

Lavoro e disoccupazione: questioni di misura e di analisi

Progetto di ricerca cofinanziato dal Ministero per l'Università
e la Ricerca Scientifica e Tecnologica - Assegnazione: 1998
Coordinatore: Ugo Trivellato

**Un modello di mistura per
l'analisi della disoccupazione di
lunga durata**

Adriano Paggiaro

Dip. di Scienze Statistiche, Univ. di Padova

Working Paper n. 12

maggio 1999

Unità locali del progetto:

Dip. di Economia Politica, Univ. Di Modena

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

Dip. Di Statistica, Univ "Ca' Foscari" di Venezia

Dip. di Metodi Quantitativi, Univ. di Siena

Dip. di Scienze Statistiche, Univ. di Padova

(coord. Michele Lalla)

(coord. Bruno Contini)

(coord. Tommaso Di Fonzo)

(coord. Achille Lemmi)

(coord. Ugo Trivellato)

Dip. di Scienze Statistiche
via S. Francesco 33, 35121 Padova

1. Introduzione

Il sensibile incremento nel livello della disoccupazione, osservato negli anni recenti nei paesi dell'Unione Europea, è stato accompagnato da profondi mutamenti nelle caratteristiche e nella distribuzione della disoccupazione stessa. In particolare, si osserva una marcata tendenza all'aumento della quota di coloro che sperimentano una lunga durata dell'episodio di disoccupazione, e ciò sembra riguardare principalmente specifici settori della popolazione, per i quali il prolungarsi del periodo di ricerca di un'occupazione può essere associato a situazioni di grave disagio sociale.

Una testimonianza dell'interesse per tali mutamenti si riscontra nella tendenza ad affiancare agli usuali indicatori della situazione del mercato del lavoro, ed in particolare al tasso di disoccupazione, la quota di *disoccupati di lunga durata*, definiti convenzionalmente come individui che sperimentano episodi superiori ai 12 mesi di ricerca ininterrotta di un'occupazione.

Il mutamento nella distribuzione delle durate assume particolare rilevanza alla luce del fatto che, come è noto, ad uno stesso livello del tasso di disoccupazione possono corrispondere scenari marcatamente differenti: ad un mercato con la presenza di elevata mobilità, in cui ogni individuo sperimenta brevi episodi di disoccupazione, se ne può contrapporre uno in cui un gruppo della popolazione è costretto a sopportare gran parte del peso della disoccupazione totale. Di conseguenza, l'aumento osservato della quota di disoccupati che sperimentano lunghe durate di disoccupazione può comportare maggiori costi sociali ed economici, sia dal punto di vista della collettività che del singolo individuo.

Ulteriori implicazioni si hanno per quanto riguarda alcune letture della disoccupazione fornite dalla teoria economica. Se è ragionevole ritenere che la disoccupazione di lunga durata sia in gran parte involontaria, ne discende che, verosimilmente, una parte rilevante di disoccupati è relegata al margine del mercato del lavoro, con notevoli barriere all'entrata. Tale assunzione è avvalorata dai dati, che evidenziano come all'aumento delle durate della disoccupazione sia spesso associata una diminuzione della probabilità di transizione all'occupazione.

In questa situazione, possono venire a mancare le basi per le argomentazioni classiche sulla relazione fra disoccupazione ed inflazione, e ciò può indurre a cambiare le prospettive per le principali politiche sul mercato del lavoro. Ad interventi generalizzati mirati a ridurre l'incidenza della disoccupazione si possono contrapporre politiche attive, mirate agli individui più svantaggiati, che in questo contesto possono non avere gli usuali effetti inflazionistici a livello aggregato.

Questo lavoro si propone di individuare una classe di modelli adeguati alle mutate esigenze di analisi della durata della disoccupazione. L'idea di fondo è l'utilizzo di *modelli di mistura* che tengano in esplicita considerazione la presenza di eventuali sottogruppi della popolazione con differenti distribuzioni della disoccupazione, prevedendo la presenza di un gruppo, identificabile con i cosiddetti disoccupati di lunga durata, con rischio di transizione verso un'occupazione molto basso. In particolare, i metodi utilizzati rappresentano una generalizzazione dei modelli per l'analisi di dati di durata in presenza di *long-term survivors*, dove si assume che una parte della popolazione abbia rischio nullo di transizione.

