residenza (2011 e 2012)



Fonte: nostre elaborazioni su dati Istat

Resta da stabilire quale dei luoghi di lavoro possa determinare lo spostamento di residenza. Sembra ragionevole supporre che sia il luogo in cui si è svolta l'attività lavorativa negli ultimi mesi dell'anno ad essere eletto più probabilmente residenza nell'anno successivo.

5. La nuova selezione di soggetti

Se l'obiettivo dell'analisi è caratterizzare il ruolo dell'attività lavorativa come determinante del trasferimento di residenza, occorre però tenere presente che il trasferimento può essere legato a motivazioni diverse: familiari, fiscali oltre che lavorative. Per isolare l'effetto del luogo di lavoro nella scelta del trasferimento, si è dunque osservato che gli eventi lavorativi, ricavabili dagli archivi e capaci di determinare almeno in parte un trasferimento in un luogo che sia in una certa misura prevedibile, sono l'aver cambiato comune di lavoro o l'aver iniziato a lavorare.

Partendo da questa osservazione, si è cercato di individuare il sottoinsieme di individui che più probabilmente ha sperimentato tali eventi in ragione delle informazioni lavorative ad essi associabili. L'analisi vista fin qui suggerisce che il trasferimento di residenza risulta maggiormente legato al profilo lavorativo fra coloro che risiedono in un comune che non coincide con nessuno dei luoghi di lavoro.

La tabella 2 mette in evidenza come i dipendenti che si trasferiscono lavorano più spesso su più comuni, più facilmente se il comune di residenza coincide con uno dei comuni di lavoro. La presenza di 2 o più profili può corrispondere ad una nuova attività lavorativa che spinge poi il soggetto a trasferirsi.

Tabella 2. Distribuzione del numero di comuni di lavoro per trasferiti e non trasferiti e per residenza al 2011 in uno dei luoghi di lavoro del 2012

N. comuni lavoro	Residenza 2011 = Comune lavoro 2012				Residenza 2011 != Comune lavoro 2012				Totale	
	Trasferiti		Non trasferiti		Trasferiti		Non trasferiti		n.	%
	n.	%	n.	%	n.	%	n.	%		
1	12.801	70,7%	659.125	87,1%	46.350	77,7%	707.928	86,1%	1.426.204	86,1%
2	4.043	22,3%	75.977	10,0%	10.175	17,1%	90.152	11,0%	180.347	10,9%
3	926	5,1%	15.316	2,0%	2.258	3,8%	17.211	2,1%	35.711	2,2%
4-29	330	1,8%	6.125	0,8%	855	1,4%	6.944	0,8%	14.254	0,9%
Totale	18.100	100,0%	756.543	100,0%	59.638	100,0%	822.235	100,0%	1.656.516	100,0%

Fonte: nostre elaborazioni su dati Istat

5.1. Profilo principale

Viene indicato con “profilo principale” la sequenza con il più lungo pacchetto di 1 consecutivi che comprende il segnale di attività lavorativa nel mese di dicembre.

Sulla base di queste osservazioni sono stati selezionati gli individui per i quali è ragionevole interpretare l’eventuale cambio di residenza legato al cambio del comune di lavoro, come coloro che presentano le seguenti caratteristiche:

- Il comune di residenza al 2011 diverso dal comune del lavoro relativo al profilo principale
- Il profilo principale con una interruzione di segnale
- Gli eventuali altri profili che terminano per 0.

Nella tabella 3 si può notare come la percentuale dei trasferiti aumenti man mano che si costruisce il sottoinsieme di analisi, arrivando quindi ad individuare il sottoinsieme con percentuale più alta di trasferimenti proprio nell’insieme che presenta contemporaneamente tutte e tre le caratteristiche sopra elencate. Infatti, il sottoinsieme identificato, formato da 118.212 individui (228.005 record³) è più sensibile al problema del trasferimento di residenza: il numero di individui trasferiti fra il 2011 e il 2012, è pari a 13.174 (11,1%) mentre quella dei trasferiti sul totale della popolazione è pari al 4,7%.

Tabella 3. Schema riassuntivo dei passi effettuati nella selezione del sottoinsieme di analisi e relativo numero di trasferimenti

Numerosità popolazione	Trasferimenti 2011-2012
1.656.516 individui con lavoro dipendente nel 2012	77.738 (4,7%)
➤ 881.873 risiedono in un comune diverso dal comune del profilo principale	59.638 (6,8%)
➤ 178.805 con profilo principale con una interruzione di segnale	18.543 (10,4%)
➤ 118.212 eventuali altri profili che terminano per 0	13.174 (11,1%)

Fonte: nostre elaborazioni su dati Istat

5.2. Seconda variabile risposta

Considerando il totale dei trasferimenti di residenza, si può notare come il 46% sia verso uno dei comuni di lavoro, con solo il 9% di questi verso un comune di lavoro diverso da quello del profilo principale.

³ Ad ogni individuo possono essere associati più record. L’identificativo univoco di record è la combinazione di soggetto e comune in cui il soggetto ha segnali di lavoro.

