



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

Corso di Laurea Specialistica in
SCIENZE STATISTICHE, ECONOMICHE, FINANZIARIE ED AZIENDALI

TESI DI LAUREA

**ANALISI LONGITUDINALE DELL'OFFERTA
DI LAVORO A PARTIRE DALLA
NUOVA RILEVAZIONE CONTINUA
DELLE FORZE DI LAVORO**

Relatore: Ch.mo Prof. ADRIANO PAGGIARO

Laureando: CLAUDIO MANZAN

ANNO ACCADEMICO 2006 – 2007

INDICE

Capitolo 1. Introduzione e sommario	1
Capitolo 2. La Rilevazione Continua Forze Lavoro	5
2.1 <i>Le novità della RCFL</i>	6
2.1.1 Piano di campionamento	7
2.1.2 Le tecniche di rilevazione	9
2.1.3 Il questionario	11
2.2 <i>L'abbinamento longitudinale</i>	12
2.2.1 Abbinamenti e informazioni invarianti	15
2.3 <i>Condizione lavorativa al momento dell'intervista</i>	19
2.4 <i>Informazioni retrospettive sulla vita lavorativa: tassi di risposta</i>	22
2.4.1 Durata della ricerca di lavoro	24
2.4.2 Fine dell'attività lavorativa	29
2.4.3 Inizio dell'attività lavorativa	33
2.5 <i>Incoerenze nei dati retrospettivi</i>	37
2.5.1 Incoerenze sull'inizio dell'attività lavorativa	38
2.5.2 Incoerenze sulla fine dell'attività lavorativa	41
2.5.3 Durata della ricerca di lavoro	45
Capitolo 3 Ricostruzione e analisi della durata della disoccupazione	49
3.1 <i>Modelli di durata della disoccupazione</i>	49
3.1.1 Problemi di censura	51
3.1.2 Rischi competitivi	52
3.2 <i>Dati dell'indagine e ricostruzione degli episodi</i>	53
3.2.1 Selezione del campione	54
3.2.2 Ricostruzione degli episodi di disoccupazione	57
3.2.3 Transizione verso l'occupazione	58
3.2.4 Transizione verso l'inattività	61

3.2.5 Individui censurati a destra	62
3.2.6 Il dataset	63
<i>3.3 Informazioni disponibili e modelli di durata</i>	<i>64</i>
3.3.1 Scelta della funzione di rischio di base	65
3.3.2 Stima non parametrica della sopravvivenza	67
3.3.3 Stima parametrica	69
<i>3.4 Utilizzo di altre informazioni per la determinazione della durata</i>	<i>74</i>
3.4.1 Informazioni aggiuntive: dati ricostruiti	77
<i>3.5 Possibili sviluppi</i>	<i>79</i>
Bibliografia	81

CAPITOLO 1

Introduzione e sommario

Per comprendere le dinamiche nel mercato del lavoro di un paese come l'Italia, che negli ultimi anni ha registrato un netto calo del tasso di disoccupazione, è fondamentale svolgere un'analisi longitudinale dell'offerta di lavoro. In particolare, risulta di interesse analizzare come la durata della disoccupazione influisca sulla probabilità di transitare verso un'altra condizione (occupazione o inattività).

Le fonti da cui attingere informazioni per poter ricostruire le durate delle storie occupazionali sono varie, ma è indubbio che la numerosità del campione e la quantità di informazioni raccolte dalla nuova Rilevazione Continua delle Forze Lavoro (RCFL) condotta dall'Istat sono nettamente superiori a qualsiasi altra indagine svolta sul mercato del lavoro italiano.

La RCFL, fonte dei dati congiunturali sul mercato del lavoro dal primo trimestre del 2004, presenta rispetto alla vecchia Rilevazione Trimestrale delle Forze Lavoro novità decisive per poter svolgere questa analisi in quanto la struttura dell'indagine pone maggiore attenzione sulla possibilità di ottenere dati di tipo longitudinale.

In particolare, ogni individuo della famiglia soggetta a rilevazione ha un codice identificativo unico che non varia nel tempo e ciò permette la costruzione di un dataset longitudinale tramite abbinamento deterministico dei dati sezionali. La frequenza di errori di abbinamento, verificata a partire dai dati invarianti dell'analisi, risulta minima (nell'ordine dello 0.05%).

Nel ricostruire la durata del periodo di disoccupazione di un individuo è necessario confrontare le informazioni retrospettive fornite nelle differenti rilevazioni ed è pertanto opportuno eseguire un preventivo controllo sulla loro qualità. Si sono analizzate non risposte e incoerenze longitudinali sulla durata di ricerca di lavoro, sull'inizio dell'attività lavorativa (per gli occupati) o sulla sua fine (per gli inoccupati).

Queste analisi hanno evidenziato che la qualità dei dati non è molto elevata, in particolare per quanto riguarda la coerenza longitudinale: la durata della ricerca di lavoro è coerente, in rilevazioni consecutive, solo in poche occasioni e anche la risposta sull'anno in cui l'intervistato ha iniziato a lavorare sono spesso non coerenti (anche oltre il 50% in alcuni casi) con la condizione occupazionale emersa nella rilevazione precedente.

La qualità dei dati dipende, inoltre, in modo rilevante dalle caratteristiche individuali del soggetto a cui è riferita l'intervista: i giovani e le persone con un maggiore livello di istruzione tendono ad avere un tasso di risposta più alto e a commettere un minor numero di incoerenze.

La ridotta qualità dei dati è un dato inatteso dal momento che l'utilizzo del questionario elettronico, altra importante novità introdotta dalla RCFL, dovrebbe garantire un buon controllo longitudinale delle risposte. I dati a disposizione, che riguardano il 2004 e il 2005, presentano da questo punto di vista un'anomalia che sembra poter essere causata dalla tecnica di indagine. Il tasso di risposta sull'inizio e sulla fine dell'attività lavorativa sono significativamente diversi a seconda dell'anno in cui si svolge l'intervista. In particolare per quanto riguarda l'anno di inizio dell'occupazione è evidente, anche dalla sola analisi descrittiva, che nel 2005 la tecnica CATI (l'intervista telefonica che è utilizzata dalla seconda rilevazione in poi) prevede una domanda di tipo confermativo rispetto alla risposta precedente per chi era già occupato, mentre nel 2004 questa opzione non era attivata.

Una volta analizzati i tassi di risposta e le incoerenze dei dati retrospettivi si è andati a ricostruire la durata degli episodi di disoccupazione. Il campione considerato per l'analisi è composto dagli individui abbinati in tre rilevazioni consecutive, che vengono pertanto seguiti per un anno. Dove possibile sono stati utilizzati gli abbinamenti tra le prime tre interviste in quanto i dati raccolti con la tecnica CAPI (intervista faccia a faccia), che avviene solo nella prima rilevazione dell'indagine, sembrano avere un maggiore tasso di risposta sulle domande di interesse per la ricostruzione.

Nel campione sono state quindi ricostruite le durate degli episodi di disoccupazione degli individui per cui sono presenti tutti i dati retrospettivi necessari e che risultano essere coerenti. Nel caso di episodi di disoccupazione multipli nel corso dell'anno di osservazione si è scelto di considerare solo il primo.

Dal momento che le analisi precedenti hanno suggerito la presenza di correlazione tra caratteristiche individuali e qualità delle risposte si è confrontata la composizione del campione "selezionato" con quella del campione che comprende tutti i disoccupati in almeno una rilevazione. Le caratteristiche che possono essere controllate non sono risultate essere distribuite in modo diverso nel sottocampione: si può quindi affermare che la presenza di selection bias sia almeno limitata.

Nella ricostruzione della durata della disoccupazione si è dovuto tener conto dell'assenza, nei dati a disposizione, della settimana a cui fa riferimento l'intervista. Questa carenza informativa non consente di determinare l'esatta durata dell'episodio di disoccupazione per

tutti gli individui: è possibile determinare, in quasi la metà dei casi, solo i limiti (inferiore e superiore) della lunghezza dell'episodio.

Si è quindi specificato un modello adatto ad analizzare delle durate che sono, in parte, non determinate univocamente e che utilizza tutte le informazioni che si hanno a disposizione.

Si è successivamente verificato come cambiano le stime del modello al variare delle informazioni disponibili per determinare la durata degli episodi di disoccupazione. Utilizzare la media dei due limiti come durata unica degli episodi sembra essere una buona soluzione che non porta a particolari cambiamenti nelle stime, mentre utilizzare, ad esempio, solo il limite superiore porta a stime molto diverse. Cambiando la quantità di informazione con l'aggiunta di un dato ricostruito dall'Istat, le stime ottenute dal modello non sono significativamente diverse. Si è così ottenuta una parziale conferma di correttezza del metodo di ricostruzione scelto per la durata degli episodi.

Dal punto di vista sostanziale le evidenze emerse dall'analisi portano ad affermare che esiste una dipendenza positiva dalla durata: la probabilità di uscire dalla disoccupazione, condizionata all'essere ancora disoccupati, cresce con l'aumentare della durata dell'episodio. Inoltre, se si assume che i rischi siano competitivi, si stima che il rischio di transitare verso l'occupazione è diverso da quello di transitare verso l'inattività (nel campione il secondo è mediamente più elevato) e la loro relazione cambia a seconda delle caratteristiche personali dell'individuo. Ad esempio la probabilità di trovare un'occupazione per un disoccupato diplomato del centro-nord con meno di 45 anni è sensibilmente superiore a quella di passare alla condizione di inattività. Per una disoccupata di più di 45 anni con lo stesso livello di istruzione e residente nelle stesse regioni, invece, il rischio di transitare verso l'inattività è di molto maggiore a quello di iniziare una nuova occupazione.

La struttura della tesi rispecchia, in gran parte, il processo con cui si è svolta l'analisi.

Nel secondo capitolo si sono inizialmente riassunte le principali caratteristiche della Rilevazione Continua delle Forze Lavoro, evidenziando le differenze rispetto alla Rilevazione Trimestrale svolta in precedenza. Si è inoltre descritto l'abbinamento longitudinale dei dati, verificando la correttezza del processo confrontando le informazioni anagrafiche degli abbinati, e si è quindi svolta l'analisi sulla presenza di risposte mancanti ed eventuali incoerenze sui dati retrospettivi.

Nel terzo capitolo, dopo una breve descrizione dei metodi di analisi della dipendenza dalla durata nell'ambito socio-economico, si è illustrato come sono state ricostruite le lunghezze degli episodi di disoccupazione utilizzando le informazioni disponibili. Dopo una descrizione

dei modelli utilizzati per analizzare queste durate si sono commentate le stime, anche tenendo conto della destinazione dopo l'uscita dalla disoccupazione. Si sono inoltre confrontati i modelli stimati utilizzando informazioni diverse . Infine si sono descritti i possibili sviluppi dell'analisi effettuata.

CAPITOLO 2

La Rilevazione Continua delle Forze Lavoro

La Rilevazione delle Forze Lavoro (RFL) è la principale fonte informativa congiunturale sullo stato e la dinamica del mercato di lavoro. I dati raccolti nella rilevazione permettono di fornire con cadenza trimestrale le stime ufficiali del numero di occupati e disoccupati presenti in Italia e un insieme di altri indicatori, come il tasso di attività o di disoccupazione, che sono fondamentali per comprendere se c'è equilibrio o meno nell'offerta di lavoro.

Le stime prodotte dalla RFL sono relative a tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente all'estero, e possono essere disaggregate sia per zona territoriale, fino a un livello provinciale, sia per caratteristiche socio-demografiche come l'età, il sesso e il livello di istruzione, ma anche per stato civile e tipologia familiare.

La Rilevazione delle Forze Lavoro è stata effettuata per la prima volta nel 1959, ma da allora ha subito innumerevoli cambiamenti sia nel metodo di rilevazione sia nei contenuti, per adeguarsi alla continua evoluzione della società e del mercato del lavoro come, ad esempio, l'aumento della diffusione dei contratti di lavoro a tempo determinato nell'ultima decade. Altre motivazioni che hanno portato a dei cambiamenti nell'indagine sono legate alla necessità di raccogliere informazioni sempre più dettagliate e di maggiore qualità per poter comprendere meglio le dinamiche socio-economiche. Per un approfondimento sull'evoluzione della RFL si veda Di Pietro (1993) e, per gli sviluppi più recenti "La nuova rilevazione sulle forze lavoro", Istat(2004).

L'ultimo di questi numerosi cambiamenti è avvenuto nel 2004 col passaggio dalla Rilevazione Trimestrale delle Forze Lavoro alla Rilevazione Continua delle Forze Lavoro. Questo passaggio, che ha coinvolto sia i contenuti sia i metodi dell'indagine, è stato reso necessario da direttive di tipo amministrativo provenienti dalla Comunità Europea e in particolare dall'Eurostat, ma è stato anche sfruttato dall'Istat per aumentare la quantità di informazioni raccolte e soddisfare le crescenti esigenze informative di chi è chiamato a decidere politiche nazionali e locali riguardanti il mercato del lavoro.

L'ampliamento dei contenuti informativi ha dovuto tener conto di vari aspetti:

- la maggiore complessità del mercato del lavoro dovuto alla sempre più diffusa flessibilità, con conseguente varietà nella tipologia dei contratti di lavoro;

- il progressivo fenomeno di terziarizzazione. ovvero del peso sempre più importante dei servizi nell'economia e quindi nella domanda di lavoro;
- l'evoluzione della struttura demografica della popolazione italiana che, oltre al fenomeno dell'invecchiamento, ha registrato negli ultimi anni un aumento di flussi migratori dall'estero e una crescente mobilità interna causata soprattutto da ragioni occupazionali.

2.1 Le novità della RCFL

Le novità della Rilevazione Continua delle Forze Lavoro (RCFL) rispetto alla Rilevazione Trimestrale delle Forze Lavoro (RTFL) riguardano sia il modo in cui sono rilevati i dati sia la struttura e le informazioni raccolte col questionario. Nel seguito si descrivono brevemente le novità principali con particolare riferimento a quelle di interesse per questo studio. Per maggiori dettagli si veda Istat(2004).

Il cambiamento principale, evidente anche dal nome della rilevazione, è il periodo di riferimento in cui si svolge l'indagine. Nella RTFL le interviste erano effettuate nella prima settimana dei quattro trimestri, ovvero nella prima settimana (senza festività) dei mesi di gennaio, aprile, luglio e ottobre. Nella RCFL il periodo di riferimento sono tutte le tredici settimane di ogni trimestre: una famiglia viene intervistata una volta per trimestre, ma le interviste nel campione sono distribuite tra le tredici settimane.

Altra differenza nei metodi di rilevazione è la modalità con cui è sottoposto il questionario: si è passati da un'intervista faccia a faccia con supporto cartaceo in ogni rilevazione a una tecnica di intervista mista: la prima rilevazione per chi entra nell'indagine è di tipo CAPI (Computer Assisted Personal Interviewing) e consiste in un'intervista faccia a faccia svolta con questionario elettronico, dalla seconda rilevazione in poi l'intervista è di tipo CATI (Computer Assisted Telephonic Interviewing) e viene svolta via telefono, sempre con supporto informatico per il rilevatore.

Queste novità hanno portato l'Istat a cambiare anche la rete di rilevazione, che è passata dall'essere gestita dai comuni tramite propri dipendenti a una rilevazione professionale formata direttamente dall'Istat.

Le differenze sulle informazioni raccolte non riguardano solo la loro quantità e qualità ma anche il tipo di quesiti che determinano la condizione lavorativa dell'intervistato. Le norme

dell'Eurostat impongono infatti un approccio oggettivo, ovvero una verifica di condizioni oggettive dell'intervistato in merito alla sua posizione nel mercato del lavoro, mentre nella RTFL la condizione lavorativa si basava anche sull'autopercezione dell'intervistato.

2.1.1 Piano di campionamento

Non ci sono stati invece cambiamenti di tipo strutturale sul disegno campionario della rilevazione, anche se nella RCFL si deve tenere conto dei vincoli sulla conduzione dell'indagine e quindi le famiglie e i comuni campione devono essere uniformemente ripartiti nelle tredici settimane di riferimento. Inoltre il disegno campionario deve essere strutturato in modo da rispettare alcuni vincoli sulla precisione delle stime e sulla procedura di riporto all'universo basati sulla distribuzione della popolazione per sesso e età proveniente da fonti esterne (l'anagrafe). Questi vincoli però non sono di grande interesse per il nostro studio in quanto si analizzeranno solo i dati campionari senza utilizzare pesi per riportarli alla popolazione.

In modo simile a quanto accadeva nella RTFL il disegno campionario consiste nel campionamento a due stadi. Nel primo stadio si selezionano i comuni campione (le unità primarie) che sono stratificati all'interno di ciascuna provincia sulla base della loro dimensione demografica. Nel secondo stadio viene estratta una lista anagrafica casuale di famiglie per ciascun comune campione selezionato nel primo stadio. Da notare che vengono considerate le famiglie "di fatto": un insieme di persone coabitanti, legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi. L'indagine passa dalle sei possibili relazioni di parentela presenti nella RTFL alle diciassette presenti nella RCFL.

Nel primo stadio del campionamento ha un ruolo fondamentale la stratificazione dei comuni a livello provinciale, che viene effettuata sulla base della dimensione demografica e di una soglia determinata sul numero di famiglie da intervistare. Questa stratificazione porta ad identificare due tipi di comuni a seconda degli abitanti. I comuni che superano un determinato numero di abitanti (che varia da provincia a provincia) sono autorappresentativi (AR) e sono tutti selezionati come comune campione. I comuni che hanno una popolazione che non supera la soglia vengono raggruppati in strati e da ciascun strato è estratto un solo comune campione. La probabilità di un comune di essere incluso nel campione sarà proporzionale alla sua dimensione demografica.

Una volta selezionati i comuni campione si deve fare un'ulteriore stratificazione, introdotta solo con la nuova rilevazione, di tipo mensile affinché tutti gli strati territoriali siano presenti almeno una volta in ciascun mese del trimestre. I comuni AR di maggiori dimensioni campionarie saranno presenti nella rilevazione ogni settimana, mentre i restanti comuni saranno rilevati una volta al mese (tre al trimestre).

Il secondo stadio del campionamento consiste nell'estrazione delle famiglie campione dalle liste anagrafiche comunali. L'estrazione genera un campione casuale semplice di famiglie in forma di quartine, in quanto per ogni famiglia estratta si estraggono le tre successive in qualità di famiglie sostitutive, da utilizzare al posto di quella base (estratta) in caso non possa o voglia partecipare all'indagine. Risulta importante, per il nostro studio, il fatto che ogni famiglia intervistata è identificata da un determinato codice che riguarda ogni passaggio dell'estrazione oltre alla regione e alla provincia. Inoltre ogni componente della famiglia ha un codice identificativo unico e invariato nelle quattro interviste basato sul rapporto di parentela con la persona di riferimento e sull'età anagrafica. Questo codice unito a quelli familiari permette di formare una stringa che consente di identificare in modo univoco l'intervistato nelle quattro rilevazioni e rende quindi possibile la ricostruzione di dati longitudinali per ogni individuo con un abbinamento deterministico

La presenza di un codice identificativo è un'altra importante novità della RCFL: nella Rilevazione Trimestrale delle Forze Lavoro si poteva identificare solo la famiglia, non l'individuo all'interno della famiglia se non ricorrendo ad algoritmi di tipo probabilistico (si veda Paggiaro e Torelli, 2002, e la relativa bibliografia).

Dopo essere state estratte, le famiglie campione vengono suddivise in gruppi di rotazione che prevedono la partecipazione all'indagine per due trimestri consecutivi e, dopo una pausa di due trimestri, una nuova partecipazione per altri due trimestri consecutivi. Le interviste nei trimestri consecutivi sono svolte a tre mesi di distanza l'una dall'altra, mentre l'intervista dopo i due trimestri di pausa avviene a nove mesi di distanza da quella precedente. Se, ad esempio, una famiglia è entrata nell'indagine nel primo trimestre del 2004 ed è intervistata a febbraio le sue successive interviste avverranno nel maggio dello stesso anno, nel febbraio e nel maggio del 2005.

In ogni trimestre il campione su cui sono calcolate le stime è formato da quattro gruppi di rotazione in uguale proporzione, formato da famiglie che partecipano rispettivamente alla prima, seconda, terza e quarta intervista dell'indagine.

Il numero di famiglie intervistate realmente è inferiore a quello stabilito dal disegno campionario (76872 le famiglie che dovrebbero essere soggette a rilevazione ogni trimestre) in quanto si verifica una certa frequenza di mancata risposta totale della famiglia, ovvero non è possibile intervistare nessuna famiglia di quelle presenti nella quartina per irreperibilità o per rifiuto. Le mancate risposte totali possono essere dovute a un fenomeno di attrition di alcune tipologie di famiglie, anche se questa evenienza è poco probabile dal momento che le quartine sono costruite in modo casuale e difficilmente la tipologia di famiglie è la stessa per tutti gli elementi della quartina. Nel processo di riporto all'universo l'Istat utilizza, comunque, una correzione per le mancate risposte totali.

2.1.2 Le tecniche di rilevazione

Un'altra delle novità introdotte dalla Rilevazione Continua delle Forze Lavoro è stata, come già detto, la creazione di una rete di rilevazione che dipende direttamente dall'Istat e non più dai comuni. L'organizzazione di una nuova rete di rilevazione si è resa necessaria per standardizzare la selezione e la formazione degli intervistatori e per avere un monitoraggio uniforme della loro attività, dal momento che è il comportamento dei rilevatori a causare la maggior parte degli errori non campionari. Le due tecniche di intervista sono gestite in modo diverso nella rete di rilevazione: il sistema CAPI dipende direttamente dall'Istat con le interviste effettuate da più di trecento rilevatori professionali presenti nelle varie province, mentre il sistema CATI è affidato a una ditta esterna i cui rilevatori sono comunque formati dall'Istat.

L'utilizzo di questa tecnica mista (CAPI nella prima intervista, CATI nelle successive) oltre a dare dei vantaggi economici (adottare solo la tecnica CAPI comporterebbe alti costi di gestione dell'indagine) dovrebbe dare dei vantaggi anche nella qualità dei dati raccolti.

Dopo aver avuto un contatto diretto nella prima intervista, in cui si è aiutato il rispondente a comprendere le domande presenti nel questionario in modo di limitare il numero di non risposte, nelle interviste successive l'intervista telefonica permette di limitare la molestia statistica e di aumentare le possibilità di controllo sui dati forniti in passato.

La rilevazione CATI oltre a essere più tempestiva ed economica consente infatti di sfruttare al meglio il potenziale del questionario elettronico. Ogni intervistato ha un percorso personalizzato all'interno del questionario a seconda delle risposte e questo permette di limitare considerevolmente il numero di quesiti: il questionario, composto in totale di oltre

200 quesiti, può ridursi a un'intervista di sole dieci domande (se l'intervistato è disabile) e comunque non supera mai le novanta domande complessive (per chi è occupato e ha un secondo lavoro). Secondo l'Istat la media dei tempi di rilevazione nella fase CATI è di dieci minuti per un'intervista in cui sono rivolti 35 quesiti. Per ridurre il disagio provocato dall'intervista il questionario elettronico dà anche la possibilità di somministrare una serie di quesiti chiedendo conferma della risposta data nella precedente occasione di rilevazione. Tra questi quesiti, oltre a quelli riguardanti le caratteristiche individuali stabili nel tempo, ci dovrebbe essere anche l'anno di inizio dell'attività lavorativa per chi è occupato in interviste consecutive.

L'intervista effettuata con tecnica CATI, grazie al software costruito appositamente, permette inoltre un controllo di completezza e coerenze in tempo reale: è possibile verificare l'ammissibilità della risposta prima che sia posta la domanda successiva. L'ammissibilità di una risposta è controllata da due tipi di regole: hard e soft. Se il rispondente infrange una regola del primo tipo viene riposta la domanda per risolvere l'incoerenza, se viene infranta una regola soft, che indica una risposta improbabile ma non del tutto incompatibile, l'intervista prosegue lasciando la possibilità al software di eseguire una correzione automatica.

La possibilità di effettuare domande di conferma e l'applicazione di regole di controllo dovrebbe portare l'indagine a raccogliere dati di buona qualità, ovvero le risposte dovrebbero essere coerenti sia all'interno del questionario sezionale sia dal punto di vista longitudinale. Si vedrà, nel corso di questo capitolo, che questa aspettativa non sarà del tutto soddisfatta, almeno per quanto riguarda la coerenza longitudinale: non sempre le possibilità offerte dal questionario elettronico sono state sfruttate appieno nell'implementazione sul campo dell'indagine.

Un'altra novità nella tecnica di intervista è l'introduzione della modalità "non sa" in parecchi quesiti (quasi in tutti quelli retrospettivi sulla vita lavorativa, ad esempio), ma in molti casi questa modalità è ammessa solo se chi risponde non è il diretto interessato. Ciò avviene per limitare la presenza di risposte "forzate" che risulterebbero probabilmente errate e allo stesso tempo evitare le risposte dette "di rifugio" se l'intervistato è direttamente il soggetto della rilevazione.

Il questionario elettronico permette inoltre la registrazione di ogni evento che riguarda l'intervista (data, ora, interruzioni) e quindi garantisce un maggiore controllo sull'operato dei rilevatori, che dovranno compilare una specifica scheda di rifiuto nel caso in cui il questionario si interrompa in maniera definitiva.