Modelli di tale natura sono stati sviluppati in ambito biometrico (Maller e Zhou, 1996); non è tuttavia immediato l'adattamento all'analisi della durata della disoccupazione per tener conto delle caratteristiche dei dati che correntemente sono disponibili. In particolare: (i) va ricercato, dove possibile, un legame con la teoria economica, in modo da trarne indicazioni utili per l'interpretazione dei risultati e per proporre conseguentemente adeguate misure di politica; (ii) occorre estendere i modelli in modo da renderne possibile la stima nelle diverse situazioni che consentono di ottenere dati per l'analisi della disoccupazione (ad es., nel caso di campionamento da stock); (iii) è necessario fornire specificazioni sufficientemente generali, che rendono agevole l'introduzione di forme di eterogeneità osservata e non osservata.

Il risultato finale consente di scomporre un modello di durata della disoccupazione in modo da ottenere congiuntamente: (a) una stima del modello di durata solamente per gli individui destinati ad una transizione verso l'occupazione entro un determinato periodo di tempo; (b) l'individuazione delle caratteristiche dell'offerta di lavoro (sesso, età, istruzione, posizione geografica, esperienze lavorative, ecc.) legate alla segmentazione della popolazione, separando tali effetti da quelli tipicamente individuati all'interno di un modello di durata per valutare il "rischio" di transizione verso un'occupazione.

Il lavoro è organizzato come segue. Nel paragrafo 2 si richiamano i principali sviluppi della teoria economica che consentono di inquadrare i fenomeni di disoccupazione persistente in un ambito generale di segmentazione del mercato. In particolare, si analizzano i possibili supporti teorici ed interpretativi ai modelli in forma ridotta proposti nel seguito, e si forniscono alcune evidenze empiriche sulla consistenza del fenomeno della disoccupazione di lunga durata.

Nel paragrafo 3 si introducono i modelli di mistura per l'analisi dei dati di durata in presenza di *long-term survivors*, e se ne specificano alcune caratteristiche appropriate per l'analisi della disoccupazione di lunga durata, mentre il paragrafo 4 descrive nel dettaglio i relativi metodi di stima. I modelli considerati prevedono la presenza di sottogruppi non direttamente osservabili, e il procedimento inferenziale è pertanto facilitato dall'utilizzo di metodi di stima in presenza di variabili latenti, con particolare riferimento all'algoritmo EM (Dempster *et al.*, 1977).

Nel paragrafo 5 si presenta un'applicazione dei modelli di mistura a dati provenienti dalla Rilevazione Trimestrale delle Forze di Lavoro (RTFL) dell'ISTAT, con gli adattamenti necessari per utilizzare i dati ottenibili dalla particolare struttura di tale indagine; in particolare, si ottengono le correzioni necessarie per un procedimento inferenziale che tenga conto della provenienza di parte del campione dallo *stock* di disoccupati all'inizio del periodo di osservazione.

Il paragrafo 6 presenta infine una serie di riflessioni sui risultati complessivi ottenuti e sulle possibilità interpretative offerte dai modelli proposti.

2. La disoccupazione di lunga durata

2.1 Alcune evidenze empiriche

Dopo aver segnalato il problema in varie occasioni negli anni precedenti, OECD (1993) presenta una dettagliata fotografia dell'allarmante situazione relativa alla crescente quota di disoccupati di lunga durata, tentando di avanzare spiegazioni e di individuare

possibili soluzioni. Da tale analisi emerge che il tasso di disoccupazione aggregato e la disoccupazione di lunga durata non sono necessariamente due facce della stessa medaglia, e nella realtà sono presenti differenti combinazioni dei due indicatori; è pertanto evidente come un'analisi attenta delle caratteristiche della disoccupazione di lunga durata possa fornire utili informazioni per determinare la dimensione dei processi di *turnover*, nonché l'eventuale presenza di un gruppo di *hard-core unemployed* ai margini del mercato del lavoro.

Per quanto riguarda la *distribuzione territoriale* del fenomeno, tra i singoli paesi appartenenti OECD si notano caratterizzazioni differenti. Innanzitutto, il fenomeno si è recentemente acuito nei paesi dove era già presente una quota elevata di disoccupati di lunga durata, come nel sud dell'Europa ed in Irlanda: in l'Italia, ad esempio, l'incidenza sul totale dei disoccupati è passata dal 51% nel 1979 al 71% nel 1990; in quest'ultimo anno, in particolare, 1.648.000 italiani erano disoccupati da più di 1 anno, pari all'8,5% della forza lavoro, una percentuale nettamente superiore all'intero tasso di disoccupazione di molti paesi OECD. Per quanto riguarda gli altri paesi europei, nel 1990 l'incidenza dei disoccupati di lunga durata era superiore al 30% ovunque, con le sole eccezioni di Svezia e Norvegia.