Dunque la scelta di concentrarsi su quest'ultimo profilo ed il luogo ad esso associato risulta essere fin qui un buon compromesso tra la necessità di non perdere troppa informazione e quella di semplificare l'analisi utilizzando un solo profilo per individuo. Infine, è pari al 54% il gruppo di individui che si trasferisce ma per i quali non è possibile ricavare un'indicazione sul luogo di trasferimento dalla lista dei luoghi lavorativi.

Tabella 4. Distribuzione dei trasferiti in comuni di lavoro principale, altri comuni di lavoro, altri comuni non di lavoro

<i>Trasferiti per Comune di destinazione</i>	<i>n.</i>	<i>%</i>
Trasferiti nel comune del profilo principale	4.938	37,5%
Trasferiti in altro comune di lavoro	1.146	8,7%
Trasferiti in altro comune non di lavoro	7.090	53,8%
Totale trasferiti	13.174	100,0%

Fonte: nostre elaborazioni su dati Istat

La centralità dell'informazione legata al profilo principale risulta anche se si considera l'informazione di area. Introducendo per i comuni di residenza e di lavoro l'informazione sui Sistemi Locali di Lavoro (SLL) di appartenenza, si nota che la maggior parte dei trasferimenti di residenza è verso comuni che sono all'interno del Sistema locale di uno dei comuni di lavoro, con il 59% dei casi di trasferimento nel SLL del luogo di lavoro principale (tabella 5).

Tabella 5. Distribuzione dei trasferiti in SLL del lavoro principale, altro SLL di lavoro, SLL non di lavoro

<i>Trasferiti per SLL di destinazione</i>	<i>n.</i>	<i>%</i>
Trasferiti nel SLL del profilo principale	7.753	58,9%
Trasferiti nel SLL (diverso dal principale) di altro comune di lavoro	1.679	12,7%
Trasferiti nel SLL di altro comune non di lavoro	3.742	28,4%
Totale trasferiti	13.174	100,0%

Fonte: nostre elaborazioni su dati Istat

Sebbene il SLL non consenta l'individuazione del potenziale comune di residenza, esso sottolinea ancora una volta l'importanza indiretta dell'attività lavorativa nella scelta del comune di residenza, con "solo" un 28% di cambi di residenza che non possono essere associati in modo diretto all'attività lavorativa.

6. L'analisi di regressione

Si vuole a questo punto stimare la relazione funzionale che lega le due variabili risposta "Trasferimento" e "Trasferimento nel comune del profilo principale", con le informazioni disponibili a livello di record (identificato univocamente dalla chiave individuo/comune di lavoro) sulle caratteristiche dell'individuo (età, sesso, cittadinanza), su quelle del profilo lavorativo (sequenza mensile di attività, fonti) e sui luoghi (comune di lavoro, residenza 2011).

5.3. I regressori

Le elaborazioni riportate nella tabella 6 suggeriscono una diversa caratterizzazione sia dei trasferiti rispetto ai non trasferiti sia dei trasferiti nel comune di lavoro principale rispetto al loro insieme complementare, per le 3 variabili di soggetto considerate.

In particolare, la distribuzione percentuale per sesso non sembra differenziarsi molto tra trasferiti e non trasferiti ma quando si confrontano i trasferiti nel comune del lavoro principale e gli altri, si nota una inversione di percentuali tra Maschi e Femmine. Ciò potrebbe essere interpretato come una maggiore propensione delle donne a trasferirsi per motivi di lavoro.

Riguardo al SLL della residenza e del lavoro, è più probabile che spostino la residenza coloro che l'avevano "più lontana" dal luogo di lavoro, cioè appartenevano ad un SLL diverso da quello del luogo di

lavoro principale, mentre coloro che non si spostano hanno più frequentemente comune di residenza del 2011 e comune di lavoro principale che partecipano dello stesso SLL.

Tabella 6. Distribuzione di genere, SLL di residenza 2011 e cittadinanza per trasferiti e non trasferiti, trasferiti nel comune principale e altri