2.1.3 Il questionario

Il questionario è composto da una scheda generale e da dodici sezioni, ognuna caratterizzata da un preciso obiettivo conoscitivo (Figura 2.1)

Nella scheda generale sono presenti alcune caratteristiche di tutti gli individui componenti della famiglia (data di nascita, sesso, stato civile, livello di istruzione) e le relazioni di parentela rispetto alla persona di riferimento.

Nei dati resi disponibili dall'Istat per la nostra analisi non c'è la sezione A: chi risponde ai quesiti. Ciò non ci consentirà di fare un'analisi completa sulla qualità dei dati raccolti, in particolare sulla risposta "non sa" ad alcune domande, in quanto non sarà chiaro se saranno dovute a un'effettiva mancanza di risposta del soggetto dell'intervista o di chi sta rispondendo al suo posto.

Nelle sezioni successive inizia il questionario vero e proprio che riguarda il rapporto dell'intervistato col mondo del lavoro fino alla sezione I. Le ultime tre sezioni riguardano nuovamente notizie anagrafiche e dettagli tecnici sul come si è svolta la rilevazione.

Le sezioni determinanti per stabilire la condizione lavorativa nella settimana di riferimento dell'intervista sono la sezione B, in cui si stabilisce se l'individuo è occupato basandosi su dati oggettivi, e la sezione F in cui si accerta se l'individuo inoccupato sta cercando attivamente lavoro.

Nel nostro studio saranno considerate anche altre due sezioni del questionario: la sezione C sull'attività lavorativa principale e la sezione E sulla precedente esperienza lavorativa. In queste due parti sono infatti presenti dei quesiti che forniscono dati retrospettivi importanti per ricostruire la durata degli episodi di disoccupazione.

Questionario individuale: diagramma di flusso

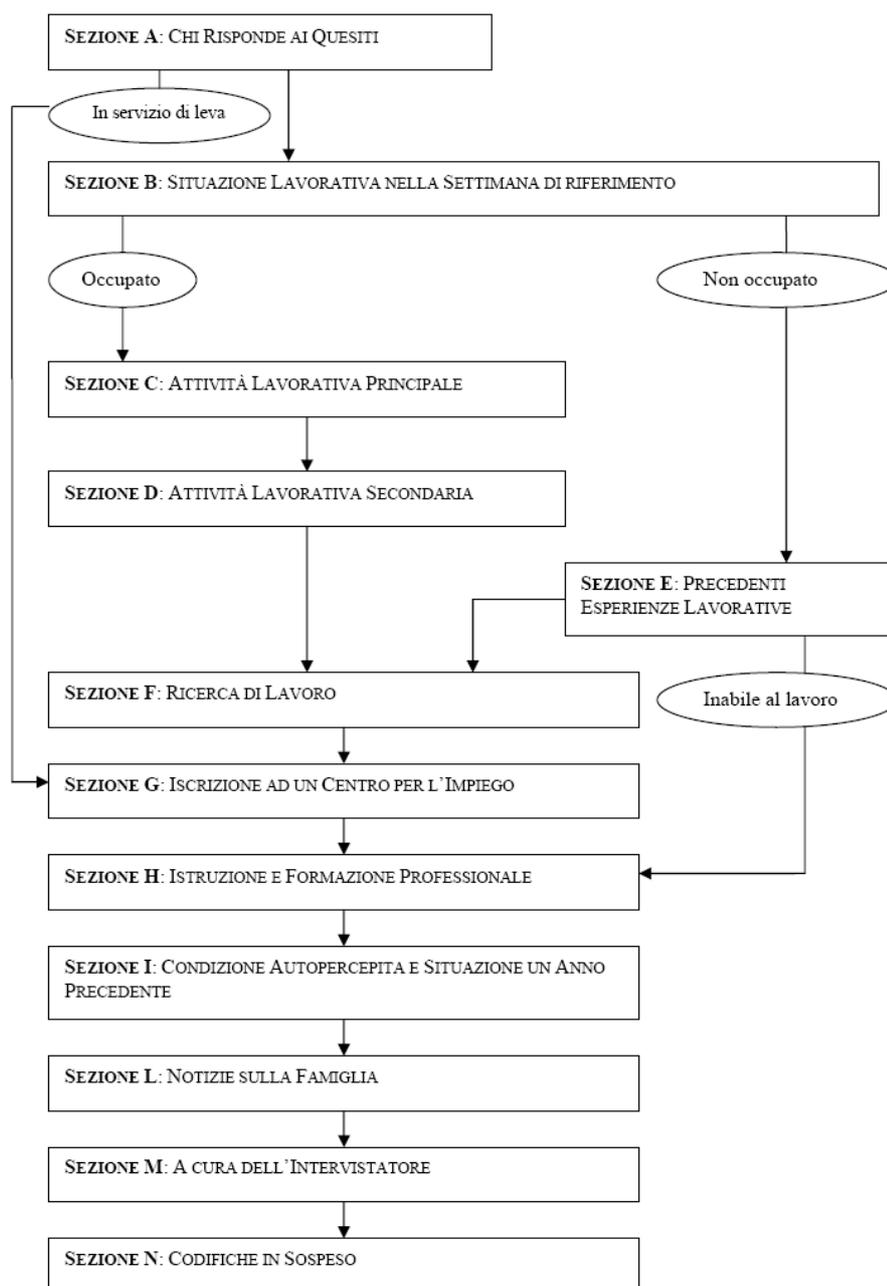


Figura 2.1: diagramma di flusso del questionario individuale della Rilevazione Continua Forze Lavoro.

2.2 L'abbinamento longitudinale

I dati disponibili nello studio provengono dalle rilevazioni effettuate negli otto trimestri degli anni 2004 e 2005. Per ogni individuo presente nel dataset sono disponibili alcuni elementi

che ne permettono l'identificazione in diverse occasioni: i codici di regione, provincia e comune di residenza, la quartina di estrazione, l'ordine della famiglia all'interno della quartina e il numero individuale all'interno del nucleo familiare. Non sono purtroppo presenti campi che permettano di capire a quale delle quattro possibili rilevazioni stia rispondendo l'intervistato, per cui tale informazione è disponibile esclusivamente a posteriori per gli individui abbinati.

Per gli obiettivi del nostro studio si è scelto di abbinare tre occasioni (consecutive) delle quattro disponibili per ogni individuo in modo da poter seguire la sua vita lavorativa per un buon periodo di tempo (un anno) e disporre comunque di un campione abbastanza numeroso. Gli abbinamenti a tre rilevazioni ricostruibili nei due anni e che coinvolgono campioni distinti sono cinque (otto se si considerano anche gli abbinamenti a tre che portano informazioni in parte ridondanti), mentre se si utilizzassero gli abbinamenti a quattro a fronte di un trimestre aggiuntivo nella finestra di osservazione si potrebbero ricostruire solo tre abbinamenti.

Inizialmente si sono quindi creati tutti gli abbinamenti a tre possibili abbinando gli individui che presentano gli stessi valori nei codici di identificazione, alternativamente nelle prime o nelle ultime tre rilevazioni presenti nel corso dei due anni disponibili. Il numero degli individui presenti nei vari abbinamenti varia tra le 35150 e le 37800 unità (il 20-22% del totale degli intervistati in ogni trimestre di indagine: la quota teorica dovrebbe essere un quarto). La numerosità degli individui che si riescono ad abbinare tende ad aumentare nei campioni di rotazione che sono intervistati per la prima volta negli ultimi trimestri del 2004, a testimonianza che c'è un continuo miglioramento da parte dell'Istat nel limitare le cause che portano le famiglie a uscire dall'indagine.

I vari campioni ottenuti da questo processo di selezione possono, comunque presentare fenomeni di *attrition* rispetto al campione trimestrale sezionale da cui si attingono le stime ufficiali dell'Istat. Nei vari campioni prodotti dagli abbinamenti non sono presenti, infatti, coloro che cambiano la propria residenza tra una rilevazione e l'altra: il rischio è quello di non considerare nelle analisi la parte di popolazione maggiormente portata alla mobilità occupazionale in quanto alcuni trasferimenti potrebbero essere per motivi di lavoro. Escono dal campione anche famiglie che interrompono l'indagine per altri motivi (continua irreperibilità o indisponibilità a continuare). Interruzione che potrebbe non essere casuale ma correlata con determinate caratteristiche dei componenti della famiglia.

In entrambi i casi nel campione ottenuto dall'abbinamento il rischio di selection bias è concreto anche se non è verificato empiricamente e se ne dovrà, eventualmente, tener conto in sede di analisi. Per la RTFL un'analisi dettagliata del problema, con una parziale soluzione in termini di riporto alla popolazione compresente, si trova in Istat (2002).

Gli abbinamenti a tre episodi possibili nei due anni, data la struttura dell'indagine, sono otto. Alcuni di questi provengono però dallo stesso campione di famiglie e quindi ci si trova davanti a informazioni ridondanti. I gruppi di rotazione indipendenti da cui si possono trarre abbinamenti a tre episodi sono solo cinque:

- gli intervistati nei primi due trimestri del 2004 e del 2005;
- gli intervistati nel secondo e terzo trimestre del 2004 e del 2005;
- gli intervistati nel terzo e quarto trimestre del 2004 e del 2005;
- gli intervistati nel quarto trimestre del 2003 (osservazione non disponibile), nel primo e nel quarto del 2004 e nel primo del 2005;
- gli intervistati nell'ultimo trimestre del 2004, nel primo e nel quarto del 2005 e nel primo del 2006 (osservazione non disponibile).

Per i primi 3 campioni possono essere effettuati abbinamenti sia tra le prime tre osservazioni sia tra le ultime tre dal momento che sono disponibili tutte le rilevazioni. Nella ricostruzione delle storie lavorative operato nel capitolo 3 si utilizzeranno, dove possibile, le prime tre occasioni.

Tuttavia nello studiare la qualità dei dati (espressa, ad esempio, dal numero di mancate risposte ad alcuni quesiti) utilizzare solo gli abbinamenti a tre sarebbe stato troppo vincolante e non avrebbe permesso di sfruttare tutta l'informazione disponibile.

Si è quindi deciso di abbinare le rilevazioni a due a due, determinando però in che punto queste avvengono all'interno dell'indagine. In questo modo è possibile individuare coppie di rilevazioni effettuate in trimestri consecutivi provenienti da gruppi di rotazione differenti dai cinque visti in precedenza e da cui sono stati tratti gli abbinamenti a tre.

Individui intervistati in trimestri consecutivi possono appartenere a due campioni indipendenti di famiglie poiché possono essere o le prime due o le ultime due rilevazioni a cui un individuo partecipa. Gli individui che sono presenti negli abbinamenti a tre già effettuati potranno essere identificati e quindi sarà possibile selezionare le risposte di chi è presente in solo due trimestri nei due anni di indagine disponibili.

Ad esempio gli intervistati nel primo e nel secondo trimestre del 2004 possono far parte di due diversi campioni a seconda che siano le prime o le ultime due rilevazioni per l'individuo.

Dal momento che sono già stati abbinati coloro che sono stati intervistati anche nel primo trimestre del 2005 (e che quindi sono entrati nell'indagine all'inizio del 2004) si possono dividere i due campioni. Gli individui appaiati, che non sono intervistati nel primo trimestre 2005, saranno per la maggior parte alla loro terza o quarta intervista. Purtroppo una piccola frazione sarà composta dalle famiglie che escono, per vari motivi, dall'indagine dopo due sole rilevazioni, ma si è scelto di tollerare questa distorsione nella costruzione del campione per poter avere indicazioni sulla qualità dei dati in più fasi dell'indagine.

Utilizzando questo metodo, infatti, si possono confrontare in quasi tutti i trimestri risposte provenienti da campioni indipendenti e quindi verificare la presenza di trend o anomalie nei dati prodotti da modifiche in itinere dell'indagine, ovvero se rilevazioni diverse effettuate nello stesso trimestre o anno portano a risultati differenti.

2.2.1 Abbinamenti e informazioni invarianti

Una volta abbinati i dati, utilizzando i codici creati dall'Istat per identificare in modo univoco ogni individuo del campione in tutte le rilevazioni, può essere utile verificare la bontà dell'abbinamento.

Una verifica immediata è data dal confrontare informazioni invarianti degli individui nelle tre occasioni. Fra i dati anagrafici si analizzano la data di nascita e il sesso dell'individuo, dati che sono richiesti all'inizio dell'intervista, nella scheda generale, e che devono essere comunicati per ogni membro della famiglia.

Per i tre elementi che compongono la data di nascita (giorno, mese, anno) l'individuo può anche non rispondere, anche se, nel caso in cui non risponda all'anno di nascita, è comunque obbligato a indicare l'età. La presenza di non risposte è molto probabilmente legata al fatto che nella scheda generale le notizie anagrafiche possono essere date da un singolo individuo per tutti i membri della famiglia. La variabile che indica chi abbia risposto effettivamente ai quesiti non è però disponibile nei dati oggetto di questo studio.

La possibilità di non rispondere ad alcune domande anagrafiche comporta il dover considerare due casi di possibile incongruenza nelle risposte: la risposta diversa in senso stretto e l'assenza di risposta in almeno una delle due occasioni che si vanno a confrontare.

Nel primo caso l'incongruenza è evidente, nel secondo non è possibile affermare che l'individuo coinvolto nelle due interviste abbia la stessa data di nascita e che quindi l'abbinamento sia sicuramente avvenuto in modo corretto.

L'analisi sulla correttezza dei dati è effettuata sugli abbinamenti a tre occasioni non ridondanti in quanto saranno l'oggetto dell'analisi svolta nel capitolo 3. Inoltre utilizzando gli abbinamenti a tre occasioni si possono verificare meglio eventuali anomalie per gli stessi individui a seconda della distanza tra interviste consecutive.

Nella Tabella 2.1 sono presenti le frequenze percentuali negli abbinamenti degli errori e delle non risposte. È stato considerato errore una risposta diversa in due osservazioni consecutive, mentre è stato considerato non risposta il caso in cui l'individuo non ha risposto alla domanda in almeno una delle due occasioni.

RISPOSTA ABBINAM.	GIORNO		MESE		ANNO		SESSO
	ERRORI	NON RISP.	ERRORI	NON RISP.	ERRORI	NON RISP.	ERRORI
TRA LE PRIME DUE RILEVAZIONI:							
041/042/051	26 0.07	90 0.26	19 0.05	62 0.18	119 0.34	24 0.07	12 0.03
042/043/052	22 0.06	136 0.38	13 0.04	107 0.30	129 0.36	42 0.12	11 0.03
043/044/053	23 0.06	298 0.83	16 0.04	164 0.46	114 0.32	45 0.12	7 0.02
044/051/054	20 0.05	326 0.86	16 0.04	229 0.61	147 0.39	42 0.11	6 0.02
TRA LA SECONDA E LA TERZA RILEVAZIONE:							
041/044/051	30 0.08	120 0.33	21 0.06	95 0.26	80 0.22	28 0.08	5 0.01
041/042/051	29 0.08	76 0.22	25 0.07	49 0.14	69 0.20	17 0.05	6 0.02
042/043/052	31 0.09	114 0.32	23 0.06	89 0.25	76 0.21	28 0.08	7 0.02
043/044/053	11 0.03	277 0.77	7 0.02	145 0.40	65 0.18	31 0.09	9 0.02
044/051/054	2 0.01	317 0.84	3 0.01	220 0.58	50 0.13	33 0.09	8 0.02
TRA LA TERZA E LA QUARTA RILEVAZIONE:							
041/044/051	2 0.01	109 0.30	5 0.01	85 0.24	22 0.06	19 0.05	3 0.01
042/051/052	7 0.02	76 0.21	5 0.01	48 0.14	22 0.06	14 0.04	9 0.03
043/052/053	4 0.01	118 0.33	4 0.01	93 0.26	26 0.07	24 0.07	3 0.01
044/053/054	6 0.02	249 0.67	2 0.01	126 0.34	25 0.07	21 0.06	1 0.00

Tabella 2.1: frequenza assoluta e percentuale di errori e non risposte alle domande su data di nascita e sesso per gli individui presenti negli abbinamenti a tre. Nell'abbinamento le prime due cifre indicano l'anno, la terza il trimestre. In neretto i trimestri confrontati

Dalla Tabella 2.1 si può notare come le modalità di risposta abbiano frequenza diversa a seconda della domanda: il numero di non risposte è molto più elevato quando si chiede il giorno o il mese di nascita rispetto a quando si chiede l'anno. In alcuni abbinamenti si ha al massimo una risposta sul giorno di nascita in due occasioni vicine in più dello 0.8% dei casi, mentre per quanto riguarda l'anno di nascita non si supera mai lo 0.12%.

Questo andamento si riflette sulla frequenza di risposte diverse in occasioni contigue: si verificano giorni o mesi di nascita diversi in un numero sempre inferiore allo 0.1% dei casi mentre si riscontra un anno di nascita differente con frequenze anche superiori allo 0.3%.

La diversità nelle modalità di risposta sull'anno di nascita può essere dovuto sia al fatto che il questionario prevede di dichiarare comunque l'età, sia alla maggiore facilità, nel caso in cui chi risponde non sia il soggetto a cui è rivolta l'intervista, nell'indicare l'anno di nascita rispetto al giorno e al mese.

Si può notare, inoltre, come le incongruenze sull'anno di nascita siano più elevate tra le prime due rilevazioni dell'indagine, dove sono più dello 0.3% sul totale, mentre calano vistosamente in quelle successive: tra le ultime due rilevazioni il numero di incongruenze è un quinto di quello osservato tra le prime due. Ciò può essere una conseguenza dell'intervista di tipo CATI e quindi della possibilità di controllare in questo tipo di quesiti la congruenza con le risposte precedenti mentre viene effettuato il questionario.

Se si confrontano i vari abbinamenti sembra esserci un aumento della presenza di non risposte sul giorno e mese di nascita a seconda dell'inizio dell'indagine. Tra le prime due interviste c'è almeno una non risposta sul giorno di nascita nello 0.26% dei casi se sono svolte nei primi due trimestri del 2004, mentre la frequenza sale allo 0.86% se sono coinvolti gli intervistati dell'ultimo trimestre del 2004 e del primo del 2005. Questo aumento di non risposte è presente anche nella domanda sul mese di nascita, ma le cause non sono di facile comprensione dal momento che non risultano cambiamenti nel questionario o nell'intervista. Inoltre non sembrano esserci elementi sufficienti per affermare la presenza di un trend dal momento che tra la seconda e la terza rilevazione del campione entrato nell'indagine nell'ultimo trimestre del 2003 (l'abbinamento 041/044/051 in Tabella 2.1) si registrano più non risposte rispetto al campione entrato nel primo trimestre 2004 (041/042/051).

Le incongruenze sul sesso dell'individuo non superano mai lo 0.03% del campione, anche se nei pochi casi in cui si hanno risposte diverse risulta difficile poter distinguere l'errore durante la rilevazione da un errore in fase di abbinamento e quindi da un qualche codice identificativo attribuito in modo errato.

Gli altri casi in cui non è chiaro se l'abbinamento è stato eseguito in modo corretto sono quelli in cui gli elementi diversi nella data di nascita sono più di uno oppure quelli in cui la presenza di non risposte è tale da non rendere possibile un confronto tra le due date di nascita (Tabella 2.2).

	ERRORI			NON RISPOSTE		
ABBINAMENTO	1	2	3	1	2	3
TRA LE PRIME DUE RILEVAZIONI:						
041/042/051	91 0.26	14 0.04	15 0.04	36 0.10	34 0.10	24 0.07
042/043/052	108 0.30	13 0.04	10 0.03	44 0.12	59 0.16	41 0.11
043/044/053	89 0.25	11 0.03	14 0.04	154 0.43	109 0.30	45 0.12
044/051/054	120 0.32	18 0.05	9 0.02	120 0.32	180 0.48	39 0.10
TRA LA SECONDA E LA TERZA RILEVAZIONE:						
041/044/051	50 0.14	9 0.03	21 0.06	33 0.09	63 0.18	28 0.08
041/042/051	38 0.11	14 0.04	19 0.05	35 0.10	28 0.08	17 0.05
042/043/052	43 0.12	12 0.03	21 0.06	39 0.11	54 0.15	28 0.08
043/044/053	53 0.15	6 0.02	6 0.02	150 0.42	105 0.29	31 0.09
044/051/054	46 0.12	3 0.01	1 0.00	117 0.31	177 0.47	33 0.09
TRA LA TERZA E LA QUARTA RILEVAZIONE:						
041/044/051	20 0.06	3 0.01	1 0.00	32 0.09	62 0.17	19 0.05
042/051/052	15 0.04	5 0.01	3 0.01	36 0.10	30 0.08	14 0.04
043/052/053	21 0.06	2 0.01	3 0.01	45 0.13	59 0.16	24 0.07
044/053/054	19 0.05	4 0.01	2 0.01	143 0.38	95 0.25	21 0.06

Tabella 2.2: frequenza assoluta e percentuale di errori e di non risposte alle tre domande sulla data di nascita in interviste consecutive. I tre errori indicano la presenza di date di nascita diverse nelle due rilevazioni, mentre tre non risposte indicano che l'individuo non ha risposto in almeno un'occasione a ciascun elemento che forma la data di nascita.

Dalla Tabella 2.2 si può notare che negli abbinamenti sia la frequenza in cui le date di nascita tra osservazioni sono completamente diverse, sia quella in cui sono diversi due elementi su

tre non superano mai lo 0.06%. Inoltre questa tipologia di errori è quasi assente tra la terza e la quarta rilevazione.

Le situazioni in cui si può, quindi, ragionevolmente ipotizzare che l'abbinamento sia errato – la presenza di un solo errore nella data di nascita può essere dovuto a qualche distrazione – sono presenti in quantità trascurabile nel campione.

Sono più frequenti le mancate risposte a più elementi della data di nascita. Tra la prima e la seconda rilevazione la data di nascita non è confrontabile, ovvero per ogni quesito non è presente una risposta in almeno una delle due occasioni, anche nello 0.12% dei casi. Ancora più numerosi (in un abbinamento quasi lo 0.5%) sono i casi in cui tra due occasioni non sono confrontabili almeno due elementi su tre, come ci si poteva attendere dal maggior numero di non risposte per giorno e mese di nascita rispetto all'anno. L'impossibilità di confrontare le date di nascita non può comunque portare ad affermare che l'abbinamento sia stato eseguito in modo errato.

Concludendo, l'analisi sui dati anagrafici degli individui presenti nel dataset longitudinale, formato in base ai codici identificativi forniti dall'Istat, mostra che i problemi di abbinamento, se ci sono, sono minimi e si può affermare che per la quasi totalità degli individui le notizie anagrafiche sono le stesse in ogni occasione in cui risultano intervistati.

Nei pochi casi in cui sembra possibile che siano state abbinate rilevazioni effettuate su individui diversi, dal momento che la data di nascita o il sesso dell'individuo risultano incongruenti tra le osservazioni, non è comunque possibile affermare con certezza che si tratti di un errore nell'abbinamento e non un errore nel riportare le informazioni da parte del rilevatore, in quanto risposte ad altri quesiti presenti nella scheda generale (ad esempio il titolo di studio) risultano, a volte, congruenti.

Non sembra quindi essere necessario escludere dal campione che si andrà ad analizzare i pochi individui che presentano incongruenze anagrafiche o mancate risposte nelle interviste effettuate nell'indagine.

2.3 Condizione lavorativa al momento dell'intervista

Un primo passo per ricostruire la lunghezza degli episodi di disoccupazione consiste nell'individuare la condizione degli individui in ogni rilevazione: occupato, disoccupato, inattivo. Tali stati sono definiti a partire dalle definizioni seguenti (Istat, 2004).

Occupato: un individuo risulta occupato (Figura 2.2) se ha più di 15 anni e nella settimana di riferimento dichiara, in alternativa, di:

- aver svolto almeno un'ora di lavoro in un'attività che preveda un corrispettivo monetario o in natura;
- aver svolto almeno un'ora di lavoro non retribuito nella ditta di un familiare nella quale collabora;
- essere assente dal lavoro con opportune condizioni:
 - a) la sua assenza non è superiore ai tre mesi oppure se durante l'assenza continua a percepire almeno il 50% della retribuzione nel caso in cui sia un lavoratore dipendente;
 - b) nel suo periodo di assenza conserva la sua attività nel caso in cui sia un lavoratore autonomo;
 - c) la sua assenza non supera i tre mesi nel caso in cui sia un coadiuvante familiare.