A conferma della tendenza alla diffusione della disoccupazione di lunga durata, è interessante osservare come questa inizi ad interessare anche paesi caratterizzati da bassi livelli di disoccupazione e da un mercato del lavoro fortemente dinamico: la Norvegia ad esempio, pur rimanendo su livelli inferiori rispetto alle medie europee, presenta un notevole incremento nel tempo dell'incidenza della disoccupazione di lunga durata, passando dal 2,9% del 1979 a quasi il 20% nel 1990. I restanti paesi OECD presentano invece incidenze vicine al 5%, mostrando come esistano ancora situazioni in cui i disoccupati sperimentano un succedersi di episodi brevi ma ripetuti.

Il problema principale in tutti i paesi, al di là dei livelli stessi, è comunque il continuo spostamento verso destra della distribuzione delle durate, che non si arresta nemmeno durante i periodi espansionistici; una volta che l'incidenza della disoccupazione di lunga durata aumenta, non appare esserci, nelle fasi di crescita economica, l'attesa pressione alla diminuzione verso i livelli del ciclo precedente, con un conseguente progressivo aumento della quota di disoccupazione considerata strutturale.

2.2 Caratteristiche individuali dei disoccupati di lunga durata

In questo contesto appare evidente l'importanza di individuare alcune caratteristiche della popolazione particolarmente legate al fenomeno della disoccupazione di lunga durata; a tale scopo, risultano utili le analisi offerte da OECD (1993) e, con particolare riferimento all'Italia, da Ricciardi (1991).

Nei diversi paesi OECD, il gruppo di disoccupati di lunga durata presenta una composizione molto variabile. Per quanto riguarda l'età, ad esempio, nella maggior parte dei paesi la classe più rappresentata è la *prime-age* (25-44), mentre i giovani presentano solitamente numerosi episodi di breve durata; differenze sostanziali si trovano però nel Sud Europa, con una presenza superiore di giovani (15-24), ed in Giappone e Svezia, con una prevalenza dei lavoratori più anziani (45-64). Altre disaggregazioni mostrano come il tasso di disoccupazione di lunga durata delle donne sia superiore a quello degli uomini in gran parte dei paesi europei, mentre livelli di

istruzione e professionalità meno elevati per i disoccupati di lunga durata sono riscontrabili in tutti i paesi OECD.

Per quanto riguarda l'Italia, Ricciardi (1991) analizza l'evolversi delle distribuzioni della durata di ricerca dichiarata nel periodo fra il 1977 ed il 1989. Le differenze principali nella distribuzione delle durate si hanno distinguendo gli individui in base alla tipologia di disoccupazione: i gruppi per i quali si osserva una notevole crescita negli anni della percentuale di durate lunghe sono le persone in cerca di prima occupazione e le altre persone in cerca di occupazione (studenti, casalinghe, pensionati ed altri); presentano invece un andamento sostanzialmente costante i disoccupati in senso stretto.

Analisi più dettagliate rivelano che un fattore spesso associato alla disoccupazione di lunga durata è la mancanza di esperienze lavorative precedenti. Tale evidenza è confermata dal fatto che le notevoli differenze osservate nella distribuzione per età scompaiono se si tiene conto dell'esperienza, mostrando "la netta preferenza delle imprese verso individui che abbiano superato la fase di primo ingresso nel mondo del lavoro", indipendentemente dall'età di questi: "i giovani, infatti, risultano maggiormente colpiti dalla disoccupazione di lunga durata solo quando sono privi di precedenti esperienze lavorative".

Ricciardi documenta inoltre come la maggior presenza delle donne fra i disoccupati di lunga durata sia concentrata nella categoria di altre persone in cerca di impiego, formata in gran parte da casalinghe, mentre per le altre due tipologie di disoccupati non si osservano sostanziali differenze per sesso. Concludendo con l'istruzione, si osserva negli anni un'inversione nella distribuzione delle durate, particolarmente evidente nel caso di individui in cerca di prima occupazione; gli individui con istruzione più elevata, avendo una dinamica di crescita delle durate notevolmente inferiore, passano dalla categoria con durate maggiori a quella con episodi più brevi.