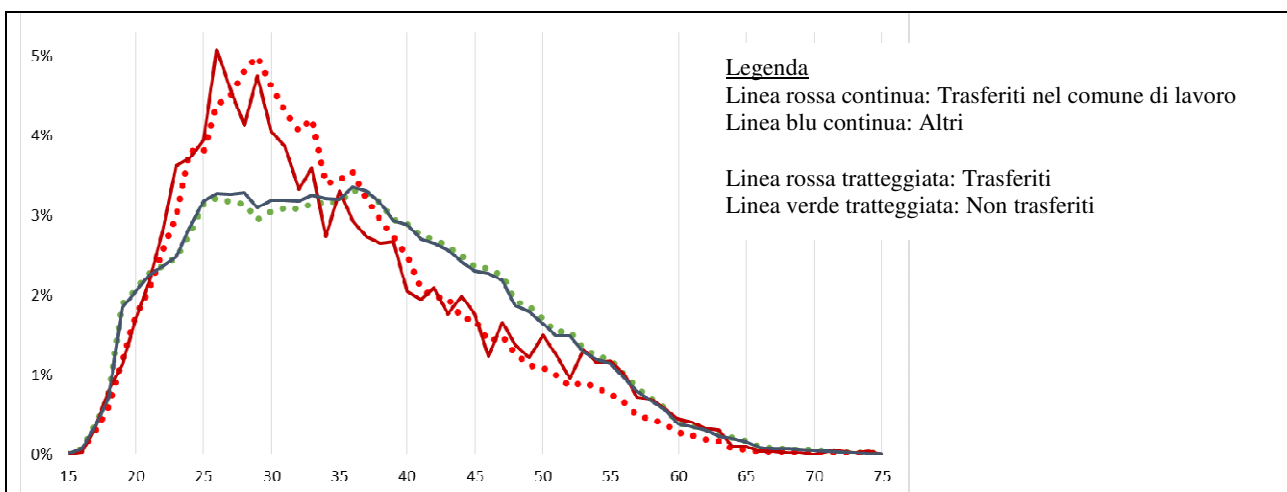
	Trasferiti		Non trasferiti		Trasferiti nel comune del lavoro principale		Altri	
	n.	%	n.	%	n.	%	n.	%
Totale popolazione	13.174	100,0%	105.038	100,0%	4.938	100,0%	113.274	100,0%
Genere								
F	6.071	46,1%	42.328	40,3%	2.818	57,1%	45.581	40,2%
M	7.103	53,9%	62.710	59,7%	2.120	42,9%	67.693	59,8%
SLL di residenza 2011								
Uguale al SLL del comune di lavoro principale 2012	2.490	18,9%	33.926	32,3%	1.058	21,4%	35.358	31,2%
Diverso dal SLL del comune di lavoro principale 2012	10.684	81,1%	71.112	67,7%	3.880	78,6%	77.916	68,8%
Cittadinanza								
Italiani	8.961	68,0%	87.624	83,4%	2.881	58,3%	93.704	82,7%
Stranieri	4.211	32,0%	17.383	16,5%	2.057	41,7%	19.537	17,2%
N.D.	2	0,0%	31	0,0%	-	0,0%	33	0,0%

Fonte: nostre elaborazioni su dati Istat

In ultimo, la cittadinanza evidenzia una maggiore presenza percentuale di stranieri tra i trasferiti (ma ancor più tra i trasferiti nel luogo di lavoro principale), rispetto ai non trasferiti (o non trasferiti nel luogo di lavoro principale) suggerendo una maggiore mobilità sul territorio dei cittadini stranieri rispetto agli italiani.

Per quanto concerne l'età (figura 1), sia i trasferiti rispetto ai non trasferiti che i trasferiti nel comune principale rispetto agli altri, risultano concentrati in classi d'età più giovani con un picco fra i 25 e i 30 anni.

Figura 7. Distribuzione della variabile età per trasferiti e non trasferiti, trasferiti nel comune principale e altri



Fonte: nostre elaborazioni su dati Istat

Volendo discretizzare la variabile per le analisi successive, si è tenuto conto di queste differenze e si sono individuate le seguenti classi d'età: <25 anni, fra 25 e 34 anni, fra 35 e 49 anni, >=50 anni (tabella 7).

Tabella 7. Distribuzione dell'età in classi per trasferiti e non trasferiti, trasferiti nel comune principale e altri

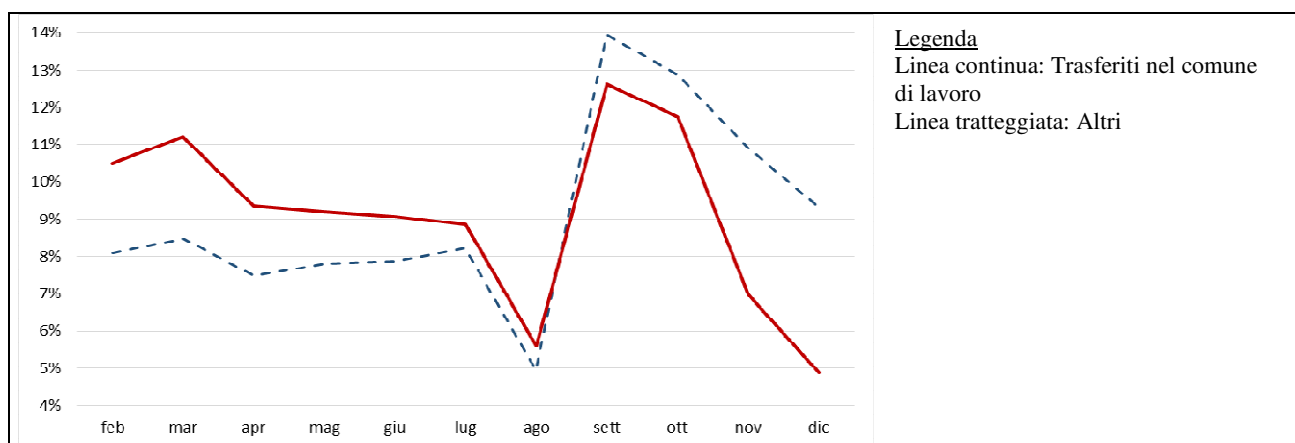
Classi di età	Trasferiti		Non trasferiti		Trasferiti nel comune del lavoro principale		Altri	
	n.	%	n.	%	n.	%	n.	%
<25 anni	2.010	15,3%	15.733	15,0%	804	16,3%	16.939	15,0%
fra 25 e 34 anni	5.677	43,1%	32.713	31,1%	1.979	40,1%	36.411	32,1%
fra 35 e 49 anni	4.360	33,1%	42.045	40,0%	1.549	31,4%	44.856	39,6%
>=50 anni	1.127	8,6%	14.545	13,8%	606	12,3%	15.066	13,3%
n.d.			2	0,0%			2	0,0%
Totale	13.174	100,0%	105.038	100,0%	4.938	100,0%	113.274	100,0%