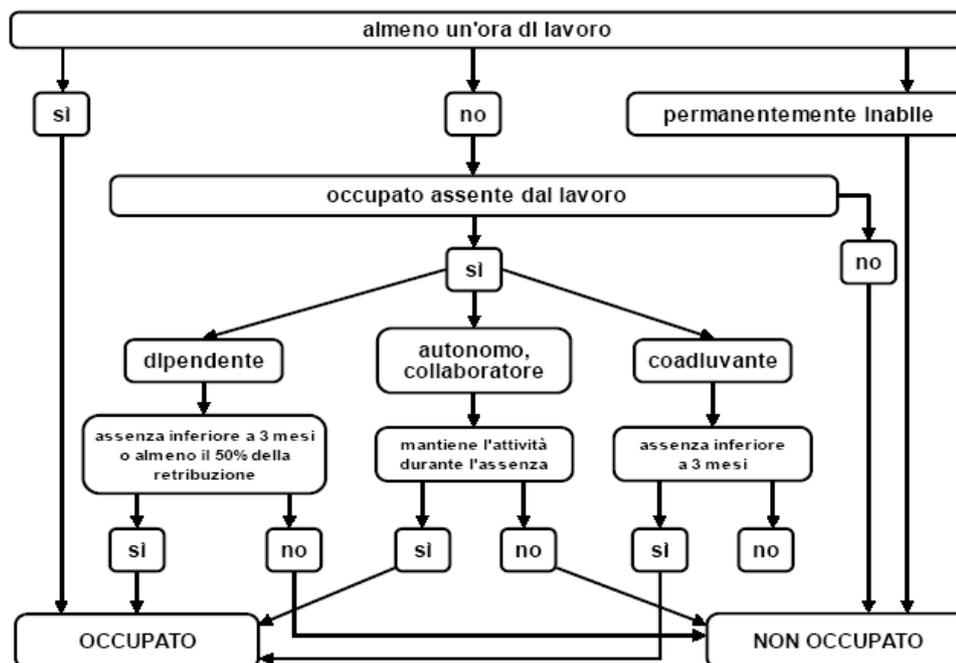


Figura 2.2: diagramma di flusso per individuare gli occupati a partire dalle domande presenti nel questionario.

Disoccupato: un individuo è classificato come disoccupato (o in cerca di occupazione, Figura 2.3) se ha un'età compresa tra i 15 e i 74 anni, non è occupato e dichiara di:

- aver effettuato almeno un'azione attiva di ricerca di lavoro nei trenta giorni precedenti all'intervista ed essere disponibile a lavorare entro le due settimane successive all'intervista;
- iniziare un lavoro entro tre mesi dalla data dell'intervista ma essere comunque disponibile a lavorare entro le due settimane successive all'intervista, qualora fosse possibile anticipare l'inizio del lavoro.

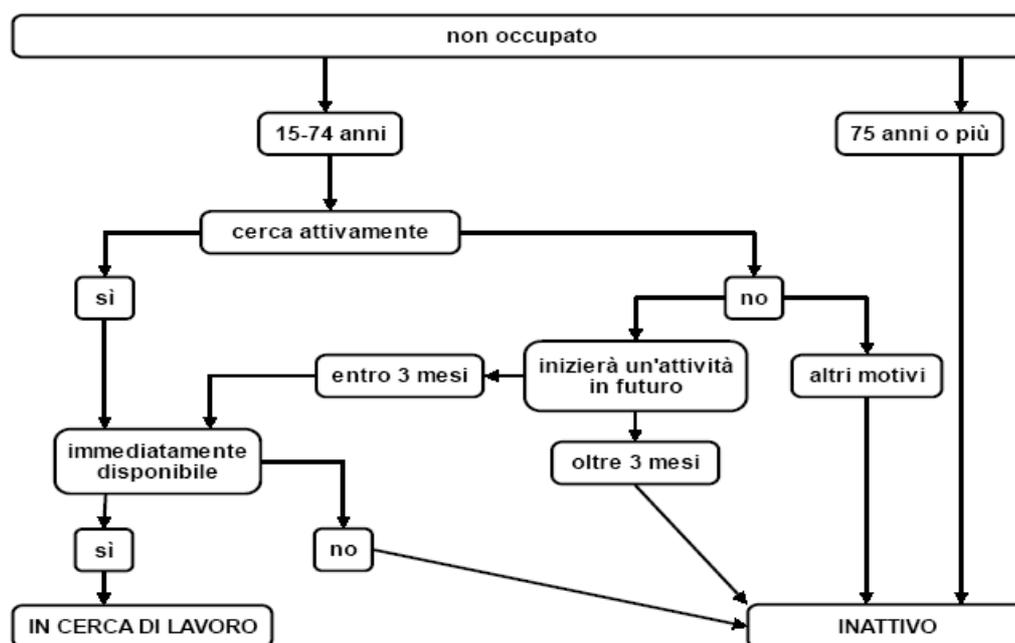


Figura 2.3: diagramma di flusso per individuare, tra i non occupati, coloro che sono alla ricerca di lavoro a partire dalle domande presenti nel questionario.

Inattivo: un individuo è considerato inattivo se non appartiene alle forze lavoro, ovvero le sue risposte ai quesiti non lo classificano né come occupato né come disoccupato.

STATO TRIMESTRE	OCCUPATI	DISOCCUPATI	INATTIVI
2004			
I	36.22	3.36	60.42
II	36.92	3.15	59.94
III	37.30	2.94	59.76
IV	36.99	3.23	59.78
2005			
I	36.31	3.27	60.42
II	36.59	3.00	60.42
III	36.30	2.73	60.97
IV	36.29	3.16	60.56

Tabella 2.3: condizione lavorativa nei due anni di indagine, dati sezionali.

Nella Tabella 2.3 sono riportate le frequenze percentuali delle condizioni rispetto al mercato del lavoro nei dati sezionali nel 2004 e 2005. Le stime sono calcolate senza utilizzare i pesi di riporto all'universo utilizzati dall'Istat per tener conto delle differenze tra la composizione della popolazione e quella del campione, per cui possono differire dalle stime ufficiali. Si noti che questo vale per tutte le analisi presentate nel seguito.

Dalla Tabella 2.3 si può notare che la percentuale di disoccupati nei dati sezionali sembra avere un andamento stagionale, con un calo nel terzo trimestre, e che nel 2005 c'è una diminuzione dei disoccupati rispetto al trimestre dell'anno precedente a fronte di un aumento degli inattivi.

2.4 Informazioni retrospettive sulla vita lavorativa: tassi di risposta

Ogni individuo intervistato, a seconda dello stato in cui si trova, risponde a domande di tipo retrospettivo relative alla ricerca di lavoro o all'attività lavorativa.

Le risposte a questi quesiti sono fondamentali per ricostruire la vita lavorativa di ogni individuo nel periodo in cui è soggetto di rilevazione (e nei mesi immediatamente precedenti) in quanto vanno ad integrarsi alle informazioni sezionali. È possibile in questo modo, ad esempio, individuare eventuali cambiamenti di stato avvenuti nel periodo che intercorre tra le interviste e determinare la lunghezza del periodo di disoccupazione per chi risulta disoccupato dalla prima rilevazione.

L'integrazione di dati sezionali e retrospettivi comporta spesso la necessità di risolvere problemi legati a risposte errate, incoerenti o mancanti. Per una rassegna dei problemi di misura nei dati campionari si veda Bound et Al. (2001).

Purtroppo nei dati in esame le informazioni retrospettive, oltre a non avere l'affidabilità dei dati sezionali in quanto condizionate dal ricordo di eventi passati dell'intervistato, sono affette da altri due problemi: la possibilità di non rispondere al quesito e la non reperibilità del mese in cui è stata effettuata l'intervista.

La possibilità di non rispondere al quesito da un lato limita l'eventualità di fornire informazioni non corrette, probabili nel caso che chi risponde non sia l'oggetto della rilevazione, dall'altro risultano irreperibili alcuni dati essenziali per determinare, ad esempio, la durata della disoccupazione, che è la quantità di maggiore interesse per questo studio. Diventa quindi fondamentale analizzare la frequenza con cui gli intervistati rispondono alle

domande retrospettive ed evidenziare eventuali caratteristiche comuni tra chi non risponde, sia per ipotizzare possibili cause alla mancanza di risposte, sia per individuare frazioni di popolazione che potrebbero essere trattate in modo differente in successivi modelli basati sulle durate ricostruite degli episodi lavorativi.

Il non conoscere in quale mese del trimestre si è svolta la rilevazione (o comunque la settimana di riferimento dell'indagine) rende impossibile determinare con certezza i mesi trascorsi dalla fine dell'ultima attività lavorativa se l'individuo è inoccupato, dall'inizio dell'attuale attività se l'individuo è occupato. Nel questionario infatti si chiede l'anno e il mese di questi due eventi e quindi può essere ricostruito solo un intervallo entro cui ci sarà la vera distanza tra il momento in cui è svolta l'intervista e l'evento in questione.

Per svolgere questa analisi, oltre alla presentazione delle principali evidenze descrittive, si utilizza un modello logit in quanto la variabile dipendente sulla presenza o meno della risposta è dicotomica. Per la teoria sui modelli per variabili dipendenti binarie si rimanda, ad esempio, a Wooldridge (2000).

Brevemente, la stima dei parametri di interesse si ottiene massimizzando la verosimiglianza del seguente modello:

$$P[y = 1 | x_1, x_2, \dots, x_k] = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)$$

dove $G(\cdot)$ è la funzione logistica:

$$G(z) = \frac{e^z}{1 + e^z}$$

Le variabili esplicative dicotomiche che verranno utilizzate nell'analisi riguarderanno:

- in quale delle quattro rilevazioni dell'indagine è stata effettuata l'intervista
- l'anno in cui si è svolta la rilevazione
- l'età (divisa in classi)
- il sesso
- la regione di residenza (raggruppate in cinque macroregioni: Nord Ovest, Nord Est, Centro, Sud, Isole)
- il titolo di studio (raggruppato in cinque livelli, il più alto raggruppa tutti i titoli di tipo universitario)

In tutte le analisi l'individuo considerato come baseline è un uomo tra i 26 e i 35 anni con un diploma di scuola superiore, residente nelle regioni centrali ed è intervistato nel 2004.

2.4.1 Durata della ricerca di lavoro

La domanda di maggiore interesse per determinare la lunghezza del periodo di disoccupazione è quella posta nella sezione F del questionario, denominata “*Ricerca di lavoro*”, e viene rivolta a tutti gli intervistati che hanno più di 15 anni di età e che dichiarano di cercare lavoro: *Da quanti mesi sta cercando lavoro?*

Dal momento che cercare attivamente un’occupazione è una condizione imprescindibile per essere considerato disoccupato la presenza della risposta a questo quesito è fondamentale per ricostruire la quantità di interesse, anche se non c’è alcuna certezza sul fatto che la ricerca sia stata attiva e continuativa nell’arco di tempo dichiarato.

Nel seguito si considera il campione composto da chi risulta disoccupato nella rilevazione, escludendo quindi chi risulta occupato o inattivo ma afferma comunque di cercare un lavoro.

RILEVAZIONE TRIMESTRE	I	II	III	IV
2004				
I	127 10.87	134 12.05	151 11.82	
II	105 9.12	147 14.22	187 13.73	169 14.32
III	128 11.91	134 13.41	175 13.38	147 12.25
IV	101 8.36	111 10.38	99 9.24	139 9.37
MEDIA	10.02	12.48	12.19	11.78
2005				
I	199 11.64	122 10.43	122 11.65	98 9.64
II	174 12.68	211 14.18	130 13.47	142 14.88
III	154 13.50	164 14.08	119 12.66	133 15.54
IV		130 10.81	106 9.42	124 11.12
MEDIA	12.48	12.48	11.70	12.61

Tabella 2.4: frequenza, assoluta e percentuale, di non risposte alla domanda “*Da quanti mesi sta cercando lavoro?*” da parte di chi nella rilevazione risulta disoccupato. Le frequenze sono suddivise per rilevazione e trimestre.

La frequenza di mancate risposte (Tabella 2.4) è nell'ordine del 10-15% e non sembra avere un andamento particolare a seconda del trimestre, dell'anno o del fatto che l'individuo sia alla prima o all'ultima intervista.

L'unica anomalia sembra essere una percentuale di non risposte inferiore nell'ultimo trimestre del 2004 rispetto agli altri trimestri in cui è svolta l'indagine, ma non ci sono né evidenze né spiegazioni plausibili all'esistenza di questa differenza.

Nel considerare il numero di mancate risposte si deve tener conto che dalla lettura del questionario sembra emergere che la possibilità di non rispondere sia riservata solo a chi risponde per conto di terzi, mentre il diretto interessato non può esimersi dal fornire una risposta.

VAR. INDIPENDENTE = P[NON RISPONDERE]		N° CAMPIONE = 34966		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	II RILEVAZIONE	.253	.068	.000
	III RILEVAZIONE	.223	.066	.001
	IV RILEVAZIONE	.188	.070	.007
ANNO	ANNO 2005	.246	.068	.000
RILEVAZIONI 2005 (INTERAZIONI)	(II RILEV.)*2005	-.253	.093	.006
	(III RILEV.)*2005	-.294	.094	.002
	(IV RILEV.)*2005	-.168	.097	.084
AREA RESIDENZA	NORD OVEST	-.088	.068	.195
	NORD EST	.392	.068	.000
	SUD	.045	.056	.420
	ISOLE	.115	.061	.060
ETÀ	15-25	-.214	.043	.000
	36-45	-.130	.047	.006
	46-55	-.062	.058	.290
	56-75	-.153	.089	.086
LIVELLO ISTRUZIONE	ELEMENTARE	.076	.061	.219
	MEDIA	.018	.040	.641
	DIPLO. 3 ANNI	.016	.078	.837
	LAUREA	-.138	.062	.027
SESSO	DONNA	-.379	.034	.000
	COSTANTE	-1.984	.077	.000

Tabella 2.5: modello logit per la probabilità di non rispondere al quesito sulla durata della ricerca di lavoro da parte dei disoccupati.

Dalle stime riportate nella Tabella 2.5 si può ricavare che per l'individuo di baseline la probabilità di non rispondere al quesito sulla durata di ricerca di lavoro stimata dal modello è del 12.1%.

Dai risultati della regressione si può notare, inoltre, come sia statisticamente significativo in quale rilevazione, nel corso dell'indagine, è effettuata l'intervista. Nel 2004 la probabilità di non rispondere nella prima rilevazione è inferiore rispetto alle successive (la probabilità

stimata di non rispondere passa, per la baseline, al 15.1% nella seconda rilevazione del 2004). Nel 2005, invece, la probabilità stimata di non rispondere non varia molto nelle diverse rilevazioni e solo il coefficiente per chi è intervistato per la terza volta sembra significativamente diverso da quello stimato per la seconda rilevazione del 2004 (i coefficienti sono: 0.246 se è intervistato per la prima o seconda volta, 0.175 nella terza e 0.266 nella quarta). Infatti per l'individuo di baseline intervistato nel 2005 la probabilità stimata risulta essere attorno al 15% nelle prime due e nell'ultima rilevazione, al 14.1% nella terza. Al netto dalle caratteristiche individuali il tasso di risposta è quindi più elevato nella prima intervista nel 2004.

Oltre al momento in cui si entra nell'indagine la probabilità di dare una risposta al quesito è condizionata anche da altri fattori. Si può notare, ad esempio, come chi risiede al Nord Est risulti avere una probabilità di non risposta superiore a chi risiede nelle altre regioni italiane. Il fattore che più incide è però il sesso di chi risponde. Una donna con le stesse caratteristiche dell'individuo di baseline alla prima intervista del 2004 ha una probabilità di non risposta stimata solo del 8.6% (rispetto al 12% di un uomo). Anche l'età incide sulla probabilità di ottenere una risposta sulla durata della ricerca di lavoro da parte dei disoccupati: i disoccupati sotto i 25 anni rispondono più frequentemente a parità delle altre condizioni.

Non ci sono grandi diversità per quanto riguarda l'istruzione, se si esclude un tasso di risposta significativamente superiore per chi ha un'istruzione di tipo universitario.

Dal momento che, molto probabilmente, la possibilità di non risposta a questo quesito è consentita solo se chi risponde non è l'oggetto dell'intervista, l'interpretazione dell'impatto delle caratteristiche individuali non è facile in quanto è influenzata dal chi risponde realmente alle domande.

Altri fattori possono influire sulla probabilità di rispondere o meno alla domanda sulla ricerca di lavoro. Date le diverse modalità dell'intervista tra la prima (di persona) e le altre rilevazioni (telefoniche) e il possibile diverso atteggiamento verso la ricerca di lavoro da parte di chi è inoccupato da un breve periodo rispetto a chi lo è da più tempo (in questo caso non tanto l'atteggiamento solo dell'individuo, ma anche dei suoi familiari), può essere interessante confrontare la frequenza di non risposte in base alla condizione di lavoro dell'intervistato nella rilevazione precedente (Tabella 2.6).

RILEVAZIONE TRIMESTRE	II		III		IV	
	SI	NO	SI	NO	SI	NO
2004						
ERA OCCUPATO?	SI	NO	SI	NO	SI	NO
II	32 19.16	115 13.26			36 19.15	133 13.41
III	44 27.85	90 10.70			35 15.91	112 11.43
IV	25 13.02	86 9.81	20 8.23	79 9.54	40 12.99	99 8.42
MEDIA	19.54	11.26	8.23	9.54	15.50	10.93
2005						
I	35 16.75	87 9.05	50 17.99	72 9.36	31 16.94	67 8.03
II	40 16.88	171 13.67	42 18.92	88 11.84	24 18.05	118 14.37
III	35 16.59	129 13.52	40 18.10	79 10.99	41 25.00	92 13.29
IV	33 16.75	97 9.64	32 14.75	74 8.15	27 14.36	97 10.46
MEDIA	16.74	11.60	17.48	9.97	18.41	11.42

Tabella 2.6: frequenza assoluta e percentuale di non risposte alla domanda "Da quanti mesi sta cercando lavoro?" da parte di chi nella rilevazione risulta disoccupato. Le frequenze sono condizionate all'occupazione nella rilevazione precedente.

Dalla Tabella 2.6 si può notare come l'essere occupati nella rilevazione precedente porta i disoccupati a rispondere con più difficoltà al quesito sulla durata in mesi della ricerca di lavoro a prescindere dal momento in cui sono intervistati (l'unica eccezione è la terza rilevazione nel 2004).

Questa diversità nel tasso di risposta potrebbe essere dovuta al modo in cui è svolta l'intervista: dalla seconda rilevazione l'intervista è CATI ed è possibile che porre per la prima volta una domanda al telefono, soprattutto se chi risponde non è il diretto interessato, porti a incrementare le non risposte.

Per provare a comprendere meglio l'influenza dello stato precedente sulla probabilità di risposta si può inserire nel modello di regressione logistica la variabile PREINOC che assume il valore uno se l'individuo risultava non occupato nello stato precedente, zero altrimenti. L'analisi va ovviamente limitata alle rilevazioni successive alla prima.

VAR. INDIPENDENTE = P[NON RISPONDERE]		N° CAMPIONE = 21082		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	-.385	.119	.001
	IV RILEVAZIONE	-.090	.074	.225
ANNO	ANNO 2005	.072	.106	.499
RILEVAZIONI 2005 (INTERAZIONI)	(III RILEV.)*2005	.278	.136	.041
	(IV RILEV.)*2005	.109	.098	.266
STATO PRECEDENTE	PREINOC	-.395	.083	.000
	PREINOC*2005	-.120	.103	.242
AREA RESIDENZA	NORD OVEST	-.161	.088	.066
	NORD EST	.266	.089	.003
	SUD	.077	.072	.283
	ISOLE	.170	.079	.030
ETA'	15-25	-.233	.056	.000
	36-45	-.135	.060	.024
	46-55	-.148	.076	.051
	56-75	-.141	.116	.223
LIVELLO ISTRUZIONE	ELEMENTARE	.047	.080	.561
	MEDIA	.042	.051	.411
	DIPLO. 3 ANNI	.037	.100	.713
	LAUREA	-.141	.079	.074
SESSO	DONNA	-.310	.044	.000
	COSTANTE	-1.42	.112	.000

Tabella 2.7: modello logit per la probabilità di non rispondere al quesito sulla durata della ricerca di lavoro da parte dei disoccupati. Non sono presenti i dati della prima intervista.

La baseline in questo modello (presentato nella Tabella 2.7) è intervistata nella seconda rilevazione dell'indagine che la coinvolge ed è transitata dall'occupazione alla disoccupazione tra la prima e seconda intervista.

Avendo eliminato la prima rilevazione l'andamento della frequenza delle non risposte non è influenzato dall'anno in cui si svolge l'intervista. Il fatto che sia significativo il coefficiente che riguarda la terza rilevazione del 2004 non è di grande interesse in quanto è presente solo un trimestre di osservazione.

Come ci si poteva attendere dai dati presenti in Tabella 2.6 l'essere inoccupati nella rilevazione precedente ha un effetto fortemente negativo sulla probabilità di non rispondere. Un effetto che non varia nei due anni di rilevazione. La probabilità di non ottenere una risposta stimata per la baseline è del 19.5% se era occupato nell'osservazione precedente, mentre è del 14% se non lo era.

I coefficienti per le altre variabili coinvolte nell'analisi sono simili a quelli ottenuti nel modello precedente, l'unica variazione si ha sul titolo di studio: l'essere laureati non è più statisticamente significativo.

2.4.2 Fine dell'attività lavorativa

Un'altra domanda funzionale a ricostruire la vita lavorativa è quella sull'anno - ed il mese se ha smesso di lavorare da meno di tre anni – in cui si è terminata l'ultima attività lavorativa: *“In che anno ha smesso di lavorare?”*

Il campione esaminato sarà quello composto da chi risulta non essere occupato e risponde affermativamente al quesito presente all'inizio della sezione E del questionario: *“Nel corso della sua vita, Lei ha mai svolto un lavoro?”*

RILEVAZIONE TRIMESTRE	I	II	III	IV
2004				
I	763 6.50	2,585 21.98	2,961 21.87	
II	934 8.20	2,837 24.55	3,174 23.54	3,744 27.77
III	1,539 13.10	2,879 25.41	3,433 23.97	3,491 26.15
IV	1,491 12.30	3,104 26.49	3,096 26.13	4,032 27.49
MEDIA	10.06	24.60	23.81	27.15
2005				
I	1,523 10.20	1,598 12.97	2,891 24.40	3,104 25.83
II	1,295 9.65	1,667 11.18	2,907 25.38	2,828 24.09
III	1,813 14.23	1,389 10.29	3,077 26.13	2,785 23.94
IV		1,852 14.54	1,605 13.03	3,002 24.63
MEDIA	11.27	12.17	22.11	24.63

Tabella 2.8: frequenza assoluta e percentuale di non risposte alla domanda sull'anno in cui ha terminato la sua ultima attività per gli inoccupati che hanno già avuto un'esperienza lavorativa,

Dalla Tabella 2.8 si può notare come la frequenza di non risposte sull'anno di fine lavoro aumenti di molto tra la prima e la seconda rilevazione per le interviste effettuate nel 2004, mentre nel 2005 l'aumento avviene tra la seconda e la terza rilevazione.

Per comprendere meglio questo differente andamento tra i due anni di indagine si possono controllare le stime prodotte da un modello che assume come variabile dipendente la presenza o meno di risposte al quesito (Tabella 2.9).

VAR. INDIPENDENTE = P[NON RISPONDERE]		N° CAMPIONE =377565		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	II RILEVAZIONE	1.190	.020	.000
	III RILEVAZIONE	1.122	.019	.000
	IV RILEVAZIONE	1.322	.020	.000
ANNO	ANNO 2005	.103	.023	.000
RILEVAZIONI 2005 (ITERAZIONI)	(II RILEV.)*2005	-1.080	.029	.000
	(III RILEV.)*2005	-.211	.028	.000
	(IV RILEV.)*2005	-.256	.028	.000
AREA RESIDENZA	NORD OVEST	.029	.014	.043
	NORD EST	.020	.015	.194
	SUD	.374	.014	.000
	ISOLE	.233	.018	.000
ETA'	15-25	-.699	.065	.000
	36-45	.429	.040	.000
	46-55	.569	.037	.000
	56-65	.854	.035	.000
	>65	1.987	.034	.000
LIVELLO ISTRUZIONE	ELEMENTARE	.935	.018	.000
	MEDIA	.421	.020	.000
	DIPLO. 3 ANNI	.254	.033	.000
	LAUREA	-.341	.036	.000
SESSO	DONNA	.469	.009	.000
	COSTANTE	-4.749	.040	.000

Tabella 2.9: modello logit per la probabilità di non rispondere al quesito sull'anno in cui hanno terminato l'ultima occupazione da parte degli inoccupati che hanno dichiarato di aver già avuto un'esperienza lavorativa,.

Per la regressione presente nella Tabella 2.9 la baseline è un non occupato che ha dichiarato di aver svolto un lavoro nel corso della sua vita con le usuali caratteristiche individuali.

Anche tenendo conto delle caratteristiche individuali, nel 2004 la probabilità di non rispondere è molto più alta nelle ultime tre interviste dell'indagine, mentre nel 2005 la frequenza di non risposte aumenta solo nella terza e quarta intervista.

Non è chiaro quali sono le cause di questa diversità tra il 2004 e il 2005. L'aumento tra la prima e la seconda rilevazione registrato nel 2004 potrebbe essere giustificato dal passaggio dal CAPI al CATI, ma nel 2005 l'aumento si registra tra la seconda e la terza intervista in cui non risulta esserci alcun cambiamento nella modalità di indagine.

La frequenza di non risposte, oltre al momento in cui si svolge l'indagine, è fortemente influenzata dalle caratteristiche individuali dell'intervistato, soprattutto dall'età. La frequenza

stimata di non risposte nella prima intervista per un giovane diplomato con meno di 25 anni residente nel centro è appena dello 0.4%, mentre per un uomo con più di 65 anni è del 6%. Un altro elemento che condiziona la probabilità di non rispondere è il grado di istruzione: più è alto maggiore è il tasso di risposta.