2.3 Alcuni riferimenti teorici

La crescente incidenza della disoccupazione di lunga durata osservata in numerosi mercati del lavoro dei paesi industrializzati indebolisce le basi empiriche su cui si basano (a) le teorie legate ad un normale *turnover* ed alla natura volontaria della disoccupazione e (b) le relazioni macroeconomiche che prevedono una diminuzione della disoccupazione nei periodi di ripresa dell'economia a livello aggregato, con un assestamento di lungo termine sul tasso di disoccupazione strutturale.

Una teoria che permette di spiegare perché, in realtà, i previsti meccanismi di aggiustamento non funzionino e la disoccupazione sia persistente è legata ai modelli *insider-outsider* (Lindbeck e Snower, 1986). Gli *insiders* sono lavoratori che dall'interno del mercato possono sfruttare la loro posizione per erigere delle barriere all'entrata di coloro che sono fuori dal mercato, gli *outsiders*. Nel caso più semplice, ad esempio, tali barriere possono derivare dagli elevati costi di assunzione e licenziamento, dovuti in alcuni paesi (fra i quali l'Italia) anche alla severa legislazione a protezione degli occupati. Modelli più complessi (Lindbeck e Snower, 1988) tengono inoltre esplicitamente in considerazione le possibili ritorsioni degli *insiders* nei confronti di quegli *outsiders* che dovessero essere assunti per aver accettato salari inferiori, tradendo una "norma" sociale non scritta; gli effetti sarebbero ovviamente negativi per gli individui assunti, ma ci sarebbero ripercussioni anche sulla produttività complessiva,

per cui la possibilità di un'assunzione con salari inferiori è ridotta sia dal punto di vista dell'offerta che della domanda. Gli effetti di un modello *insider-outsider* sui livelli aggregati sono evidenti, poiché gli *outsiders* non hanno alcuna possibilità di effettuare pressione sul sistema dei salari, e gli *insiders* possono sfruttare i costi del *turnover* per mantenere i propri salari al di sopra di quelli di equilibrio in un mercato perfettamente concorrenziale. Di conseguenza, vengono a mancare le basi per l'attuarsi del meccanismo che in un sistema in ripresa lega le spinte inflazionistiche ad una diminuzione della disoccupazione, poiché le nuove risorse create si traducono in aumenti salariali piuttosto che in creazione di nuovi posti di lavoro.

Clark e Summers (1979) presentano evidenze empiriche per gli Stati Uniti che supportano l'interpretazione secondo la quale gran parte del peso della disoccupazione totale è sopportata da un ridotto numero di persone e, di conseguenza, solo una minima parte del problema può essere spiegata da un normale *turnover*. Gli effetti politici e sociali di una distribuzione fortemente ineguale della disoccupazione sono ovviamente notevoli, con elevati costi da sopportare per un gruppo di *hard-core unemployed* che si trovano al margine del mercato del lavoro e, spesso, in situazioni di estremo disagio. Di conseguenza, i tipi di intervento più efficaci sarebbero quelli mirati ai sottogruppi svantaggiati, con la creazione di nuovi posti di lavoro dedicati e la predisposizione di piani volti ad aumentare, ad esempio attraverso la formazione, le competenze specifiche degli individui più a rischio.

In questo contesto, si può assumere che una disoccupazione di lunga durata di tipo strutturale sia strettamente legata alla categoria degli *outsiders*, per la ridotta competitività e la presenza di numerose barriere all'entrata; la conseguenza principale è la mancanza di un legame definito fra il ciclo economico, regolato essenzialmente dagli *insiders*, e la disoccupazione di lunga durata.

3. Specificazione di un modello di mistura

3.1 L'analisi della durata della disoccupazione in una popolazione eterogenea

Si supponga di voler analizzare per un singolo individuo la probabilità di transitare dallo stato di disoccupato a quello di occupato, distinguendo gli effetti su tale *rischio di transizione* di differenti fattori quali (a) la durata T dell'episodio di disoccupazione in corso e (b) alcune variabili esplicative X . Un modo tradizionale di stimare tali effetti prevede l'utilizzo di modelli di durata (Cox e Oakes, 1984, Lancaster, 1990), specificati attraverso la *funzione di rischio* $\lambda(t) = f(t)/S(t)$, che rappresenta la probabilità istantanea di transizione condizionata alla sopravvivenza nello stato fino al tempo t . Una tipica assunzione prevede che le variabili esplicative agiscano moltiplicativamente su una *funzione di rischio di base* $\lambda_B(t)$ (*modello con rischi proporzionali*):