Fonte: nostre elaborazioni su dati Istat

Oltre alla caratterizzazione degli individui, è possibile caratterizzare le sequenze dei segnali lavorativi mensili. Nella presente analisi, si è scelto di considerare la variabile relativa al numero di mesi di attività. Poiché il profilo principale ha una sola interruzione e il mese di dicembre è sempre attivo, la variabile descrive non solo la persistenza del segnale lavorativo ma lo associa ad un preciso intervallo dell'anno, per cui ad esempio un numero di segnali mensili pari a 4 indica che il dipendente ha lavorato da settembre a dicembre nel 2012.

La figura 2 mostra una più alta percentuale di trasferiti nel comune principale (linea continua) rispetto ai trasferiti in altro luogo ed ai non trasferiti se il dipendente ha lavorato fin dai primi mesi dell'anno; se invece ha lavorato solo l'ultimo quadrimestre del 2012 o meno, allora è più alta la percentuale di dipendenti trasferiti in altro luogo ed i non trasferiti (linea tratteggiata). Sembra dunque che la persistenza del segnale di attività nel luogo in cui si lavora favorisca il trasferimento.

Figura 8. Distribuzione dei trasferiti e non trasferiti, trasferiti nel comune principale e altri per mese di inizio dell'attività lavorativa



Fonte: nostre elaborazioni su dati Istat

Sulla base di queste osservazioni, si sono individuate 3 classi per la variabile che conta il numero di mesi di attività: 1-4, 5-8, 9-11 (tabella 8).

Tabella 8. Distribuzione del numero di mesi di attività, per trasferiti e non trasferiti e per trasferiti nel comune principale e altri

Numero di mesi	Trasferiti		Non trasferiti		Trasferiti nel comune principale		Altri	
	n.	%	n.	%	n.	%	n.	%
1-4	5.814	44,1%	49.306	46,9%	1.789	36,2%	53.331	47,1%
5-8	4.020	30,5%	30.278	28,8%	1.615	32,7%	32.683	28,9%
9-11	3.340	25,4%	25.454	24,2%	1.534	31,1%	27.260	24,1%
Totale	13.174	100,0%	105.038	100,0%	4.938	100,0%	113.274	100,0%

Fonte: nostre elaborazioni su dati Istat

Un'altra variabile in grado di caratterizzare il profilo lavorativo è quella relativa alle fonti da cui si è tratta l'informazione della sequenza mensile. Essa può essere indicativa del tipo di attività svolta dal dipendente. Sulla base della combinazione delle fonti osservate sui record, sono stati identificati 5 gruppi di profili riconducibili a tipologie specifiche di attività lavorativa: Dipendenti privati, Dipendenti pubblici, Dipendenti agricoli, Lavoratori domestici, Altri. Nella tabella 1 è riportata la distribuzione della variabile aggregata secondo le 5 classi identificate.

Tabella 9. Distribuzione delle fonti aggregate per trasferiti e non trasferiti, trasferiti nel comune principale e altri

Fonti aggregate	Trasferiti		Non trasferiti		Trasferiti nel comune principale		Altri	
	n.	%	n.	%	n.	%	n.	%
Dipendenti privati	9.982	75,8%	85.377	81,3%	3.144	63,7%	92.215	81,4%
Dipendenti pubblici	520	3,9%	3.674	3,5%	233	4,7%	3.961	3,5%
Dipendenti agricoli	454	3,4%	3.728	3,5%	199	4,0%	3.983	3,5%
Lavoratori domestici	1.022	7,8%	844	0,8%	916	18,6%	950	0,8%
Altri	1.196	9,1%	11.415	10,9%	446	9,0%	12.165	10,7%
Totale	13.174	100,0%	105.038	100,0%	4.938	100,0%	113.274	100,0%

Fonte: nostre elaborazioni su dati Istat

La distribuzione delle diverse tipologie lavorative fra trasferiti e non trasferiti non presenta particolari differenze, se non per i lavoratori domestici che mostrano una maggiore propensione allo spostamento, soprattutto se si tratta di spostarsi nel comune di lavoro principale. Occorre però notare che il significato del luogo di lavoro indicato nelle diverse fonti amministrative considerate, può essere diverso tra una fonte e l'altra. Mentre per i lavoratori domestici il Comune riportato nella fonte è quello in cui viene effettivamente svolto il lavoro, in altri casi, come per i dipendenti pubblici, il comune riportato nella fonte può essere quello della sede centrale dell'Ente e non il luogo effettivo dove il dipendente svolge la sua attività.