Dal modello emerge che l'intervistato con la minor probabilità di rispondere al quesito è una donna oltre i 65 anni di età residente in una regione del sud e che non è andata oltre la quinta elementare. La frequenza stimata di non risposte è del 27% nella prima intervista effettuata nel 2004 e del 58% quando è intervistata per la quarta volta nel 2004.

La probabilità di non rispondere è quindi molto variabile e dipende sia dal momento in cui si svolge la rilevazione (anno e a quale punto dell'indagine) sia dalle caratteristiche personali dell'individuo.

RILEVAZIONE TRIMESTRE	II		III		IV	
	SI	NO	SI	NO	SI	NO
2004						
ERA OCCUPATO?						
II	38 7.2	2,799 25.36			57 9.91	3,687 28.57
III	68 10.66	2,811 26.30			49 6.47	3,442 27.34
IV	55 8.04	3,049 27.63	41 5.44	3,055 27.53	54 5.78	3,978 28.97
MEDIA	8.73	26.43	5.44	27.53	7.06	28.31
2005						
I	46 6.58	1,552 13.35	40 4.83	2,851 25.87	27 4.64	3,077 26.91
II	42 6.05	1,625 11.44	45 5.87	2,862 26.79	20 5.33	2,808 24.71
III	48 6.88	1,341 10.48	37 4.77	3,040 27.63	20 3.76	2,765 24.91
IV	40 6.05	1,812 15.01	38 5.44	1,567 13.49	27 4.55	2,975 25.66
MEDIA	6.40	12.48	5.21	23.28	4.51	25.55

Tabella 2.10: frequenza assolute e percentuali di non risposte alla domanda sull'anno in cui ha terminato la sua ultima attività in base allo stato precedente.

Dalla Tabella 2.10 è evidente che l'aver smesso di lavorare da poco tempo condiziona la possibilità di rispondere al quesito: chi esce dall'occupazione nel periodo che intercorre tra due rilevazioni dichiara l'anno in cui ha terminato la propria attività in più del 90% dei casi. Mentre chi nell'occasione precedente era già inoccupato non risponde in più del 20% dei

casi, con l'unica eccezione degli intervistati per la seconda volta nel 2005 che in media non rispondono nel 12.5%. Questo dato mostra come la fine dell'attività lavorativa non sia una risposta "a conferma" per chi era già inoccupato, altrimenti ci si sarebbe aspettati una percentuale di non risposte molto inferiore per gli inoccupati. Si va quindi a stimare un regressione di tipo logit introducendo la variabile binaria PREOC che assume il valore uno se l'individuo era occupato nell'intervista precedente (Tabella 2.11).

VAR. INDIPENDENTE = P[NON RISPONDERE]		N° CAMPIONE = 236373		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	.018	.026	.503
	IV RILEVAZIONE	.080	.018	.000
ANNO	ANNO 2005	-1.035	.019	.000
RILEVAZIONI 2005 (INTERAZIONI)	(III RILEV.)*2005	.788	.032	.000
	(IV RILEV.)*2005	.876	.025	.000
STATO PRECEDENTE	PREOC	-.523	.058	.000
	PREOC*2005	.128	.077	.097
AREA RESIDENZA	NORD OVEST	-.017	.017	.313
	NORD EST	.004	.018	.810
	SUD	.319	.017	.000
	ISOLE	.162	.022	.000
ETA'	15-25	-.800	.081	.000
	36-45	.425	.048	.000
	46-55	.570	.045	.000
	56-65	.809	.042	.000
	>65	1.956	.041	.000
LIVELLO ISTRUZIONE	ELEMENTARE	.943	.021	.000
	MEDIA	.424	.023	.000
	DIPLO. 3 ANNI	.247	.039	.000
	LAUREA	-.314	.042	.000
SESSO	DONNA	.468	.011	.000
	COSTANTE	-3.431	.047	.000

Tabella 2.11: modello logit per la probabilità di non rispondere al quesito sull'anno in cui hanno terminato l'ultima occupazione da parte degli inoccupati che hanno dichiarato di aver già avuto un'esperienza lavorativa. Sono considerate le ultime tre interviste dell'indagine.

L'analisi delle stime presenti nella Tabella 2.11 conferma che essere occupati nella rilevazione precedente diminuisce significativamente la probabilità stimata di non rispondere, indifferentemente dall'anno in cui si svolge l'indagine: una diplomata inoccupata tra i 25 e i 35 anni residente nel sud ha una probabilità stimata di non rispondere nella sua seconda intervista del 2004 del 6.6% se era inoccupata anche nelle prima intervista, del 4% se era occupata. Risulta evidente anche dal modello la diversità tra la seconda intervista effettuata nel 2005 e le altre rilevazioni, successive alla prima, svolte nei due anni. Le altre caratteristiche individuali influiscono sulle non risposte in modo simile a quello visto nel modello precedente: il tasso di risposta diminuisce con l'avanzare dell'età e aumenta con un maggior grado di istruzione.

La differenza nelle risposte tra chi era occupato o meno nella rilevazione precedente può essere dovuta alla difficoltà di ricordare un episodio lontano nel tempo da parte di chi ha smesso di lavorare da molto tempo – e il fatto che il numero di non risposte sia molto più elevato tra chi ha più di 65 anni lo conferma –, ma anche a una modalità dell’indagine a noi ignota come induce a pensare la netta differenza tra la seconda rilevazione del 2005 e il resto del campione nella frequenza di mancate risposte proprio da parte di chi non era occupato.

2.4.3 Inizio dell’attività lavorativa

Un altro quesito fondamentale per ricostruire gli episodi lavorativi è quello sull’anno (e il mese) in cui è iniziata l’attività lavorativa, formulato in modo diverso se l’individuo svolge un lavoro dipendente (“*In che anno ha cominciato a lavorare per il datore di lavoro attuale?*”) o autonomo (“*In che anno ha cominciato questo lavoro?*”). Il campione analizzato sarà composto dagli occupati.

RILEVAZIONE TRIMESTRE	I	II	III	IV
2004				
I	417 3.30	1,317 10.09	1,477 10.32	
II	548 4.10	1,320 10.36	1,624 10.23	1,519 10.51
III	660 4.98	1,466 10.99	1,657 9.61	1,734 10.88
IV	629 4.52	1,302 9.89	1,291 9.83	1,740 10.16
MEDIA	4.24	10.33	9.99	10.51
2005				
I	705 4.00	286 2.10	396 3.18	486 3.77
II	578 3.73	405 2.31	510 3.86	252 1.98
III	752 5.23	306 2.00	443 3.39	287 2.20
IV		362 2.53	261 1.89	254 1.88
MEDIA	4.28	2.24	3.06	2.45

Tabella 2.12: frequenza assoluta e percentuale di non risposte sull’anno di inizio dell’attività lavorativa da parte degli occupati.

Dalle medie presenti nella Tabella 2.12 è evidente come l'andamento delle non risposte nelle diverse interviste sia diverso tra il 2004 e il 2005.

Nel 2004 nella prima rilevazione le non risposte al quesito sull'anno di inizio del proprio lavoro sono il 4.24%, mentre dalla seconda rilevazione in poi la frequenza è attorno al 10%.

Nel 2005 la percentuale degli occupati per cui non ci sono risposte nella prima rilevazione è simile (4.28%), ma dalla seconda rilevazione si ha un calo di mancate risposte, non un aumento come si è verificato nell'anno precedente.

VAR. INDIPENDENTE = P[NON RISPONDERE]		N° CAMPIONE =426520		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	II RILEVAZIONE	.983	.026	.000
	III RILEVAZIONE	.950	.026	.000
	IV RILEVAZIONE	1.005	.026	.000
ANNO	ANNO 2005	.016	.031	.618
RILEVAZIONI 2005 (INTERAZIONI)	(II RILEV.)*2005	-1.656	.044	.000
	(III RILEV.)*2005	-1.298	.043	.000
	(IV RILEV.)*2005	-1.577	.045	.000
AREA RESIDENZA	NORD OVEST	.025	.022	.264
	NORD EST	-.016	.023	.494
	SUD	.362	.022	.000
	ISOLE	.339	.027	.000
ETA'	15-25	-.233	.030	.000
	36-45	-.015	.019	.435
	46-55	.115	.020	.000
	56-75	.520	.023	.000
LIVELLO ISTRUZIONE	ELEMENTARE	.731	.022	.000
	MEDIA	.410	.017	.000
	DIPLO. 3 ANNI	.168	.028	.000
	LAUREA	-.502	.028	.000
SESSO	DONNA	-.587	.015	.000
	COSTANTE	-3.368	.032	.000

Tabella 2.13: modello di tipo logit per la probabilità di non rispondere al quesito sull'anno in cui hanno iniziato la l'occupazione da parte degli occupati.

La differenza tra i due anni di indagine risulta netta anche dalle stime prodotte dal modello logit (Tabella 2.13). La probabilità di non rispondere stimata dal modello per la baseline è attorno al 3.3% nella prima rilevazione dei due anni (i due coefficienti non sono significativamente diversi), mentre sale al 8.56% nella quarta intervista dell'indagine del 2004 e invece scende al 1.9% nella corrispondente intervista del 2005.

Differenze nei due anni di indagine che possono far pensare a un cambiamento nel questionario o nella modalità di intervista, almeno per quanto riguarda questo quesito. In particolare la diversità è evidente dalla seconda rilevazione in poi, ovvero da quando l'individuo è intervistato telefonicamente con tecnica CATI. Dal momento che l'anno di

inizio lavoro dovrebbe essere una delle domande “a conferma” nella rilevazione CATI, questi dati fanno pensare che questa modalità sia entrata in funzione solo nel 2005.

Anche in questo caso sulla frequenza delle non risposte influiscono l’età, l’istruzione e il sesso dell’individuo. L’età e l’istruzione hanno un effetto simile a quello visto nelle risposte sulla fine dell’attività lavorativa: il tasso di risposta aumenta col grado di istruzione e diminuisce con l’avanzare dell’età. Diversa, invece, la probabilità di rispondere a seconda del sesso: in questo caso le donne rispondono più frequentemente degli uomini (per il quesito visto in precedenza era l’opposto).

Il rispondere o meno al quesito può essere influenzato anche dall’aver iniziato recentemente la propria attività. Nella Tabella 2.14 si va quindi a confrontare il tasso di non risposta in base allo stato precedente.

RILEVAZIONE TRIMESTRE	II		III		IV	
	SI	NO	SI	NO	SI	NO
2004						
ERA OCCUPATO?	SI	NO	SI	NO	SI	NO
II	1,246 10.39	74 9.79			1,437 10.56	82 9.77
III	1,378 10.96	88 11.49			1,649 11.07	85 8.17
IV	1,233 9.93	69 9.22	1,192 9.80	99 10.22	1,641 10.20	99 9.46
MEDIA	10.43	10.18	9.80	10.22	10.60	9.09
2005						
I	241 1.85	45 7.94	359 3.06	37 4.95	447 3.61	39 7.54
II	334 2.00	71 8.16	469 3.80	41 4.70	211 1.75	41 6.05
III	255 1.75	51 6.99	388 3.18	55 6.36	245 1.98	42 6.42
IV	303 2.24	59 7.78	196 1.53	65 6.37	209 1.64	45 6.11
MEDIA	1.96	7.73	2.88	5.65	2.24	6.46

Tabella 2.14: frequenze assolute e percentuali, divise in base allo stato precedente, di chi non sa rispondere alla domanda sull'anno in cui ha iniziato l'attività lavorativa che sta svolgendo in quel momento.

Anche dalla Tabella 2.14 risulta evidente una discontinuità tra i due anni di indagine.

In particolare si può notare come il 2004 e il 2005 differiscano soprattutto nella frequenza di non risposte da parte di chi era già occupato nella rilevazione precedente.

Nel primo anno di analisi il tasso di risposta, dalla seconda intervista in poi, è simile tra i nuovi occupati e i vecchi occupati, nel 2005 invece c'è una netta differenza tra chi era già occupato, che risponde in oltre il 97% dei casi, e chi è transitato verso l'occupazione tra le due rilevazioni, che non risponde, in alcuni trimestri, in più del 8% dei casi: questa sembra essere un'ulteriore conferma sull'introduzione delle risposte "a conferma" per questo quesito solo nel 2005.

Si introduce quindi nel modello, le cui stime sono presenti nella Tabella 2.15, una variabile dummy (GIAOCC) che assume il valore uno se l'intervistato era occupato nella rilevazione precedente, zero altrimenti.

VAR. INDIPENDENTE = P[NON RISPONDERE]		N° CAMPIONE = 265336		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	-.046	.034	.182
	IV RILEVAZIONE	.020	.023	.375
ANNO	ANNO 2005	-.515	.068	.000
RILEVAZIONI 2005 (INTERAZIONI)	(III RILEV.)*2005	.344	.051	.000
	(IV RILEV.)*2005	.078	.046	.089
STATO PRECEDENTE	GIAOC	.126	.046	.006
	GIAOC*2005	-1.223	.065	.000
AREA RESIDENZA	NORD OVEST	-.021	.029	.465
	NORD EST	-.013	.030	.675
	SUD	.277	.028	.000
	ISOLE	.238	.035	.000
ETA'	15-25	-.299	.040	.000
	36-45	-.006	.025	.801
	46-55	.137	.026	.000
	56-75	.590	.030	.000
LIVELLO ISTRUZIONE	ELEMENTARE	.766	.029	.000
	MEDIA	.428	.022	.000
	DIPLO. 3 ANNI	.174	.037	.000
	LAUREA	-.508	.037	.000
SESSO	DONNA	-.602	.020	.000
	COSTANTE	-2.480	.057	.000

Tabella 2.15: modello logit per la probabilità di non rispondere sull'anno in cui hanno iniziato l'occupazione da parte degli occupati. Sono presenti le ultime tre interviste dell'indagine.

Le stime ottenute dal modello (Tabella 2.15) confermano l'andamento profondamente diverso del tasso di risposta al quesito tra i due anni di indagine. Nel 2004 l'essere già occupati nella precedente rilevazione aumenta, anche se non di molto, la probabilità non

rispondere. Al contrario nel 2005 l'essere già occupati diminuisce, e di molto, la frequenza di non risposte. Ad esempio, per un individuo baseline la probabilità stimata di non rispondere al quesito nella seconda intervista se risulta occupato nella rilevazione precedente è 8.6% nel 2004 e 1.6% nel 2005.

L'impatto di area di residenza, istruzione ed età dell'intervistato è simile a quello stimato nel modello precedente: la frequenza di non risposte diminuisce al crescere del livello di istruzione e aumenta con l'età. Resta significativo anche il sesso: le donne rispondono più frequentemente degli uomini.

La differenza tra i due anni nell'andamento della frequenza delle non risposte sembra essere dovuto, come detto, a un cambiamento nell'indagine di tipo CATI. Si può pensare che solo dal 2005 si sia passati a una richiesta di conferma rispetto alla risposta precedente per chi risultava già occupato e quindi sia stato effettuato un maggior controllo longitudinale sui dati raccolti da parte dei rilevatori.

2.5 Incoerenze nei dati retrospettivi

La Rilevazione Continua delle Forze Lavoro è un'indagine strutturata in modo tale da poter determinare la condizione lavorativa nella settimana di riferimento per ogni individuo presente nel campione. Per chi risponde anche ai quesiti sulle esperienze passate può essere interessante controllare la coerenza longitudinale tra i dati raccolti e lo stato osservato nella rilevazione precedente.

La risposta a una domanda di tipo retrospettivo può essere definita incoerente se porta a una ricostruzione della vita lavorativa che non corrisponde con le condizioni rilevate in precedenza. Le incoerenze si determinano confrontando la condizione lavorativa nella rilevazione precedente con le domande sull'inizio della presente attività lavorativa per gli occupati e sul termine dell'ultima occupazione per disoccupati e inattivi.

Ad esempio, un individuo che è occupato nella terza rilevazione darà una risposta incoerente se dichiara di aver iniziato la sua attività lavorativa in un periodo precedente a quello in cui è stata svolta la seconda rilevazione, ma in questa era risultato essere disoccupato.

Nei dati a nostra disposizione non è presente il mese in cui si è svolta l'intervista, ma solo il trimestre di riferimento. Questa mancanza non permette di identificare tutte le incoerenze presenti nei dati. Proseguendo con l'esempio precedente si avrà che nella terza rilevazione l'occupato ha risposto in modo inadeguato solo se ha dichiarato di aver iniziato a lavorare

prima del trimestre in cui si è svolta la rilevazione, non prima del mese effettivo in cui è stato intervistato.

Due individui incoerenti possono pertanto essere considerati tali o meno nell'analisi a seconda del mese in cui sono rilevati nel trimestre: una persona intervistata a gennaio e ad aprile 2004 sarà considerata incoerente nell'analisi se è, ad esempio, inoccupata nella prima rilevazione e nella seconda dichiara di aver iniziato la sua occupazione nel dicembre 2003. Un'altra persona dello stesso campione ma intervistata a marzo e a giugno 2004 non risulterà incoerente nella nostra analisi se, inoccupata nella prima rilevazione, dichiara di aver iniziato a lavorare nel gennaio 2004 nella seconda.

L'assenza del mese in cui si è svolta la rilevazione porta quindi a stime distorte verso il basso della probabilità che un individuo risponda incoerentemente a un quesito di tipo retrospettivo. Il numero di incoerenze sarà quindi sottostimato per chi è stato intervistato nel secondo e nel terzo mese del trimestre.

Nel seguito si considerano esclusivamente le incoerenze di interesse per la ricostruzione della durata della disoccupazione. Non si analizza perciò, ad esempio, la coerenza fra date di inizio occupazione riportate in due occasioni diverse.

2.5.1 Incoerenze sull'inizio dell'attività lavorativa

Il campione analizzato è quello dei nuovi occupati, ovvero gli occupati che erano inoccupati nella rilevazione precedente, che rispondono al quesito sull'anno e sul mese di inizio della loro attività lavorativa. Un individuo del campione risulta incoerente se dichiara di aver iniziato il lavoro che sta svolgendo prima del trimestre in cui è stata effettuata l'intervista.

Il numero di occupati che sostengono di aver iniziato la loro attività lavorativa in un momento antecedente al trimestre in cui sono stati intervistati (e classificati come inattivi o disoccupati) è molto elevato, spesso superiore al 50% complessivamente.

Dalla Tabella 2.10 risultano più incoerenti coloro che transitano verso l'occupazione da uno stato di inattività (sono più del 60% in alcuni casi) rispetto a chi nella rilevazione precedente risultava essere disoccupato (la frequenza di incoerenze non supera mai il 35%). Inoltre è evidente che le incoerenze sono più frequenti tra le osservazioni a tre mesi di distanza (prima e seconda, terza e quarta) a prescindere dalla condizione lavorativa di provenienza.

RILEVAZIONE TRIMESTRE	II		III		IV	
	DISOC	NFL	DISOC	NFL	DISOC	NFL
2004						
II	75 29.18	234 63.41			77 29.84	260 58.04
III	73 34.11	255 60.57			93 32.18	352 59.46
IV	69 28.99	256 64.48	69 25.00	237 44.63	93 32.07	375 62.19
MEDIA	30.61	62.76	25.00	44.63	31.42	60.07
2005						
I	48 25.67	179 59.87	52 23.01	169 38.50	45 26.47	157 59.47
II	82 26.89	245 60.49	50 20.83	196 38.06	56 27.45	203 54.28
III	79 32.38	199 50.90	47 19.42	230 45.10	57 32.02	213 53.65
IV	67 29.39	238 56.80	74 25.78	240 40.34	44 25.43	265 55.44
MEDIA	28.63	56.87	22.41	40.55	27.86	55.39

Tabella 2.16: incoerenze sull'inizio dell'attività lavorativa da parte dei nuovi occupati, in base allo stato precedente.

Dal momento che le frequenze, a seconda dello stato precedente, sono molto diverse può essere utile analizzare separatamente il campione per la stima del modello logit.

Le stime del modello logit presenti nella Tabella 2.17 evidenziano come la probabilità di rispondere in modo coerente da parte di chi transita dall'inattività all'occupazione sia influenzata principalmente dall'età: più è alta l'età del nuovo occupato più aumenta la frequenza di incoerenze. Ad esempio, nella seconda rilevazione del 2004 un diplomato dell'Italia centro-settentrionale risponde incoerentemente nel 40% dei casi se ha meno di 25 anni, ma nel 90% delle volte se ha più di 55 anni. Quest'ultimo dato è alquanto sorprendente dal momento che una frequenza così alta di risposte non congruenti non sembra essere spiegabile da difficoltà nel rispondere a un'intervista di tipo telefonico o altro.

Non molto chiaro è l'effetto del grado di istruzione. Il modello infatti stima una maggiore correttezza delle risposte non solo per chi ha un titolo di studio superiore al diploma ma anche per chi non è andato oltre la scuola dell'obbligo: la probabilità stimata di dare

informazioni retrospettive errate nella seconda rilevazione del 2004 da parte di uomo tra i 25 e i 35 anni che vive nel Centro-Nord è del 64% se ha un diploma di 5 anni, del 59.4% se non è andato oltre la quinta elementare, del 50.8% se ha un titolo universitario.

Un'evidenza confermata dal modello è il minor numero di incoerenze nelle risposte nella terza rilevazione rispetto alle altre. Una distanza superiore tra le rilevazioni (nove mesi tra la seconda e la terza, tre mesi tra le altre) sembra portare a una maggiore correttezza nelle risposte anche al netto delle caratteristiche individuali.

Gli intervistati sembrano avere la tendenza a dichiarare di aver iniziato a lavorare in un periodo precedente a quello effettivo e questa tendenza non si riduce se il lavoro è iniziato da pochi mesi.

VAR. INDIPENDENTE = P[RISPOSTA INCOERENTE]		N° CAMPIONE =8447		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	-.767	.115	.000
	IV RILEVAZIONE	-.062	.085	.464
ANNO	ANNO 2005	-.220	.085	.010
RILEVAZIONI 2005 (INTERAZIONI)	(III RILEV.)*2005	.080	.137	.556
	(IV RILEV.)*2005	-.073	.116	.527
AREA RESIDENZA	NORD OVEST	-.013	.083	.874
	NORD EST	.011	.084	.893
	SUD	.011	.076	.887
	ISOLE	-.244	.092	.008
ETA'	15-25	-.957	.066	.000
	36-45	.324	.071	.000
	46-55	.716	.083	.000
	56-75	1.734	.104	.000
LIVELLO ISTRUZIONE	ELEMENTARE	-.201	.092	.030
	MEDIA	-.173	.059	.003
	DIPLO. 3 ANNI	-.156	.109	.152
	LAUREA	-.546	.080	.000
SESSO	DONNA	.078	.050	.118
	COSTANTE	.583	.108	.000

Tabella 2.17: modello logit per la probabilità di essere incoerenti da parte di chi transita dall'inattività all'occupazione.

Dalle stime presenti nella Tabella 2.18 si nota che la frequenza di risposte incoerenti sull'inizio dell'attività lavorativa è nettamente inferiore nel caso in cui l'occupato provenga dalla disoccupazione. Non cambia però l'andamento del numero di incoerenze registrato nell'indagine: il minimo si registra, anche in questo caso, nella terza rilevazione. Anche in questo caso la tendenza a dichiarare di aver iniziato a lavorare in un periodo precedente a quello effettivo non diminuisce nel caso in cui si il lavoro è iniziato da meno tempo.

Rispetto alla transizione dall'inattività le caratteristiche individuali hanno un effetto diverso sulla probabilità stimata di risultare incoerenti. Oltre all'età risulta significativa anche la

regione di residenza, mentre non ci sono molte differenze a seconda del grado di istruzione raggiunto, se non per chi ha un titolo universitario che risponde più frequentemente in modo coerente. L'effetto dell'età dell'individuo è simile a quanto visto in precedenza: le incoerenze aumentano all'aumentare dell'età anche se, in questo caso, non ci sono evidenti differenze tra chi transita verso l'occupazione oltre i 45 anni.

Nella seconda (o quarta) rilevazione un laureato residente in una regione settentrionale con meno di 26 anni ha una frequenza stimata di rispondere in modo non corretto al quesito del 12.3%. Per un occupato tra i 45 e i 55 anni non laureato e residente nel Sud la probabilità di indicare una data di inizio dell'attività lavorativa incompatibile con la precedente rilevazione è stimata dal modello essere il 52%.

VAR. INDIPENDENTE = P[RISPOSTA INCOERENTE]		N° CAMPIONE = 4506		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	-.368	.165	.026
	IV RILEVAZIONE	.004	.114	.970
ANNO	ANNO 2005	-.088	.111	.431
RILEVAZIONI 2005 (INTERAZIONI)	(III RILEV.)*2005	-.033	.197	.865
	(IV RILEV.)*2005	-.055	.160	.732
AREA RESIDENZA	NORD OVEST	-.293	.124	.018
	NORD EST	-.294	.129	.023
	SUD	.348	.107	.001
	ISOLE	.056	.126	.656
ETA'	15-25	-.591	.095	.000
	36-45	.282	.092	.002
	46-55	.604	.115	.000
	56-75	.502	.201	.013
LIVELLO ISTRUZIONE	ELEMENTARE	.041	.130	.754
	MEDIA	.019	.084	.817
	DIPLO. 3 ANNI	-.051	.157	.747
	LAUREA	-.217	.118	.065
SESSO	DONNA	.107	.070	.129
	COSTANTE	-.870	.143	.000

Tabella 2.18: modello logit per la probabilità di essere incoerenti da parte di chi transita dalla disoccupazione all'occupazione.