$$\lambda(t, x) = \lambda_B(t)\varphi(x).$$

La forma di $\lambda_B(t)$ assume particolare importanza se si è interessati alla *dipendenza dalla durata*, variazione del rischio di transizione per effetto del tempo trascorso nello stato iniziale. La stima della funzione $\varphi(x)$, specificata solitamente in forma pienamente

parametrica, fornisce invece gli effetti delle singole variabili esplicative come fattori di scala sulla funzione di rischio, la cui forma è definita esclusivamente da $\lambda_B(t)$.

In presenza di *eterogeneità non osservata*, la specificazione di un modello di durata deve tener conto delle distorsioni che ne possono derivare, con particolare riferimento ad una dipendenza dalla durata negativa “spuria” dovuta al progressivo aumento nel campione del peso degli individui con rischio inferiore. Tale effetto è evidente nel caso si assuma la presenza di sottopopolazioni con comportamenti differenti sul mercato del lavoro; ad esempio, quale conseguenza della teoria *insider-outsider* è ragionevole ritenere che un insieme di disoccupati al margine del mercato del lavoro abbia probabilità molto basse di transizione verso un’occupazione.

In questa situazione, l’utilizzo dei metodi standard con l’assunzione di rischi proporzionali (si veda, ad esempio, Han e Hausman, 1990) comporta una specificazione errata e, in particolare, appare restrittiva l’assunzione che gli effetti delle esplicative rimangano i medesimi nel corso dell’episodio. Risultano invece di estremo interesse specificazioni, derivate dai *modelli di mistura* (Titterington *et al.*, 1985), che costituiscono una generalizzazione dei modelli di durata con *long-term survivors* (Maller e Zhou, 1996), nei quali si assume la presenza di un sottogruppo della popolazione con possibilità nulle, o molto basse, di transizione. Esempi del possibile utilizzo di tale classe di metodi nell’ambito del mercato del lavoro si hanno in Yamaguchi (1992), con una forte evidenza di “occupazione permanente” in Giappone e, con particolare riferimento alla presenza di *hard-core unemployed* nel mercato, in Dunsmuir *et al.* (1989) e Swaim e Podgursky (1994).

3.2 Specificazione di un modello di durata in presenza di “long-term survivors”

Si definisca una variabile indicatrice latente Y (che assume valore 1 con probabilità p , 0 con probabilità $1-p$), attraverso la quale si individua l’appartenenza di ogni individuo nella popolazione ad uno specifico sottogruppo; si assuma inoltre che i due sottogruppi così definiti presentino comportamenti differenti nel mercato e di conseguenza differenti distribuzioni delle durate degli episodi di disoccupazione, $f_1(t)$ e $f_0(t)$.

Se si ammette che all’interno della popolazione esistano altre fonti di eterogeneità osservata, è possibile introdurre tali informazioni attraverso una specificazione esplicita degli effetti delle caratteristiche individuali, non solo sulla forma delle distribuzioni di durata condizionate a Y , ma anche sulla proporzione di individui con determinate caratteristiche che appartengono ai due sottogruppi latenti, e quindi su p . L’introduzione di variabili esplicative, oltre a permettere un’analisi efficiente della struttura delle durate, consente pertanto di cogliere le caratteristiche individuali che distinguono gli appartenenti ai due gruppi.

Definendo con X le variabili esplicative che entrano nella specificazione delle singole distribuzioni, e con Z quelle che influiscono su p , la distribuzione marginale osservata della variabile T diventa la *mistura* delle due distribuzioni condizionate non osservabili:

$$f(t; x, z) = p(z) f_1(t; x) + [1 - p(z)] f_0(t; x). \quad (1)$$

Con alcuni semplici passaggi è possibile ricavare anche la funzione di sopravvivenza relativa all’intero campione, che diventa una media ponderata, punto per punto, delle due funzioni di sopravvivenza condizionate:

$$S(t; x, z) = p(z)S_1(t; x) + [1 - p(z)]S_0(t; x). \quad (2)$$

Farewell (1982) osserva che, nei modelli di durata standard, si assume implicitamente che gli episodi censurati provengano tutti dalla stessa popolazione di quelli completi, e che quindi *tutti* gli episodi abbiano una durata attesa finita e siano destinati a terminare con una transizione allo stato di interesse. Per ovviare a questa restrizione, Farewell presenta un modello in cui si ammette l'esistenza di due popolazioni distinte, ed alcuni fra gli episodi censurati possono riferirsi ad individui che non saranno mai soggetti ad una transizione verso lo stato di interesse; con evidente riferimento allo specifico contesto applicativo, tali individui sono detti *immuni*. L'utilità di tale modello viene dal fatto che, se la vera distribuzione delle durate è *impropria*, e solo una parte p della popolazione è in realtà destinata ad una transizione, risulta inadeguata l'imposizione che la funzione di sopravvivenza raggiunga il limite inferiore 0. Il modello proposto da Farewell può essere ottenuto come caso particolare di (1) e (2), con le due seguenti restrizioni equivalenti:

$$\lambda_0(t; x) = 0 \Leftrightarrow S_0(t; x) = 1. \quad (3)$$

3.3 Un modello con soglia per la disoccupazione di lunga durata

Nell'ambito dell'analisi della durata della disoccupazione, appare restrittiva l'ipotesi di un gruppo di individui con probabilità nulle di transizione verso un'occupazione. Nella logica del modello *insider-outsider*, è invece possibile assumere l'esistenza di una *soglia* t_{LT} per la disoccupazione di lunga durata che permetta di discriminare al meglio le due popolazioni, con una definizione per la variabile latente Y direttamente legata alla vera durata T^* dell'episodio:

$$Y = \begin{cases} 1 & \text{se } T^* \leq t_{LT} \\ 0 & \text{se } T^* > t_{LT} \end{cases}$$

Poiché la finestra di osservazione dalla quale si ottengono i dati è spesso molto limitata, non sempre è possibile osservare T^* e, di conseguenza, il valore di Y per gli episodi censurati con durata inferiore alla soglia; si hanno invece informazioni su Y nelle altre due situazioni possibili:

$$\delta_i = 1 \Rightarrow y_i = 1 \quad (4)$$

$$LT_i = 1 \Rightarrow y_i = 0 \quad (5)$$

dove δ e LT sono due variabili indicatrici, che valgono 1 se si osserva una transizione prima della soglia o se, invece, la durata osservata è maggiore della soglia stessa.

Purché si scelga una soglia sufficientemente elevata, appare ammissibile l'assunzione $\lambda_0(t; x) = \lambda_0$, con un rischio di transizione per i disoccupati di lunga durata basso, costante nel tempo, ed indipendente dalle caratteristiche dell'individuo; queste ultime vengono pertanto ad influenzare la probabilità di diventare disoccupato di lunga durata,

ma una volta oltrepassata la soglia non hanno più alcun effetto sul rischio di transizione. L'idea è che in tale gruppo vengano a trovarsi essenzialmente gli *outsiders* del mercato, che dopo un certo periodo di ricerca diventano omogenei rispetto alle ridotte probabilità che tale ricerca vada a buon fine. Tale assunzione è supportata da Layard e Nickell (1986), i quali affermano che, essendo i disoccupati di lunga durata marginalizzati dal mercato del lavoro, una migliore misura della condizione economica generale si ottiene considerando solo la parte della disoccupazione che può effettivamente influire sulla situazione aggregata, mentre gli *outsiders* necessitano di analisi e politiche differenziate. Un modello così specificato, oltre a consentire l'analisi degli episodi di disoccupazione più interessanti, permette di individuare quali siano le caratteristiche individuali più legate alla possibilità di rimanere al margine del mercato.

Il modello di mistura proposto ha natura diversa rispetto a quello di Farewell, poiché la funzione di sopravvivenza marginale raggiunge, anche se in tempi lunghissimi, il limite 0, ed ogni episodio ha una durata attesa finita. Tale ipotesi non è comunque in contrasto con la definizione di "immuni" data da Farewell, che assume che non ci siano transizioni nel periodo di osservazione; le restrizioni (3) sono pertanto necessarie solo fino alla durata massima osservata, mentre la forma successiva della distribuzione non influenza la specificazione del modello per il periodo di riferimento. Il modello con immuni diventa in questo modo un caso particolare di quello con soglia, dove la soglia tende all'infinito o è comunque posta oltre la durata massima osservata.