Infine, si è analizzata l'informazione a livello comunale catturata dall'ampiezza demografica, sia dei comuni di residenza al 2011 che di lavoro del 2012. In generale le aree metropolitane sono caratterizzate da una maggiore mobilità, sia che si tratti del luogo di residenza (tabella 2) che di lavoro (tabella 3). La percentuale di lavoratori in comuni sotto i 50.000 abitanti al 2011 risulta meno elevata per i trasferiti nel comune principale che per gli altri, quindi la probabilità di trasferirsi nel comune di lavoro principale sembra superiore se questo è una città metropolitana.

Tabella 10. Distribuzione dell'ampiezza demografica del comune di residenza per trasferiti e non trasferiti, trasferiti nel comune principale e altri

Ampiezza del comune di residenza	Trasferiti		Non trasferiti		Trasferiti nel comune principale		Altri	
	n.	%	n.	%	n.	%	n.	%
fino 50.000	9.045	68,7%	73.783	70,2%	3.289	66,6%	79.537	70,2%
oltre 50.000	2.840	21,6%	23.457	22,3%	1.121	22,7%	25.176	22,2%
città metropolitana	1.289	9,8%	7.800	7,4%	528	10,7%	8.561	7,6%
Totale	13.174	100,0%	105.038	100,0%	4.938	100,0%	113.274	100,0%

Fonte: nostre elaborazioni su dati Istat

Tabella 11. Distribuzione dell'ampiezza demografica del comune di lavoro per trasferiti e non trasferiti, trasferiti nel comune principale e altri

Ampiezza del comune di residenza	Trasferiti		Non trasferiti		Trasferiti nel comune principale		Altri	
	n.	%	n.	%	n.	%	n.	%
fino 50.000	7.387	56,1%	63.531	60,5%	2.208	44,7%	68.710	60,7%
oltre 50.000	4.073	30,9%	29.107	27,7%	1.912	38,7%	31.268	27,6%
città metropolitana	1.714	13,0%	12.400	11,8%	818	16,6%	13.296	11,7%
Totale	13.174	100,0%	105.038	100,0%	4.938	100,0%	113.274	100,0%

Fonte: nostre elaborazioni su dati Istat

5.4. Il modello logistico

Come strumento di sintesi delle osservazioni fatte fin qui e volendo valutare la significatività delle relazioni evidenziate, si è scelto di applicare un modello di regressione logistica tra ognuna delle variabili risposta, "Trasferimento" (TRASF) e "Trasferimento nel comune del profilo principale" (TRASF_LAVp), e le variabili descrittive di soggetto, profilo e comune (Tabella 12):

Tabella 12. Nome e descrizione delle variabili utilizzate nei modelli logistici

Nome variabile	Descrizione variabile
TRASF	1 se la residenza al 2011 è diversa da quella al 2012; 0 altrimenti.
TRASF_LAVp	1 se la residenza al 2011 è diversa da quella al 2012 ma uguale al comune relativo al profilo principale dell'individuo; 0 altrimenti.
SESSO	"F" se femmina; "M" se maschio
FLG_RES11_SLLD	1 se la residenza al 2011 appartiene allo stesso SLL del comune relativo al profilo principale; 0 altrimenti
CITTAD	1 se il dipendente ha cittadinanza italiana; 2 se ha cittadinanza straniera
ETA_4C	1 se l'età al 2011 è inferiore a 25 anni; 2 se l'età al 2011 è compresa fra 25 e 34 anni; 3 se l'età al 2011 è compresa fra 35 e 49 anni; 4 se l'età al 2011 è maggiore o al più uguale ai 50 anni
FONTI12_C	1 Dipendenti privati; 2 Dipendenti pubblici; 3 Dipendenti agricoli; 4 Lavoratori domestici; 5 Altri.
NUM12_4C	1 se i mesi attivi sono da 1 a 4, compreso dicembre; 2 se i mesi attivi sono da 5 a 8, compreso dicembre; 3 se i mesi attivi sono da 9 a 11, compreso dicembre
COMUNE_3C_RES11	1 se il comune di residenza al 2011 ha fino 50.000 abitanti; 2 se il comune di residenza al 2011 ha oltre 50.000 abitanti; 3 se il comune di residenza al 2011 è considerato città metropolitana
COMUNE_3C_D	1 se il comune di lavoro al 2011 ha fino 50.000 abitanti; 2 se il comune di lavoro al 2011 ha oltre 50.000 abitanti; 3 se il comune di lavoro al 2011 è considerato città metropolitana

$$\text{logit}(Y) = \text{SESSO} + \text{ETA_4C} + \text{CITTAD} + \text{FONTI12_C} + \text{NUM12_4C} + \text{FLG_RES11_SLLD} + \text{COMUNE_3C_RES11} + \text{COMUNE_3C_D} \quad [1]$$

dove Y = TRASF o TRASF_LAVp.