2.5.2 Incoerenze sulla fine dell'attività lavorativa

Il campione è composto dagli individui inoccupati che rispondono al quesito sull'anno (ed eventualmente mese) in cui hanno smesso di lavorare e che risultano occupati nella rilevazione precedente a quella in esame. Un individuo risponde in modo incoerente se dichiara di aver terminato la sua ultima attività lavorativa in un periodo antecedente al trimestre in cui si è svolta la precedente rilevazione.

RILEVAZIONE TRIMESTRE	II		III		IV	
STATO	DISOC	NFL	DISOC	NFL	DISOC	NFL
2004						
II	50 30.67	193 44.16			48 28.40	182 39.82
III	34 22.37	183 32.28			48 23.08	203 29.29
IV	42 22.70	146 25.61	32 13.68	157 26.12	45 15.31	186 24.28
MEDIA	25.20	33.16	13.68	26.12	21.01	29.80
2005						
I	50 25.00	195 31.35	34 12.83	143 20.25	34 18.99	133 25.73
II	74 33.64	253 40.81	21 10.10	152 22.03	36 28.80	110 35.14
III	37 18.41	192 30.67	27 12.44	129 18.22	22 13.66	122 25.85
IV	32 16.75	148 24.50	31 15.35	114 20.50	29 16.11	106 22.18
MEDIA	23.77	31.88	12.67	20.23	18.76	26.46

Tabella 2.19: incoerenze dei nuovi inoccupati sulla fine dell'attività lavorativa precedente in base allo stato attuale.

Dalla Tabella 2.19 è evidente che, seppure con valori inferiori rispetto al caso precedente, anche per la fine dell'occupazione le incoerenze hanno un peso importante. La frequenza relativa delle incongruenze è superiore per chi transita verso l'inattività. Inoltre, come si era notato anche nelle risposte sull'inizio dell'attività occupazionale, si ha il minimo delle incoerenze nella terza rilevazione, ovvero quando la distanza dall'osservazione precedente è maggiore.

Dal momento che il numero delle incongruenze è diverso a seconda della condizione lavorativa dell'intervistato sembra opportuno procedere a un'analisi separata per il modello logit.

VAR. INDIPENDENTE = P[RISPOSTA INCOERENTE]		N° CAMPIONE =11003		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	-.491	.112	.000
	IV RILEVAZIONE	-.113	.077	.146
ANNO	ANNO 2005	-.088	.073	.226
RILEVAZIONI 2005 (INTERAZIONI)	(III RILEV.)*2005	-.240	.131	.066
	(IV RILEV.)*2005	-.130	.106	.221
AREA RESIDENZA	NORD OVEST	.250	.074	.001
	NORD EST	.256	.074	.001
	SUD	-.167	.072	.020
	ISOLE	-.369	.094	.000
ETÀ	15-25	-1.284	.098	.000
	36-45	.286	.075	.000
	46-55	.453	.077	.000
	56-75	.980	.072	.000
LIVELLO ISTRUZIONE	ELEMENTARE	-.134	.070	.056
	MEDIA	.004	.060	.953
	DIPLO. 3 ANNI	-.034	.105	.748
	LAUREA	-.303	.088	.001
SESSO	DONNA	-.182	.048	.000
	COSTANTE	-.809	.101	.000

Tabella 2.20: modello logit per la probabilità di rispondere in modo incoerente sulla fine dell'attività lavorativa precedente da parte dei nuovi inattivi.

Per chi transita dall'occupazione all'inattività (Tabella 2.20) non ci sono differenze significative tra i due anni di indagine nella frequenza di incoerenze sulla fine dell'ultima occupazione svolta, mentre è evidente il calo nella terza intervista.

Secondo il modello stimato hanno un effetto significativo sulla frequenza delle incongruenze l'area di residenza, il sesso e l'età e dell'intervistato, mentre per quanto riguarda l'istruzione solo l'aver conseguito titolo di studio di tipo universitario sembra influire. In particolare si può notare come la probabilità stimata di collocare la fine dell'ultima occupazione in un periodo precedente a quello indicato dall'indagine aumenta, anche in questo caso, al crescere dell'età: nella seconda intervista un uomo residente nell'Italia centrale che non ha una laurea ha una probabilità stimata di rispondere incoerentemente del 11% se ha meno di 26 anni, del 54% se ha più di 55 anni.

VAR. INDIPENDENTE = P[RISPOSTA INCOERENTE]		N° CAMPIONE =3754		
VARIABILI ESPLICATIVE		COEFFICIENTE	ERRORE STD.	P-VALUE
RILEVAZIONE	III RILEVAZIONE	-.776	.219	.000
	IV RILEVAZIONE	-.208	.143	.145
ANNO	ANNO 2005	-.044	.135	.742
RILEVAZIONI 2005 (ITERAZIONI)	(III RILEV.)*2005	-.056	.256	.827
	(IV RILEV.)*2005	-.141	.195	.468
AREA RESIDENZA	NORD OVEST	.124	.149	.403
	NORD EST	-.233	.165	.158
	SUD	.193	.134	.151
	ISOLE	-.007	.159	.963
ETA'	15-25	-.424	.129	.001
	36-45	.378	.113	.001
	46-55	.641	.131	.000
	56-75	.735	.217	.001
LIVELLO ISTRUZIONE	ELEMENTARE	.021	.152	.890
	MEDIA	.149	.105	.155
	DIPLO. 3 ANNI	.075	.185	.684
	LAUREA	-.443	.175	.011
SESSO	DONNA	-.107	.089	.229
	COSTANTE	-1.283	.181	.000

Tabella 2.21: modello logit per la probabilità di rispondere in modo incoerente sulla fine dell'attività lavorativa precedente da parte dei nuovi disoccupati.

Dalle stime prodotte dalla regressione logistica, presentate nella Tabella 2.21, si può notare che l'andamento delle incoerenze all'interno dell'indagine per chi transita dall'occupazione alla disoccupazione è molto simile a quello visto per chi transita verso l'inattività. Infatti anche in questo caso non ci sono differenze tra i due anni di indagine mentre è evidente come la frequenza di incongruenze sia minore nella terza rilevazione rispetto alle altre.

Differente è invece l'effetto delle caratteristiche individuali: per chi transita verso la disoccupazione non è significativa né l'area di residenza né l'essere uomo o donna. Resta, al contrario, significativa l'età dell'intervistato: si stima che alla seconda intervista un giovane diplomato disoccupato con meno di 26 anni di età, occupato nella prima rilevazione, risponda in modo incoerente sulla fine della sua ultima attività nel 15.4 % dei casi, mentre un individuo con le stesse caratteristiche ma con più di 55 anni ha una probabilità stimata di essere incoerente del 36.7%. Come per chi transita verso la non forza lavoro ci sono meno incoerenze da parte di chi ha un titolo di studio universitario.

Il fatto che gli individui si contraddicano maggiormente sul periodo in cui hanno iniziato o smesso di lavorare nelle rilevazioni in cui l'intervista avviene a tre mesi di distanza rispetto a quando avviene a 9 mesi di distanza (la terza rilevazione), è abbastanza sorprendente.

Gli intervistati non appaiono influenzati dalla distanza temporale tra intervista ed evento (inizio o fine del lavoro) nell'indicare quando l'evento stesso avviene, ma sembrano

collocare comunque il periodo in cui è cambiata la loro condizione lavorativa più lontano nel tempo di quello che è. Ammesso che esista questa tendenza a collocare questo tipo di eventi in anticipo da parte di alcuni individui, restano non chiare le motivazioni e perché la percentuale di incoerenze sia più elevata tra chi transita, in entrambe le direzioni, tra inattività e occupazione. Inoltre si deve sempre tener conto che il numero di incoerenze trovate è sottostimato a causa dell'assenza del mese in cui è stata svolta l'intervista e che quindi solo le risposte incoerenti di chi è intervistato nel primo mese del trimestre sono stimate correttamente.

Dall'analisi appare comunque chiaro che in questi quesiti di tipo retrospettivo non sono applicati controlli automatici di tipo longitudinale contestualmente alla rilevazione.

2.5.3 Durata della ricerca di lavoro

Un'altra incoerenza di interesse tra le rilevazioni si può avere sulle risposte al quesito che riguarda la durata di ricerca di occupazione. Questo tipo di incoerenza può anche essere considerata trascurabile in quanto non in contraddizione con gli stati rilevati, ma è comunque interessante dal momento che il dato sulla durata della ricerca di lavoro sarà utilizzato per ricostruire la lunghezza del periodo di disoccupazione.

Si vanno quindi a confrontare le risposte alla domanda "*Da quanti mesi sta cercando lavoro?*" da parte di chi:

- risulta disoccupato in due rilevazioni consecutive;
- dà una risposta in entrambe le occasioni;
- cerca lavoro da un tempo superiore alla distanza "teorica" che ci dovrebbe essere tra le rilevazioni (tre mesi tra le prime e le ultime due, nove mesi tra la seconda e la terza) in modo da non considerare coloro che escono temporaneamente dalla disoccupazione.

Il confronto tra le risposte presente nella Tabella 2.22 mostra come la quasi totalità dei disoccupati non sia coerente nelle due dichiarazioni se si accetta l'ipotesi che le rilevazioni siano effettuate esattamente a tre mesi di distanza sia tra la prima e la seconda rilevazione sia tra la terza e la quarta e a nove mesi di distanza tra la seconda e la terza rilevazione.

Le cause che possono portare a dichiarare durate della ricerca di lavoro incoerenti in due rilevazioni consecutive sono varie. Innanzitutto nei dati retrospettivi di durata è comune una forte polarizzazione delle risposte intorno al numero di mesi corrispondenti ai semestri e

all'anno (Trivellato e Torelli, 1993) e questo porta ovviamente a un'incoerenza tra le risposte.

Inoltre il processo di memoria di chi risponde può non essere accurato per chi è disoccupato da molti mesi (caso comune nel campione visto che si considera chi è disoccupato in almeno due rilevazioni consecutive) e quindi può portare a delle discrepanze tra le risposte date a distanza di tempo.

Il numero elevatissimo di incoerenze (la frequenza relativa di chi risponde coerentemente nelle due rilevazioni è quasi sempre sotto il 10%) potrebbe però anche far ipotizzare che le rilevazioni non avvengano esattamente alla distanza "teorica", ma che per vari motivi possano trascorrere un numero diverso di settimane tra le rilevazioni e che questo comporti un'ulteriore aumento delle incoerenze.

RILEVAZIONE TRIMESTRE	II		III		IV	
	$t_2 - t_1 > 3$	$t_2 - t_1 < 3$	$t_3 - t_2 > 9$	$t_3 - t_2 < 9$	$t_4 - t_3 > 3$	$t_4 - t_3 < 3$
2004						
II	165 44.59	170 45.95			188 42.63	217 49.21
III	141 38.52	198 54.10			174 43.94	182 45.96
IV	136 39.65	166 48.40	115 42.75	136 50.56	178 42.48	195 46.54
2005						
I	161 35.70	244 54.10	76 43.43	87 49.71	120 55.05	97 44.50
II	202 38.40	264 50.19	105 44.12	123 51.68	149 42.82	157 45.11
III	138 36.22	205 53.81	113 50.90	100 45.05	121 45.83	124 46.97
IV	143 38.03	190 50.53	159 52.13	125 40.98	113 35.09	178 55.28

Tabella 2.22: differenze tra le durate di ricerca lavoro dichiarate in due occasioni successive. t_i indica la durata dichiarata nella i -esima intervista.

Potrebbe quindi essere interessante considerare come incoerenze solo i casi in cui le differenze tra le durate dichiarate superino il limite minimo e massimo della distanza tra le rilevazioni. Quindi gli intervistati in due trimestri consecutivi saranno considerati incoerenti se nella seconda rilevazione dichiareranno di cercare lavoro da meno mesi rispetto alla prima

(è stato comunque lasciato il vincolo che devono cercare da più di tre mesi) o da oltre sei mesi in più, sempre rispetto a quanto dichiarato nella rilevazione precedente.

Gli intervistati nella terza rilevazione saranno considerati incoerenti se la durata di ricerca non supera di almeno sei mesi quella registrata in precedenza o se dichiarano di cercare lavoro da un anno in più di quello che avevano dichiarato nella seconda rilevazione.

RILEVAZIONE TRIMESTRE	II		III		IV	
	$t_2 - t_1 > 6$	$t_2 - t_1 < 0$	$t_3 - t_2 > 12$	$t_3 - t_2 < 6$	$t_4 - t_3 > 6$	$t_4 - t_3 < 0$
2004						
II	105 28.38	98 26.49			120 27.21	93 21.09
III	114 31.15	113 30.87			128 32.32	87 21.97
IV	92 26.82	88 25.66	70 26.02	102 37.92	124 29.59	79 18.85
2005						
I	102 22.62	138 30.60	55 31.43	69 39.43	105 48.17	65 29.82
II	119 22.62	150 28.52	68 28.57	96 40.34	92 26.44	71 20.40
III	96 25.20	112 29.40	60 27.03	68 30.63	91 34.47	60 22.73
IV	98 26.06	98 26.06	86 28.20	97 31.80	86 26.71	87 27.02

Tabella 2.23 differenze tra le durate di ricerca lavoro dichiarate in due occasioni successive. t_i indica la durata dichiarata nella i -esima intervista.

Anche escludendo l'assunzione che le rilevazioni siano effettuate sempre alla distanza prevista, la frequenza delle incoerenze resta comunque abbastanza elevata, complessivamente sempre superiore al 45%, come si può osservare nella Tabella 2.23.

Le incoerenze sulle risposte non hanno una "direzione" precisa, anche se sembra che nella terza rilevazione si tenda a dichiarare più spesso una durata di ricerca inferiore a quella "corretta" rispetto alla seconda e quarta intervista. Risultano quindi rilevanti i fenomeni di polarizzazione delle risposte con l'arrotondamento della durata rispetto a certi valori e la diversa percezione della durata della ricerca tra una rilevazione e l'altra.

Visto l'elevato numero di incoerenze si potrebbe ipotizzare che la dichiarazione più affidabile sulla durata della ricerca di lavoro sia quella data nella prima intervista in cui l'individuo

risulta disoccupato, dal momento che la rilevazione è quella più vicina, in termini temporali, all'inizio della ricerca e quindi è facilitato il processo di memoria.

Questo non è così scontato (come si è visto in precedenza con le incoerenze su inizio e fine lavoro) e l'affidabilità o meno della dichiarazione potrebbe dipendere anche dal trimestre in cui si svolge l'intervista. Tale presunzione sarà comunque mantenuta nel seguito per la ricostruzione delle durate della disoccupazione a partire dai dati longitudinali.

CAPITOLO 3

Ricostruzione e analisi della durata della disoccupazione

La seconda parte di questo studio è dedicata a un'analisi dei dati di durata ottenibili a partire dalla Rilevazione Continua delle Forze di Lavoro nel 2004 e 2005. L'analisi sarà rivolta alla durata degli episodi di disoccupazione, in particolare a quella del primo episodio osservato nei vari abbinamenti a tre episodi. Per poter svolgere l'analisi innanzitutto sarà necessario ricostruire la durata degli episodi (o spell) di disoccupazione e poi applicare un modello statistico che tenga conto della particolarità dei dati dovuta alla struttura dell'indagine.

3.1 Modelli di durata della disoccupazione

Nell'analisi dei dati di durata si considera la lunghezza osservata dello spell, indicata con t , come la determinazione di una variabile casuale T che ha funzione di ripartizione $F(T)$, funzione di densità $f(t)$ e che può assumere solo valori positivi. Ci sono altre funzioni legate alla variabile casuale T di grande interesse nell'analisi dei dati di durata. Fra queste la funzione di sopravvivenza $S(t)$, che rappresenta la probabilità che l'episodio di disoccupazione duri oltre un determinato istante t : $P[T \geq t]$. La funzione di sopravvivenza, quindi, non è altro il complemento a uno della funzione di ripartizione.

Un'altra funzione utile nello studiare l'andamento dei dati di durata è la probabilità che uno spell termini nell'istante t condizionata al fatto che l'episodio sia ancora in corso al tempo t :

$$\theta(t) = \lim_{dt \rightarrow 0} \frac{P[t \leq T < t + dt | T \geq t]}{dt} = \frac{f(t)}{S(t)}$$

$\theta(t)$ è detta funzione di rischio ed è fondamentale per stabilire in quale modo la probabilità di transitare verso un altro stato sia dipendente dalla durata dell'episodio: se la derivata di $\theta(t^*)$ è positiva allora la dipendenza dalla durata in t^* è positiva, ovvero la probabilità che uno spell termini cresce all'aumentare della durata dell'episodio. Se la derivata di $\theta(t^*)$ è negativa allora la dipendenza dalla durata è anch'essa negativa e quindi la probabilità che

l'episodio termini diminuisce all'aumentare della sua durata. Se $\frac{d\theta(t)}{dt} = 0$ in ogni t allora non c'è dipendenza dalla durata.

Nei modelli di tipo parametrico la forma della funzione di rischio dipende dalle ipotesi sulla distribuzione che genera le durate. Ad esempio se si ipotizza che esista una dipendenza dalla durata monotona sia assumerà che la v.c. T si distribuisca come una Weibull con funzione di rischio:

$$\theta(t) = \alpha t^{\alpha-1} \omega \text{ con } \alpha, \omega > 0.$$

L'andamento di $\theta(t)$ è monotono crescente se $\alpha > 1$ e monotono decrescente se $\alpha < 1$. Se $\alpha = 1$ è costante e la distribuzione diventa un'esponenziale. Il parametro α è anche detto parametro "di forma":

La probabilità di transitare da uno stato all'altro può però variare da persona a persona a seconda di determinate caratteristiche individuali. Per determinare come queste caratteristiche influenzano il rischio di uscire da un episodio di disoccupazione si dovranno fare delle assunzioni. Quella più comune per i modelli di durata è di specificare i rischi come proporzionali:

$$\theta(t, \underline{x}, \underline{\beta}, \theta_0) = \theta_0(t) \phi(\underline{x}, \underline{\beta}).$$

$\theta_0(t)$ è la funzione di rischio di base mentre $\phi(\underline{x}, \underline{\beta})$ non fa altro che moltiplicare la funzione di rischio per un fattore scalare che dipende dalle caratteristiche dell'individuo.

Una specificazione comune per conservare il vincolo di positività per la funzione è la seguente:

$$\phi(\underline{x}, \underline{\beta}) = \exp(\underline{x}'\underline{\beta})$$

Tornando al modello Weibull la funzione di rischio, una volta scelto in che modo si concretizza l'effetto delle variabili esplicative, diventa:

$$\theta(t) = \alpha t^{\alpha-1} \exp(\underline{x}'\underline{\beta}),$$

mentre la funzione di sopravvivenza è:

$$S(t) = \exp\left[-t^\alpha \exp(\underline{x}'\underline{\beta})\right].$$

Il modello Weibull non è, ovviamente, l'unico modello utilizzabile nell'analisi dei dati di durata. Un altro modello molto utilizzato in quanto richiede meno assunzioni (ed è quindi meno vincolante) sulla funzione di rischio è basato sulla distribuzione esponenziale a tratti ipotizzando una funzione di rischio costante all'interno di intervalli di tempo determinati in modo arbitrario (si veda Jenkins, 2005):

$$\theta(t, \underline{x}) \begin{cases} \bar{\theta}_1 \exp(\underline{x}'\underline{\beta}) & \text{se } t \in [0, \lambda_1] \\ \vdots \\ \bar{\theta}_k \exp(\underline{x}'\underline{\beta}) & \text{se } t \in [\lambda_{k-1}, \lambda_k] \end{cases} .$$

Riparametrizzando e definendo α_i come intercetta la funzione di rischio può essere riscritta come:

$$\theta(t, \underline{x}) \begin{cases} \exp(\alpha_1 + \underline{x}'\underline{\beta}) & \text{se } t \in [0, \lambda_1] \\ \vdots \\ \exp(\alpha_k + \underline{x}'\underline{\beta}) & \text{se } t \in [\lambda_{k-1}, \lambda_k] \end{cases}$$

3.1.1 Problemi di censura

Una delle caratteristiche principali dei dati di durata nelle scienze sociali, in particolare di quelli provenienti da indagini sezionali, è la presenza di censura. Un episodio si definisce censurato se nella finestra di osservazione non sono osservati o il suo inizio o la sua fine.

Nel caso in cui si sia osservato l'istante in cui inizia l'episodio, ma l'individuo non sia ancora transitato al termine del periodo di indagine si parla di episodi censurati a destra. Per questi individui si conosce pertanto solo la funzione di sopravvivenza $S(t)$, dove t indica la durata osservata dell'episodio nell'ultima rilevazione dell'indagine.

Si è invece di fronte a episodi censurati a sinistra nel caso in cui non si osserva il momento in cui lo spell è iniziato, in quanto ha avuto origine prima del periodo in cui si svolge l'indagine.

In questo caso la presenza della censura è maggiormente rilevante in quanto comporta il così detto "length bias problem": lo schema di osservazione tende a rendere più probabile l'ingresso nel campione di individui che sono disoccupati da più tempo. Per non sovrastimare

la durata della disoccupazione per questi individui si utilizza l'informazione $\frac{f(t)}{S(\tau)}$ con t che

rappresenta, come sempre, la durata complessiva dell'episodio e τ il numero di mesi già trascorsi nello stato di disoccupazione prima che iniziasse il periodo di osservazione. Si considera quindi la probabilità che l'individuo transiti dopo t mesi di disoccupazione condizionata al fatto che fosse già disoccupato da τ mesi (per una trattazione completa dell'argomento si veda, fra gli altri, Lancaster, 1990).

A partire dalla tipologia di informazione disponibile per ogni individuo si può esplicitare la funzione di verosimiglianza per un campione casuale di n disoccupati:

$$L(t, \tau, x) = \prod_{i=1}^n \frac{S(t_i, x_i)}{S(\tau_i, x_i)} \theta(t_i, x_i)^{c_i}$$

dove c_i è una variabile binaria che assume valore uno se l'episodio dell'individuo è osservato fino alla sua conclusione e zero se è censurato a destra. Se l'individuo entra nello stato di disoccupato durante il periodo di osservazione, si ha $\tau = 0$ e il denominatore risulta pari a 1.

3.1.2 Rischi competitivi

Fino ad ora si è analizzata la durata della disoccupazione senza distinguere verso quale condizione lavorativa transitasse l'individuo. Per poter svolgere un'analisi di interesse dal punto di vista economico potrebbe essere utile considerare separatamente il rischio di transitare verso l'occupazione, $\theta_{oc}(t, x)$, e quello di transitare verso l'inattività, $\theta_{nfl}(t, x)$.

Se si vuole stimare queste due funzioni a partire dalla verosimiglianza "globale" del modello risulta utile l'assunzione che i due processi siano indipendenti e che quindi valga:

$$\theta(t, x) = \theta_{oc}(t, x) + \theta_{nfl}(t, x)$$

Si può dimostrare che in questo caso vale la seguente scomposizione della funzione di sopravvivenza:

$$S(t, x) = S_{oc}(t, x) S_{nfl}(t, x)$$

e che quindi la funzione di verosimiglianza può essere riscritta, con una notazione semplificata, come:

$$L(t, x) = \theta_{oc}(t, x)^{c_{oc}} S_{oc}(t, x) \theta_{nfl}(t, x)^{c_{nfl}} S_{nfl}(t, x),$$

dove c_{oc} e c_{nfl} sono variabili binarie che valgono uno se l'individuo transita rispettivamente verso l'occupazione o l'inattività.

La log-verosimiglianza quindi diventa:

$$l(t, x) = \left\{ c_{oc} \ln [\theta_{oc}(t, x)] + \ln [S_{oc}(t, x)] \right\} + \left\{ c_{nfl} \ln [\theta_{nfl}(t, x)] + \ln [S_{nfl}(t, x)] \right\}$$

Dal momento che la funzione di log-verosimiglianza è fattorizzata in due parti, e in ogni blocco ci sono parametri diversi che riguardano solo una delle due destinazioni, le stime si possono ottenere massimizzando le due componenti separatamente.

3.2 Dati dell'indagine e ricostruzione degli episodi

Nello studio si è scelto di analizzare la durata degli episodi di disoccupazione per gli abbinamenti a tre osservazioni provenienti da campioni indipendenti presenti nei due anni di rilevazione. In particolare si è effettuata la ricostruzione della lunghezza del primo spell di disoccupazione rilevato per ogni individuo che è risultato disoccupato in almeno una delle tre rilevazioni di ogni abbinamento.

Le analisi svolte nel precedente capitolo hanno mostrato che in molti casi la qualità dei dati retrospettivi è superiore nella prima intervista dell'indagine, ovvero nella rilevazione di tipo CAPI. La prima intervista ha un maggior tasso di risposta sul quesito che riguarda la fine dell'ultima attività lavorativa; dato utile nel ricostruire la durata dell'episodio di disoccupazione. La scelta quindi è stata di utilizzare, dove possibile, gli abbinamenti che coinvolgono le prime tre rilevazioni: nei 5 campioni in cui si andrà a determinare la lunghezza del primo spell di disoccupazione solo in un caso si sarà costretti a usare le ultime tre rilevazioni (nell'abbinamento degli individui entrati nell'indagine nell'ultimo trimestre del 2003, che non è disponibile).