Come si vedrà, una soglia finita ha l'ulteriore pregio di rendere applicabile il modello anche in presenza di dati provenienti da campionamento da *stock*, l'analisi dei quali presenta invece notevoli problemi con l'ipotesi di distribuzione impropria e la conseguente presenza di individui con episodi di durata infinita.

4 Stima del modello

4.1 La funzione di verosimiglianza

Per poter definire la verosimiglianza del modello di mistura, ne rimangono da specificare le distribuzioni coinvolte. Per la relazione fra Y , variabile risposta dicotomica, e le esplicative Z , si utilizza in particolare una regressione logistica:

$$P(Y = 1 | Z; \gamma) = p(z, \gamma) = \frac{\exp(z\gamma)}{1 + \exp(z\gamma)}. \quad (6)$$

Il modello viene completato dalla specificazione di un modello di durata per gli individui con $y_i = 1$, per il quale si assume l'ipotesi di rischi proporzionali, meno restrittiva se limitata alle durate inferiori alla soglia; la funzione di rischio condizionata che ne deriva è la seguente:

$$\lambda(t, x; \beta | y = 1) = \lambda_1(t, x; \beta) = \lambda_b(t) \exp(x\beta). \quad (7)$$

Poiché per gli episodi con $LT=1$ si è assunto un rischio costante, tenendo conto della (5) l'unico contributo alla funzione di verosimiglianza è la probabilità $1-p$ di appartenere alla popolazione con $Y=0$; per gli altri episodi il contributo è invece quello tradizionale

per i modelli di durata, con la funzione di densità (1) per gli episodi conclusi e la funzione di sopravvivenza (2) per quelli censurati, ovviamente tenendo in esplicita considerazione le restrizioni (3) imposte al modello per la popolazione con $Y=1$. Si ottiene in questo modo la seguente verosimiglianza marginale:

$$L(\beta, \gamma) = \prod_{i=1}^n \left\{ [p(\gamma) f_1(\beta)]^{\delta_i} [p(\gamma) S_1(\beta) + 1 - p(\gamma)]^{1-\delta_i} \right\}^{1-LT_i} [1 - p(\gamma)]^{LT_i} . \quad (8)$$

Si osservi che, se la soglia viene fissata ad un valore superiore alla durata massima osservata, non si osserva alcun episodio con $LT=1$ e la (8) si riduce alla verosimiglianza del modello con immuni utilizzata da Farewell (1982). Si noti inoltre che la specificazione della verosimiglianza in (8) è valida in qualunque situazione sia possibile osservare individui con $y_i=0$, se all'indicatore LT viene data una differente interpretazione; sono pertanto riconducibili a questa specificazione, ad esempio, i contributi di eventuali episodi per i quali si osservi un'uscita definitiva dalle forze lavoro.

4.2 Stima con metodi iterativi: l'algoritmo EM

La verosimiglianza del modello di mistura definita dalla (8) può risultare particolarmente complessa, ed il reperimento del massimo della stessa può risultare poco agevole; questo accade in particolare poiché occorre stimare contemporaneamente tutti i parametri, sia quelli che governano le distribuzioni di durata che quelli relativi alla mistura. Una possibile soluzione numericamente più agevole si ottiene attraverso l'algoritmo EM (Dempster *et al.*, 1977), per un'applicazione del quale si veda, tra gli altri, Kuk e Chen (1992).

La complessità dei modelli di mistura dipende essenzialmente dalla presenza della variabile latente Y , che non consente di distinguere gli individui appartenenti alle differenti popolazioni; se fosse possibile conoscere Y , la stima del modello risulterebbe invece spesso immediata. Nel caso particolare del modello con verosimiglianza marginale (8), l'introduzione di Y consente di distinguere il differente contributo delle due popolazioni, e tenendo conto della (4) e della (5) si ottiene la seguente verosimiglianza congiunta:

$$L(\beta, \gamma; y) = \prod_{i=1}^n \lambda_1(t_i, x_i; \beta)^{\delta_i} [p(z_i; \gamma) S_1(t_i, x_i; \beta)]^{y_i} [1 - p(z_i; \gamma)]^{1-y_i} . \quad (9)$$

Con alcuni semplici passaggi è possibile osservare come la (8) sia ottenibile dalla marginalizzazione della (9) rispetto a Y ; il vantaggio della nuova specificazione è che, con la presenza esplicita di Y , la massimizzazione della (9) può avvenire in maniera agevole attraverso l'algoritmo EM, procedimento iterativo che consiste in due passi: l'imputazione dei valori assunti dai dati mancanti per Y (passo E) e la massimizzazione della verosimiglianza in presenza di informazione completa (passo M).