Come profilo di riferimento, per entrambi i modelli, si sono considerate le donne straniere di età compresa tra i 34 ed i 49 anni, che risiedevano nel 2011 in un comune di più di 50.000 abitanti (non metropolitano), e che per il 2012 presentano un profilo principale, di fonte “privati”, attestante un’attività lavorativa di al più quattro mesi (minimo dicembre, al massimo da settembre alla fine dell’anno) in un comune diverso da quello di residenza.

I parametri stimati misurano la variazione nel *logit* dovuta al possesso di una data modalità rispetto a quella posseduta dal gruppo di riferimento (Tabella 13).

Tabella 13. Stima dei parametri del modello di regressione

	TRASF	TRASF_LAVp
(Intercept)	-2.330 ***	-3.370 ***
SESSO_F	-	-
SESSO_M	-0.196 ***	-0.501 ***
FLG_RES11_SLLD_0	-	-
FLG_RES11_SLLD_1	-0.741 ***	-0.459 ***
CITTAD_italiani	-	-
CITTAD_stranieri	0.617 ***	0.799 ***
ETA_4C_<25	0.256 ***	0.466 ***
ETA_4C_25-34	0.509 ***	0.499 ***
ETA_4C_34-49	-	-
ETA_4C_>=50	-0.437 ***	-0.213 ***
FONTI12_C_1privati	-	-
FONTI12_C_2pubblici	0.378 ***	0.727 ***
FONTI12_C_3agricoli	0.053	0.403 ***
FONTI12_C_4domestici	2.088 ***	2.824 ***
FONTI12_C_5altri	-0.097 **	0.065
POS12_NUM1_12_4C_1-4	-	-
POS12_NUM1_12_4C_5-8	0.088 ***	0.288 ***
POS12_NUM1_12_4C_9-11	0.140 ***	0.524 ***
COMUNE_3C_D_<50.000	-0.064 **	-0.613 ***
COMUNE_3C_D_>50.000	-	-
COMUNE_3C_D_metro	0.021	0.055
COMUNE_3C_RES11_<50.000	0.139 ***	-0.002
COMUNE_3C_RES11_>50.000	-	-
COMUNE_3C_RES11_metro	0.382 ***	0.326 ***

*** p<0.001; ** p<0.01;

Fonte: nostre elaborazioni su dati Istat

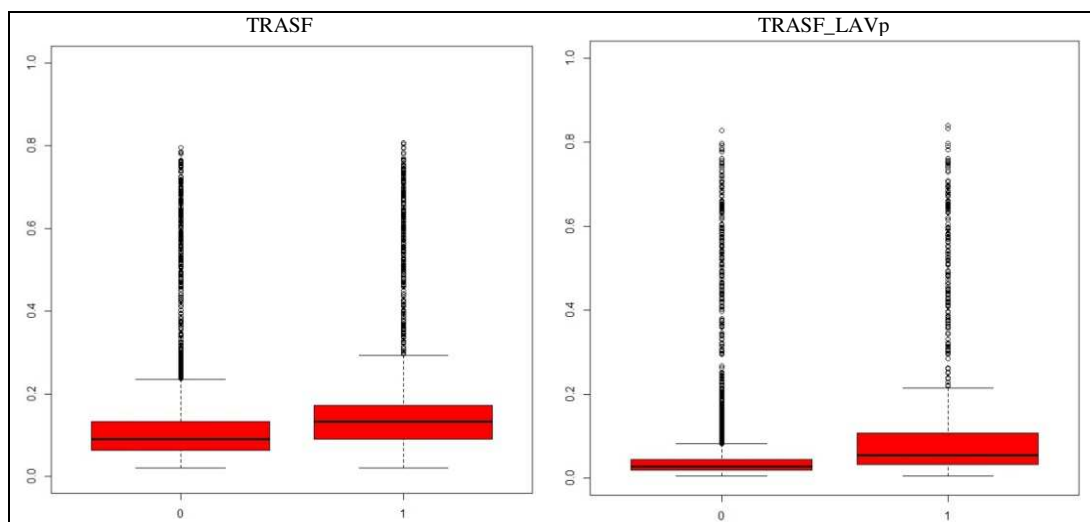
Risultano significativamente più protetti dallo sperimentare l’evento trasferimento: gli uomini rispetto alle donne; coloro che hanno età maggiore di 50 anni rispetto a coloro che hanno tra i 34 e 49 anni; i residenti nel SLL del comune di lavoro principale; i dipendenti il cui segnale è ricavato dalla fonte “altri” rispetto alla fonte “privati”; coloro che lavorano in un comune con meno di 50 mila abitanti rispetto a “> 50.000”.

Risultano più esposti gli individui che presentano le restanti modalità, cui è associata una stima significativa e di segno positivo del parametro.

I risultati per il secondo modello sono simili ma in generale gli effetti risultano più forti (valori delle stime dei parametri più elevati). Solo per la variabile “dimensione demografica della residenza”, la modalità “<50.000 abitanti” risulta non significativa nel secondo modello.