Le ricostruzioni saranno dunque basate sui dati di chi è intervistato:

- nei primi due trimestri del 2004 e nel primo del 2005;
- nel secondo e terzo trimestre del 2004 e nel secondo del 2005;
- negli ultimi due trimestri del 2004 e nel terzo del 2005;
- nell'ultimo trimestre del 2004, nel primo e nel quarto del 2005;
- nel primo e ultimo trimestre del 2004 e nel primo del 2005;

Nella ricostruzione della lunghezza della disoccupazione si è guardato innanzitutto alla condizione lavorativa rilevata nelle osservazioni dell'indagine e si sono poi considerate tutte le informazioni retrospettive utili presenti nelle tre interviste.

L'integrazione di queste informazioni con gli stati rilevati è resa però complessa dall'indisponibilità della data in cui si è svolta l'intervista all'interno del trimestre e dal fatto che non tutti gli individui, come si è visto nel precedente capitolo, forniscono tutte le informazioni di interesse.

L'assenza del mese in cui si svolge l'intervista ha come conseguenza l'impossibilità di determinare una durata univoca del periodo di disoccupazione, in quanto le date retrospettive sulla fine e sull'inizio dell'attività sono indicate dagli intervistati con mese e anno. Se sono

disponibili tutte le informazioni sarà così sempre possibile individuare un intervallo di tempo in cui l'individuo è rimasto disoccupato, ma solo in alcuni casi sarà determinata un'unica durata.

La mancanza di risposte ad alcuni quesiti rende a volte impossibile determinare un intervallo per la lunghezza dello spell e in altre rende la ricostruzione meno accurata. Per questo si è deciso di considerare, al fine di analizzare le possibilità di ricostruire le durate della disoccupazione, gli individui che sono risultati disoccupati in almeno una delle tre occasioni e di cui si hanno tutte le informazioni necessarie per ricostruire nel modo più preciso possibile la durata dell'episodio.

Nella prima osservazione in cui l'intervistato risulta disoccupato queste informazioni sono:

- la durata della ricerca di lavoro;
- l'anno (e il mese se ha smesso da meno di tre anni) in cui l'individuo ha terminato la sua attività se ha dichiarato di aver già lavorato in passato.

Per chi transita dalla disoccupazione all'occupazione un'altra risposta fondamentale per determinare al meglio la lunghezza dello spell è l'inizio della nuova attività. In questo caso, inoltre, non si considererà chi darà risposte incoerenti, ovvero chi dichiara di aver iniziato a lavorare prima del trimestre in cui risulta disoccupato.

Esiste un'ulteriore informazione retrospettiva per i disoccupati che escono verso l'inattività data dalla risposta alla domanda "*Quanti mesi sono passati dall'ultima volta che ha fatto qualcosa per cercare lavoro?*". A questa domanda risponde però esclusivamente chi dichiara di non aver cercato lavoro nelle ultime quattro settimane: solo una parte di coloro che transitano all'inattività.

3.2.1 Selezione del campione

Dai campioni in cui sono presenti tutti gli individui che risultano disoccupati in almeno in un'occasione vengono quindi esclusi coloro che non rispondono ai quesiti che permettono di ricostruire al meglio la durata dell'episodio disoccupazione secondo la sequenza illustrata nella Tabella 3.1). Dal totale di coloro che sono disoccupati almeno una volta nelle tre rilevazioni (1) non vengono considerati :

- i disoccupati che non dichiarano da quanti mesi cercano lavoro nella prima osservazione in cui sono disoccupati (2);
- i disoccupati che non dichiarano quando hanno smesso di lavorare (se hanno risposto di aver già lavorato in passato) nella prima osservazione in cui sono disoccupati (3);

- i disoccupati che, se escono verso l'occupazione, non dichiarano quando hanno iniziato a lavorare (4);
- i disoccupati che, nella rilevazione in cui transitano verso l'occupazione, danno risposte incoerenti, da un punto di vista longitudinale, sull'inizio dell'occupazione (5).

ABBINAMENTO	41/42/51	42/43/52	43/44/53	44/51/54	41/44/51
DISOCCUPATI ⁽¹⁾	2,238	2,178	2,183	2,388	2,157
- non dichiarano durata ricerca lavoro ⁽²⁾	283 12.65	270 12.72	263 12.05	244 10.22	251 11.64
-non dichiarano fine attività lavorativa ⁽³⁾	55 2.81	76 4.00	75 3.91	65 3.03	91 4.77
- non dichiarano inizio attività lavorativa ⁽⁴⁾	49 2.58	67 3.67	45 2.44	43 2.07	50 2.75
- risposte incoerenti ⁽⁵⁾	92 4.97	82 4.66	79 4.39	84 4.13	79 4.48
DISOCCUPATI "SELEZIONATI"	1,759	1,676	1,721	1,952	1,686

Tabella 3.1: determinazione del sottocampione su cui sono ricostruite le durate di disoccupazione. Le frequenze sono calcolate in modo sequenziale rispetto al campione selezionato nel passo precedente.

Nel sottocampione di disoccupati di cui sono disponibili tutte le informazioni per poter determinare in modo ottimale la lunghezza del periodo di disoccupazione sono presenti il 79% di coloro che risultano disoccupati almeno una volta nei cinque abbinamenti analizzati. Si può notare che tra coloro che rispondono al quesito sulla durata di ricerca di lavoro le frequenze di non risposte e incoerenze rispetto alle domande retrospettive sono molto più basse rispetto a quelle rilevate nel capitolo precedente. Questo può essere dovuto al fatto che la modalità "non sa" nel quesito sulla ricerca di lavoro è concessa solo se chi risponde non è il soggetto di riferimento dell'intervista.

Una volta selezionato il campione in cui sono disponibili tutte le informazioni retrospettive si può verificare se ci sono particolari diversità rispetto al campione di partenza per quanto riguarda le caratteristiche individuali che, come visto nel precedente capitolo, influenzano il tasso di risposta ad alcune domande.

VARIABILE	MODALITÀ	DISOCCUPATI	DISOCCUPATI "SELEZIONATI"
ETÀ	15-25	3,398 30.49	2,843 32.33
	25-35	3,387 30.39	2,621 29.80
	35-45	2,501 22.44	1,904 21.65
	45-54	1,358 12.19	1,041 11.84
	55-75	500 4.49	385 4.38
SESSO	UOMINI	5,052 45.33	3,768 42.85
	DONNE	6,092 54.67	5,026 57.15
AREA GEOGRAFICA	NORD OVEST	1,642 14.73	1,294 14.71
	NORD EST	1,235 11.08	964 10.96
	CENTRO	1,389 12.46	1,103 12.54
	SUD	4,657 41.79	3,659 41.61
	ISOLE	2,221 19.93	1,774 20.17
ISTRUZIONE	LICENZA ELEMENTARE	1,333 11.98	1,008 11.48
	LICENZE MEDIA	4,216 37.88	3,294 37.51
	PROFESSIONALE	600 5.39	471 5.36
	DIPLOMA	3,841 34.51	3,084 35.12
	TITOLO UNIVERSITARIO	1,140 10.24	924 10.52

Tabella 3.2: confronto tra tutti i disoccupati in almeno una rilevazione e coloro che forniscono i dati retrospettivi per ricostruire la durata della disoccupazione

Dalla Tabella 3.2 si può notare che non ci sono grandi differenze nella composizione tra il campione di chi risulta disoccupato in almeno una delle tre rilevazioni nei quattro abbinamenti e il campione di cui si andrà a ricostruire la durata degli episodi di disoccupazione. Le uniche differenze rilevanti tra le caratteristiche individuali osservabili sono la maggiore presenza nel campione "selezionato" di individui sotto i 25 anni (passano dal 30.5% al 32.3%) e di donne (dal 54.7% al 57.2%). La selezione non sembra invece aver portato a cambiamenti significativi per quanto riguarda la distribuzione geografica del campione e per il grado di istruzione degli intervistati.

I risultati che si otterranno dall'analisi non dovrebbero quindi essere affetti da gravi problemi di selection bias almeno per quanto riguarda l'eterogeneità osservabile. La mancanza di dati

retrospettivi potrebbe comunque essere legata a caratteristiche individuali dell'intervistato che influenzano la durata della disoccupazione ma che non possono essere controllate.

3.2.2 Ricostruzione degli episodi di disoccupazione

Per gli individui per cui sono disponibili tutte le informazioni si va quindi a determinare il limite inferiore e superiore di un periodo che può essere considerato una proxy dell'episodio di disoccupazione, in quanto l'individuo non svolge alcuna attività lavorativa e dichiara di cercare lavoro. Questi limiti sono determinati con assunzioni plausibili ma non tutte verificabili.

Innanzitutto si assume che la distanza tra le rilevazioni sia di tre mesi nei casi in cui le interviste sono effettuate in trimestri consecutivi e di nove mesi quando l'individuo non è osservato per due interi trimestri (tra la seconda e la terza intervista dell'indagine). Per quanto dichiarato da Istat questa distanza teorica dovrebbe essere sempre rispettata dai vari rilevatori e, in caso contrario, l'intervista è comunque riferita alla settimana di riferimento corretta. L'analisi sui dati retrospettivi vista nel capitolo precedente, in particolare sui mesi di ricerca di lavoro, pone comunque dei dubbi sul funzionamento di tale procedimento sul campo.

Si ipotizza inoltre che i dati retrospettivi, se non sono incoerenti con le condizioni lavorative determinate nei trimestri di rilevazione, siano corretti. Altra ipotesi che l'alto numero di incoerenze trovate nel precedente capitolo potrebbe mettere in discussione, ma necessaria se si vuole utilizzare tutta l'informazione disponibile.

Nel determinare la durata della disoccupazione si farà affidamento sempre all'intervista immediatamente successiva all'evento che determina la transizione da uno stato all'altro. Ad esempio per la durata della disoccupazione ci si baserà sulla prima rilevazione in cui l'intervistato risulta disoccupato. Nella ricostruzione degli episodi di disoccupazione non è stata invece considerata la durata in mesi della ricerca di lavoro dichiarata nelle interviste successive alla prima in cui l'individuo risulta disoccupato. Questa scelta è stata dettata dall'alto numero di incoerenze osservato nelle risposte sulla durata della ricerca di lavoro e dall'assunzione che la durata della ricerca di lavoro sia meglio specificata nella prima intervista in cui si è disoccupati.

La prima quantità da determinare per tutti i disoccupati è la durata dell'episodio di disoccupazione rilevata nella prima indagine in cui l'individuo è disoccupato, che

indicheremo con T_0 . Potenzialmente T_0 è dato dal minimo tra la durata della ricerca di lavoro e i mesi trascorsi da quando l'individuo ha smesso di lavorare (se dichiara di aver avuto un'occupazione in passato). Tuttavia quest'ultima grandezza, a causa dell'assenza del mese in cui si è svolta l'intervista, non può essere determinata in modo univoco a partire dalle risposte presenti nel questionario. Può essere determinata una distanza minima (T_{0min}), sotto l'ipotesi che la rilevazione sia effettuata nel primo mese del trimestre di indagine, e una distanza massima (T_{0max}) se l'intervista è effettuata nell'ultimo mese utile.

Queste due quantità coincideranno se l'individuo non ha mai lavorato in passato – la durata della disoccupazione coinciderà con la durata della ricerca di lavoro – o se la durata della ricerca di lavoro è inferiore alla distanza tra il mese in cui ha smesso di lavorare e il primo mese del trimestre in cui risulta disoccupato.

Per chi risulta disoccupato per la prima volta nelle successive interviste T_0 è corretto in base alla distanza teorica dalla rilevazione precedente in cui non risulta disoccupato.

T_{0min} e T_{0max} saranno quindi vincolati a non superare:

- i tre mesi nel caso in cui l'intervista della prima disoccupazione sia svolta nel trimestre successivo a quella precedente (in cui l'individuo non risulta disoccupato);
- i nove mesi nel caso in cui l'individuo risulti disoccupato per la prima volta nella terza rilevazione dell'indagine.

Una volta determinate queste due quantità (T_{0min} e T_{0max}) per ogni individuo nella prima rilevazione in cui risulta disoccupato, è possibile specificare il limite inferiore e superiore della durata dello spell di disoccupazione a seconda delle informazioni disponibili nelle successive rilevazioni (se ci sono), a partire dalle assunzioni specificate in seguito.

3.2.3 Transizione verso l'occupazione

Un disoccupato transita verso l'occupazione se, in una rilevazione successiva a quella in cui risulta disoccupato, o risulta essere nello stato di occupato o dichiara di aver terminato la sua ultima attività lavorativa in un mese successivo all'ultimo trimestre in cui era stato classificato disoccupato. Quest'ultimo è un caso di mobilità nel mercato del lavoro che non sarebbe individuabile se non si utilizzassero i dati retrospettivi.

Nel caso in cui la transizione sia rilevata direttamente nei dati sezionali il limite inferiore della durata della disoccupazione (che indicheremo con T_{min}) sarà il minimo tra:

- la distanza (in mesi) tra la fine dell'attività lavorativa precedente – se ha dichiarato di aver già lavorato - e l'inizio della sua nuova occupazione;
- la somma tra T_{0max} e la distanza tra il mese in cui ha iniziato a lavorare e l'ultimo mese del trimestre in cui l'individuo è stato classificato disoccupato per la prima volta.

Se l'occupazione è iniziata nel trimestre in cui l'individuo risulta disoccupato per la prima volta l'ultima quantità potrebbe anche essere minore di T_{0min} . Ciò accade, ad esempio, se una persona non ha mai lavorato e nel primo trimestre del 2004 alla prima intervista dell'indagine dichiara di cercare lavoro da un mese. Se nel secondo trimestre, alla seconda intervista, dichiara di aver iniziato a lavorare a gennaio T_{min} risulterebbe addirittura negativo. Per costruzione si è quindi imposto il vincolo che il limite inferiore del periodo di disoccupazione sia maggiore o uguale al periodo minimo, arrotondato per eccesso, in cui l'individuo risulta disoccupato a prescindere dal mese in cui esce verso l'occupazione: $T_{0min} + 1 + k$ con k che rappresenta la distanza che intercorre tra le rilevazioni in cui l'individuo risulta disoccupato.

Il vincolo di durata minima per chi transita verso l'occupazione (o l'inattività) quindi sarà:

- $T_{0min} + 1$ se l'individuo transita dopo essere risultato disoccupato in una sola rilevazione;
- $T_{0min} + 1 + 3$ se l'individuo transita verso l'occupazione nella terza intervista dell'indagine dopo essere stato disoccupato nelle prime due;
- $T_{0min} + 1 + 9$ se l'individuo transita verso l'occupazione nella quarta intervista dell'indagine dopo essere stato disoccupato nella seconda e nella terza.

Il limite superiore dello spell di disoccupazione, ovvero la durata massima in cui l'individuo è rimasto disoccupato (che indicheremo con T_{max}) prima di transitare verso l'occupazione, è il minimo tra:

- la distanza (in mesi) tra la fine dell'attività lavorativa precedente – se ha dichiarato di aver già lavorato - e l'inizio della sua nuova occupazione;
- la somma tra T_{0min} e la distanza tra il mese in cui ha iniziato a lavorare e il primo mese del trimestre in cui l'individuo risulta disoccupato per la prima volta.

In questo caso, oltre a imporre la stessa durata minima vista in precedenza, si impone una durata massima per i dati di flusso. Chi è disoccupato per la prima volta nella seconda osservazione ed esce verso l'occupazione nella terza (e ultima) osservazione non potrà avere una durata massima di disoccupazione superiore ai 12 mesi (il periodo di tempo entro cui dovrebbero essere svolte le tre rilevazioni).

Per esemplificare come sono determinati i due limiti si consideri una persona presente nell'abbinamento tra le prime due rilevazioni del 2004 e la prima del 2005. Disoccupata nel primo trimestre del 2004 dichiara di cercare un'occupazione da 5 mesi e aver smesso di lavorare nel settembre del 2003. T_{0min} sarà quindi pari a 4 mesi (la distanza tra settembre 2004 e inizio gennaio 2004) mentre T_{0max} sarà 5 mesi: se è intervistata nel marzo del 2004 la durata della ricerca di lavoro è inferiore alla distanza dalla fine dell'ultima attività.

L'individuo risulta ancora disoccupato nella seconda osservazione mentre nella terza osservazione è occupato e dichiara di aver iniziato a lavorare nel febbraio 2005. Il limite inferiore della durata di disoccupazione (T_{min}) sarà di 16 mesi (somma tra T_{0max} e i mesi tra marzo e febbraio) mentre T_{max} sarà pari a 18 mesi (somma tra T_{0min} e la distanza tra l'inizio di gennaio e la fine di marzo).

Su chi transita verso l'occupazione per un breve periodo, senza però risultare occupato in alcuna rilevazione, non si ha a disposizione il mese in cui ha iniziato la sua occupazione, ma solo quando la ha terminata. I limiti entro cui termina in periodo di disoccupazione saranno quindi meno accurati.

Il limite inferiore sarà la durata minima per chi transita verso l'occupazione del tipo $T_{0min} + 1+k$ vista in precedenza.

Il limite superiore è stato ricostruito ipotizzando che l'occupazione temporanea che ha determinato la transizione sia durata solo un mese e che l'intervista in cui l'individuo risulta disoccupato per la prima volta sia stata effettuata nel primo mese del trimestre di riferimento.

Date queste ipotesi il limite superiore è il minimo tra:

- la distanza tra la fine della penultima occupazione e l'inizio presunto dell'ultima occupazione, se aveva dichiarato di aver già lavorato nella prima rilevazione in cui è disoccupato;
- la somma tra T_{0min} e la distanza tra il mese precedente a quello in cui dichiara di aver terminato la sua ultima occupazione (l'episodio lavorativo che determina la sua uscita dalla disoccupazione) e il primo mese del trimestre in cui l'individuo è stato classificato disoccupato per la prima volta.

Inoltre si impone che il limite superiore non superi i 12 mesi per chi è soggetto a campionamento da flusso, ovvero non è disoccupato nella prima osservazione presente nell'abbinamento.

Si consideri, ad esempio, un individuo presente nell'abbinamento tra l'ultimo trimestre del 2004, il primo e il quarto del 2005. L'intervistato è inattivo nella prima e nella terza osservazione mentre nella seconda (primo trimestre del 2005) risulta disoccupato e dichiara di cercare lavoro da 6 mesi. T_0 sarà unico e sarà pari a 3 mesi in quanto si è assunta l'ipotesi che l'intervista sia svolta nella distanza "teorica" tra le rilevazioni (tre mesi in questo caso). Nella terza rilevazione dichiara di aver smesso di lavorare nel settembre del 2005: T_{\min} sarà pari a 4 mesi ($T_{0\min} + 1$) mentre T_{\max} sarà uguale a 11 mesi (somma fra $T_{0\min}$ e la distanza in mesi tra gennaio e agosto).

3.2.4 Transizione verso l'inattività

Per chi transita dalla disoccupazione all'inattività le informazioni sono inferiori in quanto non è possibile stabilire in quale momento l'individuo transiti. Gli unici dati retrospettivi utili per limitare l'intervallo in cui avviene l'uscita dalla disoccupazione sono le risposte, date però solo da una piccola frazione di coloro che transitano dalla disoccupazione all'inattività, sul periodo passato dall'ultima azione di ricerca di lavoro. Quando disponibile questa informazione verrà utilizzata, ma l'intervallo in cui un individuo può essere o meno transitato sarà in genere più ampio nel caso in cui la destinazione sia l'inattività invece dell'occupazione.

La durata minima dello spell di disoccupazione per chi passa dalla disoccupazione all'inattività sarà il massimo tra:

- $T_{0\min} + 1 + k$ con k che rappresenta la distanza che intercorre tra le rilevazioni in cui l'individuo risulta disoccupato, come spiegato in precedenza per chi transita verso l'occupazione;
- la differenza tra la durata della disoccupazione se l'individuo fosse disoccupato fino al primo mese del trimestre in cui risulta inattivo e la distanza dall'ultima volta che ha cercato lavoro.

Se nella rilevazione in cui risulta inattivo l'intervistato non risponde alla domanda "*Quanti mesi sono passati dall'ultima volta che ha fatto qualcosa per cercare lavoro?*" la seconda quantità non sarà disponibile.

La durata massima del periodo di disoccupazione prima di transitare verso l'inattività sarà data, invece, dal minimo tra:

- la durata della disoccupazione se l'individuo fosse intervistato nell'ultimo mese del trimestre in cui risulta inattivo;
- la differenza tra la durata della disoccupazione se l'individuo fosse intervistato nell'ultimo mese del trimestre in cui risulta inattivo e il numero di mesi da cui sostiene di aver smesso di cercare lavoro.

Il limite superiore sarà vincolato a essere maggiore o uguale a $T_{0min} + 1 + k$ in modo da correggere eventuali risposte incoerenti (numero di mesi superiore a quello che intercorre tra le due interviste) sul quando ha smesso di cercare lavoro da parte dell'intervistato.

Si ricostruisca, ad esempio, il primo episodio di disoccupazione per una persona presente nell'abbinamento tra il primo e l'ultimo trimestre del 2004 e il primo del 2005. L'individuo è inattivo nella prima e ultima rilevazione e disoccupato nella seconda in cui dichiara di essere alla ricerca della sua prima occupazione da 2 mesi ($T_{0min}=T_{0max}$). Nella terza osservazione non dichiara da quanto ha smesso di cercare lavoro. Il limite inferiore della disoccupazione sarà quindi posto pari a 3 mesi (la durata minima dell'episodio arrotondata per eccesso) mentre il limite superiore sarà di 5 mesi, ovvero la durata della disoccupazione se l'individuo avesse smesso di cercare lavoro pochi giorni prima della terza osservazione presente nel campione.

3.2.5 Individui censurati a destra

Gli individui che, una volta rilevati disoccupati, non transitano verso un altro stato prima della terza occasione presente nell'abbinamento, sono considerati censurati. Per chi non esce dalla disoccupazione è possibile calcolare un limite inferiore della durata dell'episodio di disoccupazione, ma non è possibile calcolarne uno superiore. Si potrebbe determinare un limite superiore oltre il quale si è certi che il disoccupato non è più nell'intervallo di osservazione: questa quantità non è però informativa sulla durata della disoccupazione ma solo sulla durata massima del periodo di disoccupazione osservabile sui dati disponibili.

La durata censurata della disoccupazione sarà quindi:

- $T_{0min} + 12$ se l'individuo è disoccupato in tutte le tre osservazioni del campione
- $T_{0min} + 9$ se l'individuo è disoccupato nella seconda e terza osservazione presente nel campione e l'abbinamento coinvolge le prime tre rilevazioni dell'indagine

- $T_{0min} + 3$ se l'individuo è disoccupato nella seconda e terza osservazione presente nel campione e l'abbinamento coinvolge le ultime tre rilevazioni dell'indagine
- T_{0min} se l'individuo entra nella disoccupazione nell'ultima osservazione.

3.2.6 Il dataset

Una volta determinati i limiti superiori e inferiori della durata degli episodi si è costruito un dataset in cui per ogni individuo, che è risultato disoccupato almeno una volta, sono presenti:

- il limite inferiore (T_{min}) e superiore (T_{max}) della durata dell'episodio di disoccupazione;
- la condizione lavorativa successiva alla disoccupazione (se non è censurato);
- gli indici di censura sinistra e destra;
- il tempo minimo (T_{0min}) e massimo (T_{0max}) già trascorso in disoccupazione per chi è censurato a sinistra (queste quantità saranno poste a zero per il campionamento da flusso);
- alcune variabili esplicative raccolte nella prima intervista in cui risulta disoccupato (età, titolo di studio, regione di residenza, sesso).

I limiti entro cui un individuo transita dalla disoccupazione a un altro stato a volte possono coincidere determinando così in modo univoco la durata della disoccupazione.

Nei casi in cui l'individuo esce verso l'occupazione i due limiti coincidono se:

- la durata della disoccupazione coincide con la distanza tra il termine di una precedente occupazione e l'inizio di quella successiva;
- le durate coincidono coi vincoli posti nella ricostruzione per rispettare le ipotesi sulla modalità dell'indagine, in particolare sul rispetto sistematico da parte dei rilevatori della distanza tra le rilevazioni.

Per chi transita dalla disoccupazione all'inattività la durata dell'episodio è unica se nella prima occasione in cui è disoccupato T_{0min} coincide con T_{0max} e quando è classificato come inattivo risponde al quesito sul numero di mesi trascorsi da quando ha smesso di cercare lavoro.

Si ricorda inoltre che per gli individui censurati a destra la durata dell'episodio di disoccupazione è unica e coincide con il limite inferiore.

A partire da queste informazioni, fortemente condizionate dalle ipotesi sulla rilevazione e sul processo di memoria che sono state assunte durante la ricostruzione, nel seguito si descrive

quali siano i contributi alla verosimiglianza per ciascuna tipologia di ricostruzione necessari per stimare in modo non distorto un modello di durata della disoccupazione.

3.3 Informazioni disponibili e modelli di durata

Per determinare in che modo la durata della disoccupazione influenza le dinamiche del mercato di lavoro si è scelto di adottare dei modelli di durata di tipo parametrico (o semiparametrico) e di supporre quindi nota la funzione di sopravvivenza a meno di un insieme di parametri che saranno stimati massimizzando la funzione di verosimiglianza.

Nel dataset che si vuole analizzare τ , ovvero il tempo già trascorso in disoccupazione per gli individui che risultano disoccupati nella prima rilevazione dell'abbinamento, è determinato da $T_{0\min}$. Non è possibile utilizzare $T_{0\max}$ per ricostruire τ poiché non si può affermare con certezza che un individuo sia disoccupato da almeno $T_{0\max}$ mesi: si andrebbe a condizionare la probabilità di sopravvivere a un evento che potrebbe non essere accaduto.

Il contributo alla verosimiglianza dato dagli individui per cui il limite inferiore e superiore del periodo di disoccupazione non coincidono, e quindi non è possibile determinare univocamente il t_i , sarà diverso da quello visto in precedenza.

Per sfruttare al meglio l'informazione a disposizione, il contributo alla verosimiglianza dell' i -esimo individuo che esce dalla disoccupazione tra t_{\min_i} e t_{\max_i} è :

$$P\left[t_{\min_i} \leq T \leq t_{\max_i} \mid T > \tau_i\right]$$

Ricordando la definizione di funzione di sopravvivenza vista nel paragrafo 3.1 il contributo alla verosimiglianza dell' i -esimo individuo può essere riformulato come:

$$\frac{S(t_{\min_i}) - S(t_{\max_i})}{S(\tau_i)} .$$

Sia d_i una variabile binaria che assume il valore uno se $t_{\min_i} < t_{\max_i}$ e si ponga, per semplificare la notazione, $t_i = t_{\min_i}$ quando d_i è pari a zero. La funzione di verosimiglianza valida per tutto il campione, tenendo conto dell'eterogeneità osservabile x è la seguente:

$$L(\underline{\alpha}, \underline{\beta}; t_{\min}, t_{\max}, \tau, x) = \prod_{i=1}^n \frac{S(t_{\min_i}, \underline{x}_i)}{S(\tau_i, \underline{x}_i)} \theta(t_{\min_i}, \underline{x}_i)^{c_i(1-d_i)} \left(1 - \frac{S(t_{\max_i}, \underline{x}_i)}{S(t_{\min_i}, \underline{x}_i)} \right)^{d_i c_i}. \quad (3.3)$$

$\underline{\alpha}$ rappresenta il vettore dei parametri di forma che si andranno a stimare a seconda della distribuzione che si assumerà per la variabile T, mentre $\underline{\beta}$ sono i parametri di scala che determinano l'effetto proporzionale delle x sul rischio di transizione.

3.3.1 Scelta della funzione di rischio di base

La scelta della distribuzione della v.c. T è essenziale nel determinare la relazione tra la durata della disoccupazione e le dinamiche del mercato del lavoro, in quanto a seconda della funzione scelta si fanno delle assunzioni sulla forma della funzione di rischio.

Un modello comune nell'analisi dei dati di durata ipotizza che la v.c. T segua una distribuzione di Weibull. Questa ipotesi però implica che la funzione di rischio del modello sia monotona, assunzione difficilmente giustificabile nei dati che si vogliono analizzare in quanto l'intervallo di tempo considerato è molto ampio (alcuni periodi di disoccupazione superano i 5 anni) e non è molto realistico pensare che la probabilità condizionata di uscire dalla disoccupazione cresca (o diminuisca) in modo costante.

Per non imporre particolari condizioni alla forma della funzione di rischio di T, di cui non si può avere nemmeno una rappresentazione empirica, dal momento che per molti intervistati non si è riusciti a determinare univocamente la durata dell'episodio, si potrebbe scegliere di utilizzare un modello basato sulla funzione esponenziale a tratti.

Questo modello, come visto nel paragrafo 3.1, impone che la funzione di rischio di base sia costante - e che quindi T si distribuisca secondo un'esponenziale - in determinati intervalli di tempo scelti arbitrariamente. Il modello va a stimare i livelli costanti di rischio negli intervalli individuando la dipendenza dell'uscita dalla disoccupazione dalla durata.

Negli intervalli in cui l'individuo non transita il contributo alla verosimiglianza sarà:

$$\frac{S(\lambda_{k-h})}{S(\lambda_{k-h-1})} = \frac{\exp\{\lambda_{k-h} e^{\alpha_{k-h} + X_i \beta}\}}{\exp\{\lambda_{k-h-1} e^{\alpha_{k-h} + X_i \beta}\}}$$

dove λ_{k-h} e λ_{k-h-1} rappresentano gli estremi dell'intervallo e α_{k-h} è il parametro da stimare che rappresenta il rischio di base per l'intervallo.

Il fatto che molti episodi non abbiano un'unica durata comporta però dei problemi nell'attribuire il contributo alla verosimiglianza nel caso in cui t_{\min_i} e t_{\max_i} si trovino in due intervalli di tempo diversi tra quelli determinati arbitrariamente.

Nel caso in cui $\lambda_{k-1} < t_{\min_i} < t_{\max_i} \leq \lambda_k$ e $d_i c_i = 1$ il contributo alla verosimiglianza nel k-esimo intervallo $(\lambda_{k-1}, \lambda_k]$ dell'i-esimo individuo sarà:

$$\frac{\exp\{t_{\min_i} e^{\alpha_k + X_i \beta}\} - \exp\{t_{\max_i} e^{\alpha_k + X_i \beta}\}}{\exp\{\lambda_{k-1} e^{\alpha_k + X_i \beta}\}}.$$

Il contributo alla verosimiglianza dell'i-esimo individuo sarà quindi la produttoria dei contributi dati in ogni intervallo di tempo in cui è disoccupato, come nel caso in cui la lunghezza dell'episodio è determinata in maniera univoca.

Nel caso in cui i limiti (inferiore e superiore) della durata dell'episodio cadano in due intervalli diversi - ovvero se $\lambda_{k-1} < t_{\min_i} \leq \lambda_k < t_{\max_i} \leq \lambda_{k+1}$ - non sarà più possibile calcolare il contributo alla verosimiglianza dell'i-esimo individuo come produttoria dei contributi calcolati separatamente in ogni intervallo, in quanto non è più possibile scomporre la probabilità condizionata:

$$P[t_{\min} < T < t_{\max} | T > \lambda_{k-1}] \neq P[t_{\min} < T < \lambda_k | T > \lambda_{k-1}] \cdot P[\lambda_k < T \leq t_{\max} | T > \lambda_k]$$

Il contributo alla verosimiglianza da parte dell'individuo alla stima del livello di rischio negli intervalli avrà quindi le stesse modalità viste in precedenza per i casi in cui l'individuo non esce dallo spell, mentre se i limiti appartengono a due intervalli diversi ma consecutivi - $t_{\min} \in (\lambda_{k-1}, \lambda_k]$ e $t_{\max} \in (\lambda_k, \lambda_{k+1}]$ - il contributo nei due ultimi intervalli dello spell avrà questa forma:

$$\begin{aligned} & \frac{S(t_{\min}; \alpha_k)}{S(\lambda_{k-1}; \alpha_k)} \frac{S(t_{\min}; \alpha_k) - S(t_{\max}; \alpha_k, \alpha_{k+1})}{S(t_{\min}; \alpha_k)} = \\ & = \frac{S(t_{\min}; \alpha_k)}{S(\lambda_{k-1}; \alpha_k)} \left[1 - \frac{S(\lambda_k; \alpha_k)}{S(t_{\min}; \alpha_k)} \frac{S(t_{\max}; \alpha_{k+1})}{S(\lambda_k; \alpha_{k+1})} \right] = \\ & = \frac{\exp\{t_{\min_i} e^{\alpha_k + X_i \beta}\}}{\exp\{\lambda_{k-1} e^{\alpha_k + X_i \beta}\}} \left[1 - \frac{\exp\{\lambda_k e^{\alpha_k + X_i \beta}\}}{\exp\{t_{\min_i} e^{\alpha_k + X_i \beta}\}} \frac{\exp\{t_{\max_i} e^{\alpha_{k+1} + X_i \beta}\}}{\exp\{\lambda_k e^{\alpha_{k+1} + X_i \beta}\}} \right]. \end{aligned}$$

In questo caso la verosimiglianza non può più essere ricostruita come prodotto dei contributi portati dagli individui nei singoli intervalli e la sua applicazione ai dati che si vogliono analizzare diventa particolarmente gravosa dal punto di vista informatico e computazionale.

Non è stato quindi possibile implementare il modello basato sull'esponenziale a tratti nel corso del nostro studio, nonostante possa essere ritenuto quello più adatto alla descrizione dell'effetto della durata della disoccupazione.

Se si vogliono utilizzare tutte le informazioni disponibili per verificare l'esistenza di una dipendenza dalla durata, controllando l'effetto dell'eterogeneità osservata, non resta che assumere che T abbia una distribuzione di tipo Weibull, nonostante le problematiche viste in precedenza.

L'obiettivo è di confrontare le stime ottenute da questo modello a seconda della quantità di informazione utilizzata e dello stato verso cui transitano i disoccupati controllando se ci sono variazioni significative sul parametro di forma della distribuzione e sugli effetti delle applicative.

3.3.2 Stima non parametrica della sopravvivenza

Prima di procedere alla stima parametrica del modello può essere utile analizzare la distribuzione empirica delle durate. La lunghezza dell'episodio non è determinata in modo univoco nel 42.7% dei casi, quindi la funzione di sopravvivenza basata sulla distribuzione empirica di uno solo dei due limiti non è molto informativa. Può essere utile confrontare le due funzioni di sopravvivenza empiriche in quanto quella effettiva dell'episodio di disoccupazione sarà compresa tra queste due funzioni.

La funzione di sopravvivenza empirica è così calcolata per ogni diversa durata osservata degli episodi (indicate con t_j):

$$\hat{S}(t_j) = \prod_{i=1}^j \left(1 - \frac{h_i}{(h_i + m_i)} \right)$$

dove h_j è il numero di individui che transita in t_j e m_j è il numero di individui che sono censurati tra t_j e t_{j+1} . $\hat{S}(t_j)$ è lo stimatore di Kaplan-Meier, per una trattazione più completa del quale si veda, tra gli altri, Kiefer (1988).

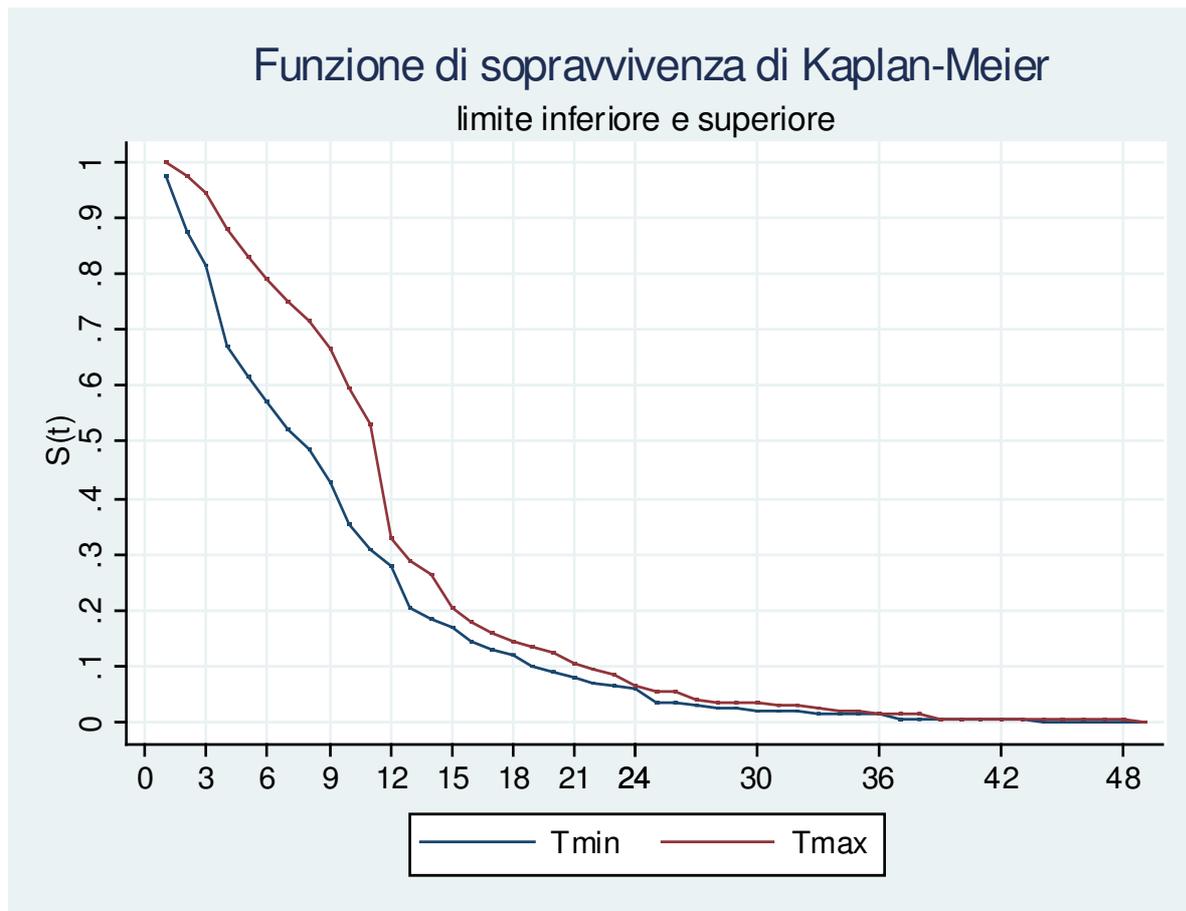


Figura 3.1: funzioni di sopravvivenza di Kaplan-Meier stimate su t_{\min} e t_{\max} . Non è rappresentato il valore della funzione quando l'episodio supera i cinquanta mesi.

Dalla Figura 3.1 si nota come le due funzioni di sopravvivenza abbiano un diverso comportamento per una durata della disoccupazione inferiore all'anno. Ad esempio dopo 6 mesi quasi il 60% del campione è sicuramente ancora disoccupato, ma solo il 20% del campione può essere considerato come certamente transitato. Inoltre, l'80% dei disoccupati è transitato dopo un anno e tre mesi e che poco più del 10% è sicuramente ancora attivo nella ricerca di lavoro dopo un anno e mezzo.

La diversità tra le due funzioni di sopravvivenza empirica conferma la necessità di tenere in considerazione entrambi i limiti della durata dell'episodio di disoccupazione.

Si possono confrontare anche le due funzioni di rischio empiriche ottenute ipotizzando che i due limiti indichino il mese esatto in cui avviene la transizione. La stima avviene non parametricamente col metodo del Kernel con assunzioni standard. Per dettagli si veda, tra gli altri, Loader (1999).

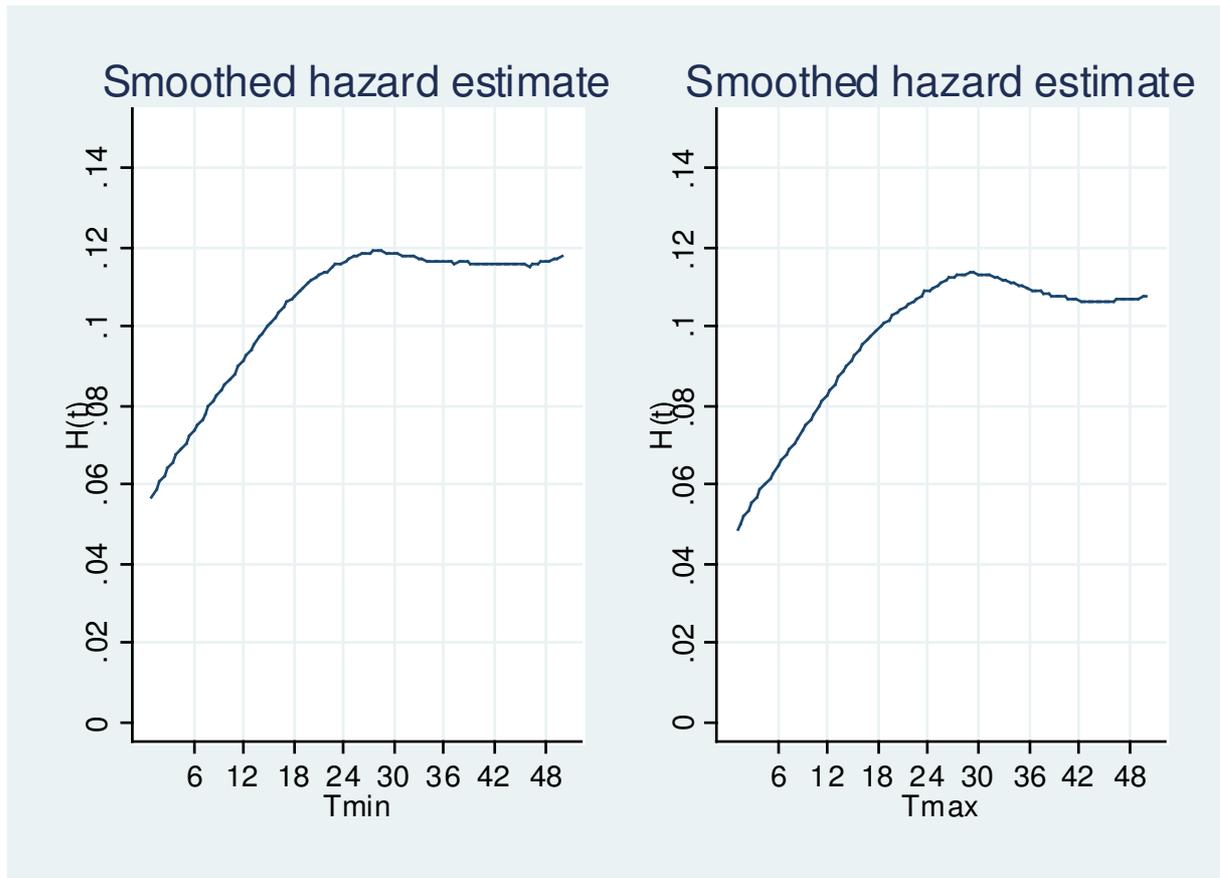


Figura 3.2: stima non parametrica delle funzioni di rischio per t_{\min} e t_{\max}

Le due funzioni di rischio empiriche calcolate sui due limiti (Figura 3.2) sembrano avere un andamento molto simile. Si può notare come il rischio di uscire dalla disoccupazione sia crescente nei primi due anni, in cui transita più del 90% del campione. Ipotizzare che la durata della disoccupazione abbia una distribuzione di tipo Weibull non è quindi in netto contrasto con i dati osservati, anche se dal terzo anno la probabilità condizionata di transitare a un'altra condizione lavorativa sembra restare stabile o addirittura diminuire col passare del tempo, soprattutto per t_{\max} .

3.3.3 Stima parametrica

Dall'analisi empirica dei due limiti entro cui si trova l'effettiva durata dei vari spell emerge la necessità di utilizzare un modello che tenga conto di tutta l'informazione disponibile. Inoltre non è da escludere che il rischio di transitare verso una condizione lavorativa diversa dalla disoccupazione possa essere monotono. Si va quindi ad applicare la verosimiglianza (3.3).

Se T si distribuisce come una Weibull la funzione di verosimiglianza da massimizzare sarà:

$$\prod_{i=1}^n \exp \left[\left(\tau_i^\alpha - t_{\min_i}^\alpha \right) e^{x_i' \beta} \right] \left(\alpha t_{\min_i}^{\alpha-1} e^{x_i' \beta} \right)^{c_i(1-d_i)} \left\{ 1 - \exp \left[\left(t_{\min_i}^\alpha - t_{\max_i}^\alpha \right) e^{x_i' \beta} \right] \right\}^{d_i c_i}$$

Il vettore x_i dell'i-esimo individuo sarà composto da una serie di variabili binarie che riguardano la sua età, zona di residenza, istruzione e sesso registrate nell'indagine nella prima rilevazione in cui è risultato essere disoccupato. Le uniche variabili potenzialmente time-varying nel campione sono l'istruzione e, ovviamente, l'età poiché coloro che cambiano zona di residenza escono dal campione di osservazione. Tuttavia, dal momento che il periodo di disoccupazione osservato è al massimo di 12 mesi il cambiamento del livello di istruzione da parte di chi è disoccupato non dovrebbe essere significativo, per cui le variabili sono assunte costanti nel tempo.

Le stime di massima verosimiglianza ottenute dal modello sono presentate in Tabella 3.3

		N° CAMPIONE =8781		
VARIABILI ESPLICATIVE (β)		COEFFICIENTE	ERRORE STD.	P-VALUE
AREA RESIDENZA	NORD OVEST	-.032	.054	.549
	NORD EST	.094	.059	.108
	SUD	-.275	.045	.000
	ISOLE	-.224	.051	.000
ETA'	15-25	.101	.037	.007
	36-45	.058	.041	.152
	46-55	.057	.050	.256
	56-75	.089	.074	.231
LIVELLO ISTRUZIONE	ELEMENTARE	.064	.051	.210
	MEDIA	-.042	.034	.217
	DIPLO. 3 ANNI	.103	.065	.115
	LAUREA	.019	.053	.714
SESSO	DONNA	.064	.029	.029
	COSTANTE	-3.349	.064	.000
PARAMETRO DI FORMA (α)				
	Ln (α)	.293	.008	.000

Tabella 3.3: Stime di massima verosimiglianza del modello Weibull che utilizza t_{\min} e t_{\max}

Dal momento che $\ln(\alpha)$ è significativamente maggiore di zero la dipendenza dalla durata della disoccupazione è positiva: la probabilità di terminare la propria disoccupazione condizionata al fatto che si è rimasti disoccupati aumenta con passare dei mesi. La stima quindi conferma ciò che era emerso dalle funzioni di rischio empiriche calcolate separatamente su t_{\min} e t_{\max} .

Rispetto all'individuo scelto come baseline (diplomato tra i 25 e i 35 anni residente nelle regioni centrali) c'è una maggiore probabilità di uscita dalla disoccupazione per chi ha meno

di 25 anni e per le donne. Inoltre ci sono importanti differenze dal punto di vista geografico: nelle regioni del Sud e delle Isole il rischio di uscire dalla disoccupazione è di molto inferiore rispetto al resto del campione.

Dal momento che si sta considerando l'uscita verso entrambe le condizioni lavorative l'interpretazione del modello non è molto chiara dal punto di vista socio-economico. Per poter dare una migliore interpretazione all'effetto della durata della disoccupazione e dell'eterogeneità osservata sulla probabilità di terminare lo spell è necessario considerare verso quale condizione lavorativa transita il disoccupato.

Se si ipotizza che i rischi siano competitivi, con la probabilità di transitare verso l'occupazione indipendente dalla probabilità di transitare verso l'inattività, si può massimizzare separatamente la funzione di verosimiglianza a seconda dello stato verso cui esce il disoccupato.

		N° CAMPIONE = 8781		
VARIABILI ESPLICATIVE (β)		COEFFICIENTE	ERRORE STD.	P-VALUE
AREA RESIDENZA	NORD OVEST	.157	.082	.054
	NORD EST	.300	.087	.001
	SUD	-.682	.075	.000
	ISOLE	-.624	.087	.000
ETA'	15-25	.054	.060	.370
	36-45	-.093	.069	.177
	46-55	-.426	.096	.000
	56-75	-.598	.155	.000
LIVELLO ISTRUZIONE	ELEMENTARE	-.124	.095	.196
	MEDIA	-.247	.058	.000
	DIPLO. 3 ANNI	.041	.106	.702
	LAUREA	.177	.080	.026
SESSO	DONNA	-.458	.050	.000
	COSTANTE	-3.425	.100	.000
PARAMETRO DI FORMA (α)				
	Ln (α)	.228	.015	.000

Tabella 3.4: Stime di massima verosimiglianza del modello Weibull che utilizza t_{\min} e t_{\max} .
Transizioni verso l'occupazione.

I risultati per la transizione verso l'occupazione sono in Tabella 3.4

La dipendenza dalla durata è significativa e positiva, ma è inferiore rispetto a quella stimata nel modello precedente: in questo caso il parametro di forma α ha un coefficiente pari a 1.26, mentre nel modello globale è di 1.34.

L'effetto dell'eterogeneità osservata è importante: la probabilità di trovare un'occupazione è influenzata in modo significativo da tutte le caratteristiche dell'individuo. Hanno minori possibilità di transitare verso l'occupazione le donne con più di 45 anni che vivono nelle

regioni del Sud o delle Isole e che non hanno un titolo di studio superiore alla licenza media. Hanno invece maggiori possibilità di transitare verso l'occupazione i disoccupati con istruzione universitaria.

VARIABILI ESPLICATIVE (β)		COEFFICIENTE	N° CAMPIONE = 8781	
			ERRORE STD.	P-VALUE
AREA RESIDENZA	NORD OVEST	-.178	.071	.013
	NORD EST	-.055	.078	.482
	SUD	-.092	.057	.109
	ISOLE	-.047	.063	.455
ETA'	15-25	.136	.047	.004
	36-45	.136	.050	.006
	46-55	.276	.060	.000
	56-75	.396	.085	.000
LIVELLO ISTRUZIONE	ELEMENTARE	.149	.061	.014
	MEDIA	.056	.041	.171
	DIPLO. 3 ANNI	.140	.082	.089
	LAUREA	-.096	.071	.175
SESSO	DONNA	.345	.037	.000
	COSTANTE	-4.280	.082	.000
PARAMETRO DI FORMA (α)				
	Ln (α)	.333	.010	.000

Tabella 3.5: Stime di massima verosimiglianza del modello Weibull che utilizza t_{\min} e t_{\max} . Transizioni verso l'inattività

Anche se si considera la transizione verso l'inattività (tabella 3.5) la dipendenza dalla durata è positiva. Anzi, in questo caso la dipendenza dalla durata è superiore a quella stimata non facendo distinzioni sul tipo di transizione: α , il parametro di forma, in questo caso è pari a 1.39.