I *boxplot* (Figura 9) delle distribuzioni dei valori delle probabilità stimate per i trasferiti e i non trasferiti, mostrano che le probabilità associate a coloro che non si trasferiscono sono tendenzialmente più basse, ma complessivamente la distribuzione dei non trasferiti si sovrappone a quella dei trasferiti, indice di una scarsa capacità discriminativa del modello. Nei *boxplot* relativi al secondo modello le parti centrali delle due distribuzioni risultano più distanziate ma le code si sovrappongono comunque.

Figura 9. *Boxplot delle distribuzioni delle probabilità stimate per i trasferiti e i non trasferiti (TRASF), trasferiti nel comune principale/altri (TRASF_LAVp)*



Fonte: nostre elaborazioni su dati Istat

5.5. La validazione del modello

La base dati AIDA contiene i dati relativi ai profili lavorativi degli individui attivi nel 2012 e nel 2013. Le precedenti analisi sono state condotte sui dati relativi al 2012. Per validare il modello stimato sono stati utilizzati i dati relativi ai lavoratori dipendenti attivi nel 2013. A tale scopo sono stati selezionati 113.051 individui che nel corso del 2013 hanno cambiato comune di lavoro o hanno iniziato a lavorare, con comune di residenza al 31 dicembre 2012 diverso dal comune di lavoro principale. Per ciascun individuo così selezionato, sono state calcolate le probabilità di trasferirsi e di trasferirsi nel comune del profilo principale, a partire dalle stime dei parametri di regressione ottenuti dai due modelli per il 2012.

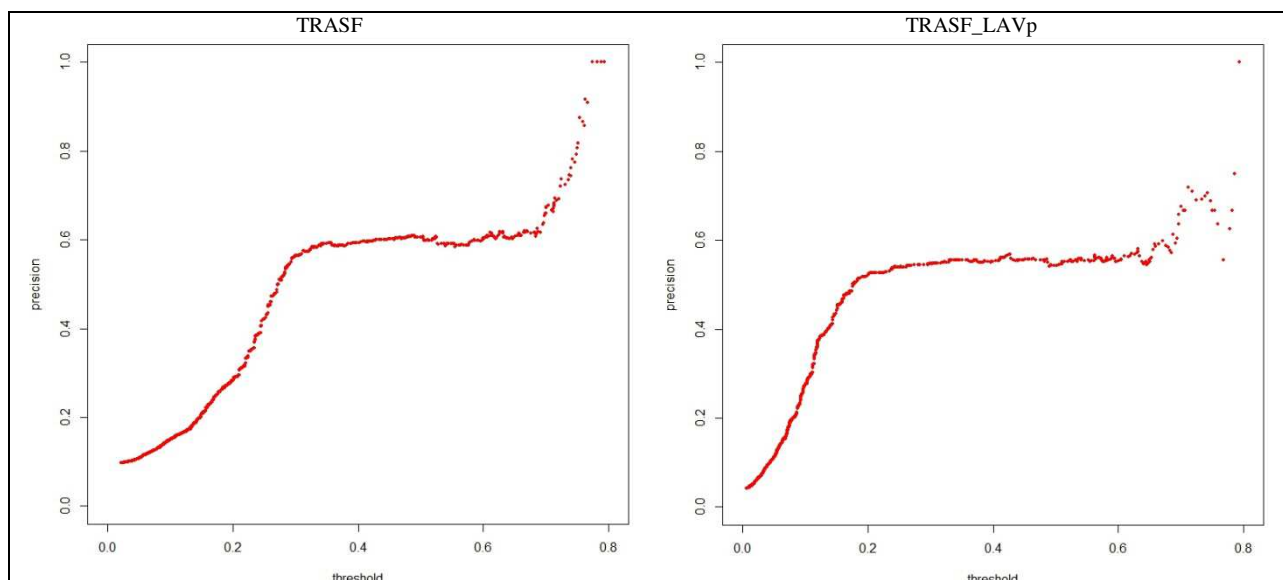
Scelto un valore soglia per la probabilità stimata, che identifichi l'appartenenza alla classe positiva (per la quale il valore predetto della variabile risposta è 1) o negativa, si costruisce una tabella di contingenza, che incrocia le modalità osservate e predette della variabile risposta. Sono così identificate le frequenze di falsi positivi, falsi negativi, veri positivi e veri negativi.

La variabile “Trasferimento” e soprattutto la variabile “Trasferimento nel comune di lavoro principale”, hanno una dimensione della sottopopolazione che sperimenta l'evento decisamente ridotta rispetto al gruppo complementare. In questo caso, nel fissare la soglia occorre prestare attenzione, non solo al numero di veri positivi identificati, ma anche al numero di falsi positivi, infatti una bassa probabilità di generare falsi positivi, applicata però a un gran numero di ‘suscettibili’, può facilmente generare un numero di falsi positivi più elevato del numero di ‘veri positivi’. Al posto dell'usuale curva ROC si è dunque scelto come criterio di

valutazione della soglia, la precisione, definita come il rapporto tra veri positivi e totale positivi (veri positivi + falsi positivi).

Nella figura 10 sono riportati, sull'asse delle ascisse, i valori delle probabilità stimate per la popolazione del 2013, e sull'asse delle ordinate la precisione calcolata assumendo quel valore di probabilità come valore soglia.

Figura 10. Precisione rispetto a probabilità stimata, per i modelli trasferiti/non trasferiti (TRASF), trasferiti nel comune principale/altri (TRASF_LAVp)



Fonte: nostre elaborazioni su dati Istat

Per entrambi i modelli la precisione raggiunge valori molto elevati, fino al massimo di 1, corrispondenti alle soglie di probabilità di 0,775 (8 veri positivi) e 0,794 (3 veri positivi). Volendo aumentare il numero di veri positivi identificati si può scendere al livello successivo di precisione (0,917 e 0,719) cui corrispondono le tabelle di contingenza riportate di seguito (Tabella 14).

Tabella 14. Tabelle di contingenza dei valori osservati e predetti per il 2013 relativi ai modelli trasferiti/non trasferiti (TRASF), trasferiti nel comune principale/altri (TRASF_LAVp)

TRASF	$P > 0,763$	$P < 0,763$	TRASF_LAVp	$P > 0,712$	$P < 0,712$
1	11	11.026	1	23	4.705
0	1	101.970	0	9	108.271

Fonte: nostre elaborazioni su dati Istat

Nella figura 10, per entrambi i modelli, si nota la presenza di un lungo *plateaux*, identificato da un ampio intervallo di valori soglia cui corrisponde un valore di precisione con poche fluttuazioni. Per comprendere l'origine di questo andamento si analizzano in dettaglio le caratteristiche dei soggetti cui corrispondono questi valori.

Ordinando gli individui in modo decrescente rispetto a precisione e probabilità di trasferirsi, i primi soggetti risultano tutti di fonte Domestici e per lo più donne straniere. Ciò suggerisce che, per questo tipo di lavoratori, un cambio del comune di lavoro equivale a un quasi immediato cambio di residenza.

7. Conclusioni

Per poter valutare la componente lavorativa nella scelta del trasferimento di residenza è importante individuare la sottopopolazione maggiormente esposta al rischio di trasferimento, tenendo anche conto della tipologia di informazione lavorativa di cui si dispone.

Il segnale mensile di attività da fonti amministrative può essere utilizzato per individuare i soggetti che hanno sperimentato un cambio di comune di lavoro o di inizio di lavoro. Questa informazione incrociata alla relazione tra luoghi di residenza e di lavoro permette di individuare un sottoinsieme di popolazione dove la percentuale di trasferiti risulta più che raddoppiata, pur rimanendo un fenomeno marginale che riguarda l'11% della popolazione.

Nelle analisi effettuate, a ciascun individuo è stato associato un solo luogo di lavoro sulla base delle caratteristiche delle sequenze, permettendo di semplificare le analisi. Questa scelta ha assicurato il mantenimento di una grossa parte delle informazioni e la possibilità di introdurre una più specifica variabile risposta che associa all'individuo una probabilità di trasferirsi in un determinato luogo.

I risultati dell'analisi di regressione indicano che il ruolo dell'informazione lavorativa è importante solo per specifici gruppi, quali i lavoratori domestici che guidano gran parte dei risultati dell'analisi di regressione. Il fatto che risultino più esposti al rischio di trasferimento le donne rispetto agli uomini e gli stranieri rispetto agli italiani è dovuto principalmente ai lavoratori domestici che sono in maggioranza donne straniere.

È possibile che l'inserimento dell'informazione sul cambio del luogo di lavoro in un contesto longitudinale possa far aumentare la capacità predittiva del luogo di lavoro per la caratterizzazione del fenomeno dei trasferimenti.

8. Bibliografia

- S. Ambroselli (2014) I codici identificativi univoci all'interno del SIM (Sistema Integrato di Microdati). *Giornate della ricerca in Istat, 10-11 novembre 2014.*
- F. Borrelli, A. Chieppa, S. Di Domenico, G. Gallo, S. Rosati, V. Tomeo (2018) Primi risultati della sperimentazione condotta su fonti amministrative capaci di valutare i segnali di dimora abituale in Italia e individuazione di sottopopolazioni critiche. *Istat Working Papers n. 23/2016.*
- A. Chieppa, G. Gallo, Borrelli F., S. Di Domenico, V. Tomeo (2018) Knowledge Discovery for inferring the Usually-Resident Population. *Mathematical population studies.*

ABSTRACT

Change of residence can be motivated by family, fiscal, job or subjective reasons. The present paper aims to evaluate the role of the workplace to plan a change of residence, using the archive information about working activity. The population analyzed is composed of individuals with at least one work contract in the administrative archives in 2012 in Emilia Romagna. The activity status (active / inactive) of the individual in a municipality refers to the 12 month of the year by a sequence of 12 indicators. The analysis of working information (activity sequence, workplaces and residence) allows identifying the events potentially determining the change of residence: starting a job in a municipality different from the residence or changing the municipality of work. We perform a logistic regression analysis to estimate the residence change probability in relation to subject, profile and municipality regressors for subjects experiencing these events.