L'effetto dell'eterogeneità osservata è molto diverso rispetto a quello registrato per chi transita verso l'occupazione. In questo caso il rischio di uscire verso l'inattività è superiore per le donne e per chi ha la licenza elementare. Per quanto riguarda l'età, la minor probabilità di transitare verso l'inattività è di chi è tra i 25 e i 35 anni, per poi aumentare con l'età. Non sembra esserci molta differenza tra le varie regioni di residenza: l'unico coefficiente significativo è negativo e riguarda chi vive nel Nord-Ovest.

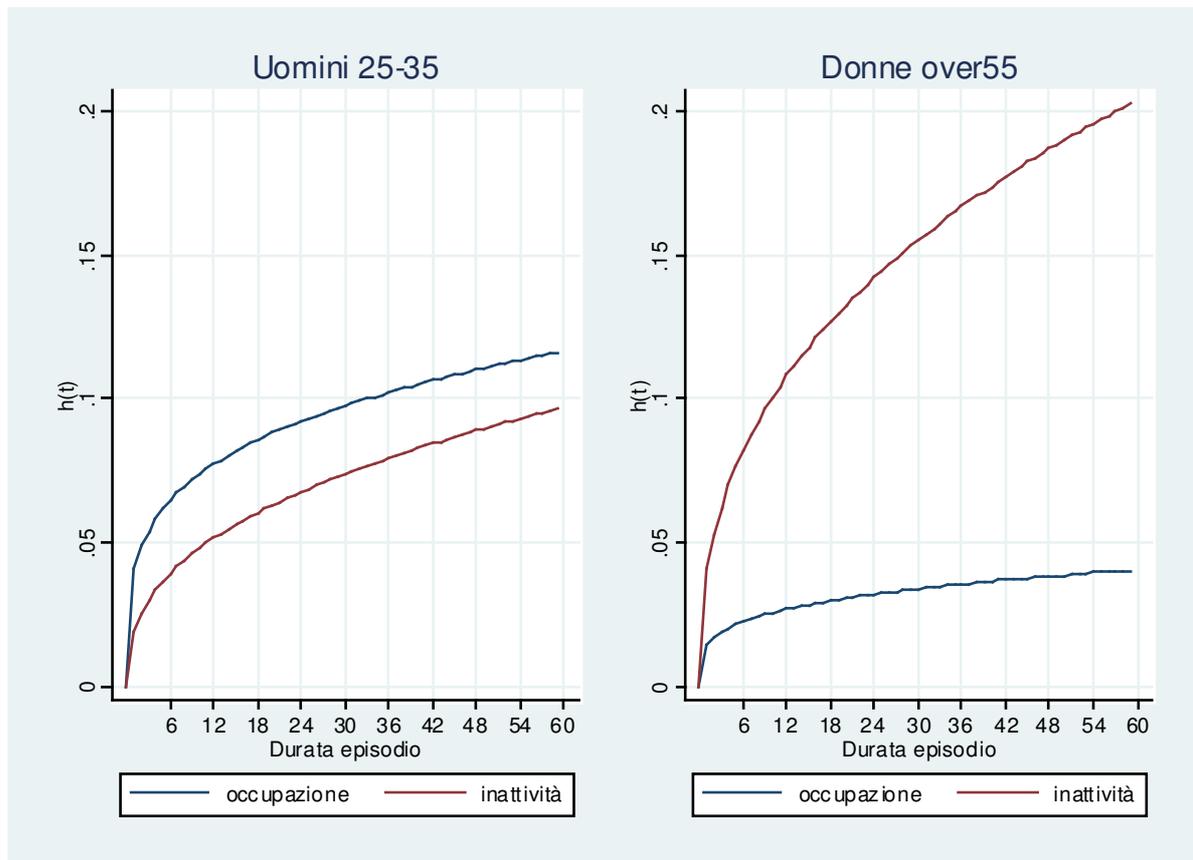


Figura 3.3: funzioni di rischio stimate per i diplomati residenti nelle regioni centrali, distinte per tipo di transizione

Dalla Figura 3.3 si nota che per l'individuo di baseline (diplomato tra i 25 e i 35 anni residente nelle regioni centrali) il rischio stimato di uscire verso l'occupazione è sensibilmente più alto di quello di uscire verso l'inattività. La differenza tra i due rischi non tende a cambiare molto all'aumentare della durata dell'episodio. Per le donne diplomate residenti nelle regioni centrali che hanno più di 55 anni la stima del rischio di transitare verso l'inattività è invece molto superiore a quella di transitare verso l'occupazione. Inoltre questa differenza aumenta al proseguire dell'episodio: mentre la probabilità condizionata di transitare verso l'occupazione cresce poco dopo il primo anno, il rischio di uscire verso l'inattività aumenta rapidamente.

Nell'interpretare queste stime delle funzioni di rischio si devono comunque sempre considerare le assunzioni su cui si basa il modello: funzione di rischio di base monotona, rischi proporzionali per gli individui con diverse caratteristiche individuali e rischi competitivi indipendenti per l'uscita verso le due possibili destinazioni. Un'analisi accurata della sensibilità dei risultati a tali assunzioni è pertanto indispensabile qualora si sia interessati agli aspetti sostanziali dell'analisi del mercato del lavoro.

3.4 Utilizzo di altre informazioni per la determinazione della durata

Fino ad ora si è utilizzato un modello che tiene conto dell'impossibilità di determinare univocamente la durata dell'episodio di disoccupazione.

Si potrebbe però pensare di utilizzare i limiti dell'episodio per costruire un'unica quantità che sintetizzi l'informazione raccolta sulla durata della disoccupazione di ogni individuo. Ad esempio si può considerare la lunghezza effettiva dello spell come la media tra la durata minima e massima:

$$t_{med_i} = \frac{(t_{min_i} + t_{max_i})}{2}.$$

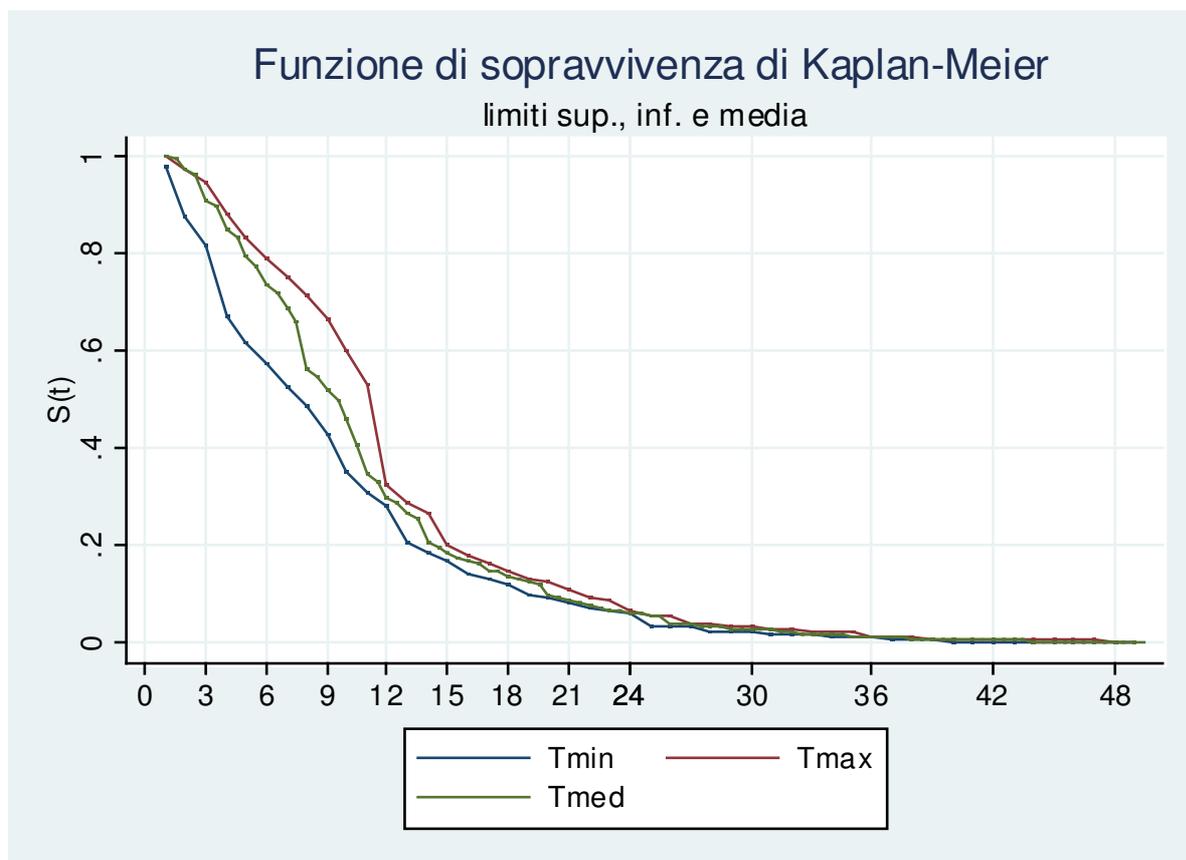


Figura 3.4: funzioni di sopravvivenza di Kaplan-Meier stimate su t_{min} , t_{max} e t_{med_i} . Non è rappresentato il valore della funzione quando l'episodio supera i cinquanta mesi.

Come atteso la funzione di sopravvivenza empirica costruita su t_{med_i} è compresa tra le funzioni di sopravvivenza calcolate sul limite inferiore e superiore degli episodi di disoccupazione (Figura 3.4)

Per verificare se l'utilizzo di questa sintesi delle informazioni disponibili porta a una differente analisi della dipendenza dalla durata si va a stimare un modello di tipo Weibull che assume t_{med_i} la durata effettivamente misurata dell'episodio di disoccupazione.

Si andrà quindi a massimizzare la seguente funzione di verosimiglianza standard:

$$\prod_{i=1}^n \exp \left[\left(\tau_i^\alpha - t_{med_i}^\alpha \right) e^{\frac{x_i \beta}{t_{med_i}}} \right] \left(\alpha t_{med_i}^{\alpha-1} e^{\frac{x_i \beta}{t_{med_i}}} \right)^{c_i} .$$

		N° CAMPIONE = 8781		
VARIABILI ESPLICATIVE (β)		COEFFICIENTE	ERRORE STD.	P-VALUE
AREA RESIDENZA	NORD OVEST	-.028	.053	.601
	NORD EST	.098	.058	.090
	SUD	-.274	.045	.000
	ISOLE	-.227	.050	.000
ETA'	15-25	.103	.037	.005
	36-45	.056	.040	.164
	46-55	.050	.050	.313
	56-75	.079	.073	.281
LIVELLO ISTRUZIONE	ELEMENTARE	.062	.051	.220
	MEDIA	-.042	.033	.214
	DIPLO. 3 ANNI	.103	.065	.111
	LAUREA	.023	.052	.659
SESSO	DONNA	.059	.029	.041
	COSTANTE	-3.380	.063	.000
PARAMETRO DI FORMA (α)				
	Ln (α)	.297	.008	.000

Tabella 3.6: Stime di massima verosimiglianza del modello Weibull che utilizza t_{med} come durata degli episodi

Le stime di massima verosimiglianza ottenute utilizzando la media delle durate come proxy per la lunghezza degli episodi (Tabella 3.6) sono molto simili a quelle ottenute nel modello in cui si utilizza tutta l'informazione disponibile. La dipendenza dalla durata, in particolare, non è significativamente diversa da quella ottenuta in precedenza (in questo caso il parametro di forma è pari a 1.346, nel modello che considera i due limiti è 1.34).

Questo risultato non è sorprendente in quanto le durate minime e massime degli episodi di disoccupazione coincidono per il 57% degli individui e la loro differenza è superiore ai tre mesi solo nel 15% dei casi.

La differenza nella stima della dipendenza dalla durata sarebbe stata probabilmente più marcata se fosse stato possibile implementare un modello basato sull'esponenziale a tratti, in cui si sarebbero potuti confrontare nel dettaglio i primi mesi di disoccupazione dove la

diversità nel determinare la lunghezza dell'intervallo risulta più evidente, come si può notare dalla Figura 3.4.

I due modelli che si sono confrontati utilizzano, anche se in modo diverso, tutte le informazioni che si è riusciti a ricostruire dai dati a disposizione. Può essere interessante verificare se le stime rimangono invariate anche utilizzando solo parte delle informazioni a disposizione. Ad esempio si può utilizzare il limite superiore come unica durata del periodo di disoccupazione, con un'ipotesi di tipo conservativo che consideri l'individuo uscito dalla disoccupazione solo nel mese in cui è certamente transitato. Le stime ottenute dal modello sarebbero quelle riassunte dalla Tabella 3.7.

VARIABILI ESPLICATIVE (β)		COEFFICIENTE	N° CAMPIONE = 8781	
			ERRORE STD.	P-VALUE
AREA RESIDENZA	NORD OVEST	-.013	.053	.808
	NORD EST	.105	.058	.068
	SUD	-.274	.045	.000
	ISOLE	-.242	.050	.000
ETA'	15-25	.106	.037	.004
	36-45	.052	.040	.194
	46-55	.021	.050	.669
	56-75	.016	.073	.831
LIVELLO ISTRUZIONE	ELEMENTARE	.047	.051	.354
	MEDIA	-.040	.033	.227
	DIPLO. 3 ANNI	.104	.065	.108
	LAUREA	.038	.052	.474
SESSO	DONNA	.033	.029	.259
	COSTANTE	-3.680	.064	.000
PARAMETRO DI FORMA (α)				
	Ln (α)	.333	.008	.000

Tabella 3.7: Stime di massima verosimiglianza del modello Weibull che utilizza solo t_{\max} .

In questo caso (Tabella 3.7) si ha una dipendenza dalla durata significativamente superiore a quella registrata nei modelli in cui si utilizzava tutta l'informazione disponibile: α è pari a 1.39.

Si può notare inoltre come risultino significativi solo l'area di residenza e l'età, mentre non è più significativo il sesso. L'utilizzare solo una parte dell'informazione porta quindi a stimare in modo significativamente diverso la dipendenza della durata e l'effetto di alcune caratteristiche individuali.

Utilizzare in modo corretto i dati retrospettivi nella ricostruzione delle durate degli spell è quindi fondamentale per determinare in modo corretto la relazione tra periodo di disoccupazione e probabilità di transitare verso un'altra condizione lavorativa.

3.4.1 Informazioni aggiuntive: dati ricostruiti

Nei dati messi a disposizione dall'Istat, oltre alle risposte al questionario sono presenti delle quantità ricostruite successivamente al momento in cui si svolge la rilevazione. La maggior parte di questi dati non aggiunge ulteriori informazioni alle risposte date alle domande del questionario, in quanto non sono altro che una loro rielaborazione.

Il dato ricostruito che riguarda la durata della disoccupazione dà, però, un contributo informativo che non è possibile ottenere dalle risposte al questionario, in quanto è il minimo tra la durata della ricerca di lavoro e la distanza effettiva dall'ultimo mese in cui l'intervistato ha lavorato. Questa quantità è ricostruita tenendo conto del mese effettivo in cui è stata svolta l'intervista, un'informazione che non è disponibile nel nostro dataset.

Si può quindi utilizzare la durata di disoccupazione fornita dall'Istat come unico t_0 (che indicheremo con t_{0ist}) nella prima intervista in cui l'individuo risulta disoccupato e risolvere parte dell'incertezza sulla durata degli episodi di disoccupazione. La ricostruzione degli spell avviene con modalità quasi identiche a quelle utilizzate in precedenza.

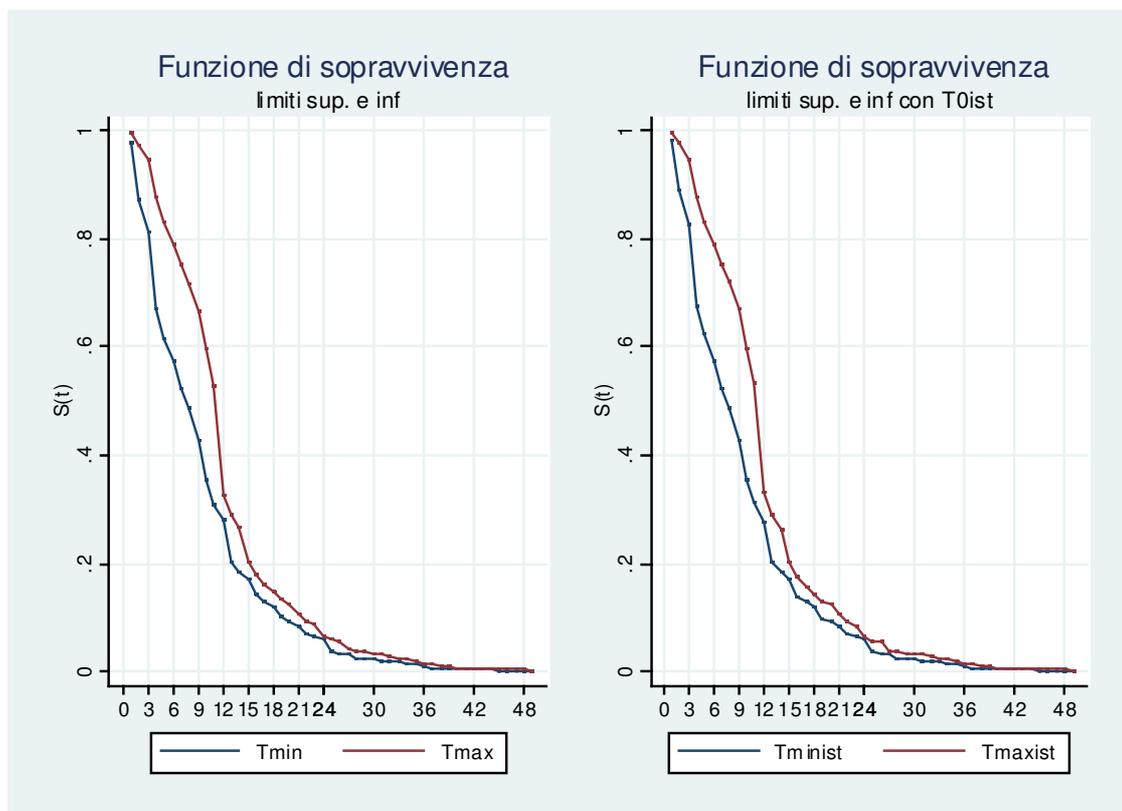


Figura 3.5: funzioni di sopravvivenza di Kaplan-Meier stimate su t_{\min} , t_{\max} e su $t_{\min ist}$, $t_{\max ist}$. Non è rappresentato il valore della funzione quando l'episodio supera i cinquanta mesi.

Le funzioni di sopravvivenza empiriche calcolate sui limiti della durata dello spell di disoccupazione ottenuti utilizzando t_{0ist} (Figura 3.5) sono molto simili a quelle ottenute senza questa informazione aggiuntiva. Il fatto che le funzioni di sopravvivenza siano quasi identiche non sorprende, dal momento che i limiti degli episodi ricostruiti nelle due diverse modalità sono identici nel 83.3% delle occasioni. Inoltre il τ_{ist} , ovvero il t_{0ist} per le osservazioni censurate a sinistra, coincide con il τ calcolato in precedenza nel 91.5% dei casi.

VARIABILI ESPLICATIVE (β)		COEFFICIENTE	N° CAMPIONE = 8781	
			ERRORE STD.	P-VALUE
AREA RESIDENZA	NORD OVEST	-.036	.054	.507
	NORD EST	.087	.059	.137
	SUD	-.282	.045	.000
	ISOLE	-.229	.051	.000
ETA'	15-25	.104	.037	.005
	36-45	.060	.040	.140
	46-55	.061	.050	.228
	56-75	.102	.074	.167
LIVELLO ISTRUZIONE	ELEMENTARE	.060	.051	.244
	MEDIA	-.046	.034	.170
	DIPLO. 3 ANNI	.098	.065	.133
	LAUREA	.019	.053	.725
SESSO	DONNA	.064	.029	.028
	COSTANTE	-3.365	.064	.000
PARAMETRO DI FORMA (α)				
	Ln (α)	.297	.008	.000

Tabella 3.8: Stime di massima verosimiglianza del modello Weibull che utilizza $t_{\min ist}$ e $t_{\max ist}$

Le stime ottenute dal modello su $t_{\min ist}$ e $t_{\max ist}$, come si può vedere nella Tabella 3.8, sono quasi identiche a quelle ottenute senza utilizzare la durata della disoccupazione ricostruita: il parametro di forma è di 1.346, identico a quello ottenuto considerando la media dei limiti come durata unica dell'intervallo e poco superiore a 1.34 che è quello stimato nel primo modello.

Quindi l'utilizzo di questa informazione aggiuntiva non influenza significativamente le stime sulla dipendenza dalla durata del rischio di uscire dalla disoccupazione, almeno nel modello che è stato adottato per l'analisi. Ciò può dare una parziale conferma sulla bontà del metodo scelto per ricostruire la lunghezza degli episodi di disoccupazione.

3.5 Possibili sviluppi

Lo studio ha possibili sviluppi sia metodologici sia dal punto di vista analitico.

Una volta implementato dal punto di vista informatico il modello basato sull'esponenziale a tratti si potrebbero confrontare nuovamente i risultati delle stime sulla dipendenza dalla durata tra i diversi modi di utilizzare le informazioni disponibili ottenendo, forse, risultati più robusti. Nonostante la dipendenza dalla durata, considerando tutto l'arco temporale, non abbia cambiamenti significativi - come è stato dimostrato in questo studio - il suo effetto potrebbe essere diverso, a seconda delle informazioni considerate, in determinati intervalli di tempo.

Il fatto che la dipendenza dalla durata sia positiva anche nel caso in cui si valuti il rischio di uscire verso l'occupazione è abbastanza sorprendente: risultati di studi svolti in passato - si veda Torelli (1987) per un'analisi svolta sulla Ricerca trimestrale delle Forze Lavoro - avevano mostrato una dipendenza negativa dalla durata. Questi studi però riguardavano solitamente una determinata fascia di popolazione e consideravano la durata della ricerca di lavoro e non una proxy della durata della disoccupazione. Potrebbe quindi essere opportuno svolgere l'analisi solo su individui con determinate caratteristiche per effettuare un confronto. Inoltre si potrebbe verificare la presenza di eterogeneità non osservata che tuttavia non cambierebbe la dipendenza positiva dalla durata: la presenza di eterogeneità non osservata porta a sottostimare il vero rischio di transitare (Heckman e Singer 1984).

Per quanto riguarda l'analisi della durata della disoccupazione e i suoi effetti specifici sulle dinamiche del mercato del lavoro, al di là degli scopi di questa tesi, sarebbe utile avere maggiori informazioni per poter analizzare in modo più dettagliato la relazione tra la durata della disoccupazione e, in particolare, la probabilità di trovare un'occupazione.

Bibliografia

- Bound J., Brown C., Marthiowetz N. (2001). Measurement error in survey data. In Heckman J. e Leamer D. (a cura di), *Handbook of Econometrics volume 5*, 59. Amsterdam, North Holland
- Di Pietro E. (1993). *La nuova indagine ISTAT sulle forze di lavoro. Economia e Lavoro*.
- Heckman, J.J. e Singer B. (1984). Econometric duration analysis. In *Journal of Econometrics*, 24, 63-132.
- Istat (2002). Le matrici di transizione della rilevazione trimestrale sulle forze lavoro .Nota metodologica. *Collana approfondimenti*, Roma.
- Istat (2004), *La Nuova Rilevazione sulle Forze Lavoro. Contenuti, metodologie, organizzazione*. Roma
- Jenkins S.P. (2005), *Survival Analysis*.
- Kiefer, N.M. (1988). Economic duration data and hazard function. In *Journal of Economic Literature*, 26, 646-679.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge, University Press.
- Loader C. (1999), *Local Regression and Likelihood*. Springer. New York,
- Paggiaro A. e Torelli N. (2002). Una procedura generalizzata per l'abbinamento esatto di record. In Camillo F. e Tassinari G. (A cura di), *Data mining, web mining e CRM, metodologie, soluzioni e prospettive*. Franco Angeli, Milano. 26-38.
- Torelli N. (1990) L'Analisi dei dati di durata nelle scienze sociali: riflessioni metodologiche e una applicazione alla durata della disoccupazione. In *Statistica Applicata*, Vol.2, n.1. 7-25.
- Torelli, N. e Trivellato, U. (1993). Data inaccuracies and sampling plan in a model of unemployment duration. In Bunzel, H, Jensen, P. e Westergard-Nielsen, N. (a cura di), *Panel Data and Labor Market Dynamics*.. North-Holland.
- Trivellato U., Marliani G. e Torelli N. (1989), Longitudinal analysis of unemployment duration from a household survey with rotating sample: a case study with Italian labor force data. In *Bureau of the Census, Fifth Annual Research Conference. March 19-22, 1989*. 408-427.
- Wooldridge J.M. (2000). *Introductory econometrics: A modern Approach*. South-Western College Publishing.17.