Al passo M , condizionalmente al valore assunto da Y nell'iterazione precedente la verosimiglianza fattorizza nel modo seguente:

$$L(\beta, \gamma | y) = \prod_{i=1}^n \left[\lambda_1(t_i, x_i; \beta)^{\delta_i} S_1(t_i, x_i; \beta) \right]^{y_i} \left\{ p(z_i; \gamma)^{y_i} [1 - p(z_i; \gamma)]^{1-y_i} \right\}. \quad (10)$$

La prima parte della produttoria in (10), dalla quale proviene la stima di β , è la verosimiglianza del modello di durata con funzione di rischio (7), pesata dal valore assunto da Y ; la seconda parte, responsabile della stima di γ , è invece quella proveniente dalla regressione logistica (6). La stima dei due vettori di parametri avviene pertanto in modo indipendente se si “aumenta” l’informazione disponibile imputando il valore di Y per gli episodi dove questo rimanga latente.

L’imputazione avviene al *passo E* dove, mantenendo valide la (4) per gli episodi conclusi e la (5) per quelli con durata superiore alla soglia, per gli episodi rimanenti si sostituisce ad Y il suo valore atteso, calcolato condizionatamente ai valori stimati precedentemente per i parametri β e γ . La verosimiglianza di interesse per y_i ($\delta_i = 0, LT_i = 0$) risulta pertanto la seguente:

$$L(y_i | \beta, \gamma) = [p(z_i; \gamma) S_1(t_i, x_i; \beta)]^{y_i} [1 - p(z_i; \gamma)]^{1-y_i}. \quad (11)$$

Il valore atteso di Y è facilmente calcolabile se si osserva che la (11) rappresenta la verosimiglianza di una distribuzione Bernoulliana con la seguente media:

$$E(y_i | \beta, \gamma) = \frac{p(z_i; \gamma) S_1(t_i, x_i; \beta)}{p(z_i; \gamma) S_1(t_i, x_i; \beta) + 1 - p(z_i; \gamma)}. \quad (12)$$

Questo risultato consente, oltre ad una semplificazione nella stima, anche un’interpretazione più esplicita del funzionamento del modello. È infatti immediato osservare che, oltre al valore p proveniente dalla distribuzione logistica, si hanno informazioni su Y anche attraverso la funzione di sopravvivenza: il denominatore in (12) è infatti la funzione di sopravvivenza *marginale* definita in (2) con le restrizioni (3), mentre il numeratore rappresenta la parte di questa “imputabile” agli individui con $y_i = 1$, data dalla sopravvivenza *condizionata* riscalata per p . Si è pertanto formalizzato anche il contributo, nella definizione dei due gruppi della popolazione, degli episodi censurati con durate inferiori alla soglia, con un valore basso della funzione di sopravvivenza che può essere indice di appartenenza al sottogruppo svantaggiato.

4.3 Campionamento da stock e distorsione da lunghezza

In campo economico e sociale, analizzando dati di durata di tipo osservazionale, accade spesso che alcuni individui si trovino già nello stato di interesse nel momento in cui la rilevazione ha inizio; in particolare, tale situazione si presenta per tutti gli individui se si utilizza un campionamento da stock. Ne deriva un caratteristico problema di selezione non casuale del campione, in quanto la probabilità di entrare nel campione è direttamente proporzionale alla durata della permanenza nello stato, con una possibile sovrastima della presenza di episodi lunghi rispetto a quelli brevi (Salant, 1977). Di conseguenza, non solo la lunghezza media degli episodi campionati risulta superiore alla reale durata media nell’intera popolazione, ma viene influenzata anche la

distribuzione nel campione delle caratteristiche individuali, osservate o meno, che siano correlate con differenti durate della disoccupazione (Chesher e Lancaster, 1983). Ne deriva una distorsione da lunghezza (*length bias*), per tener conto della quale è necessaria una specifica derivazione delle distribuzioni interessate.

Salant (1977) analizza nel dettaglio le forme funzionali implicate, nell'ipotesi di flussi costanti in entrata nello stato di interesse; tale assunzione rientra nel problema più generale per cui ogni episodio in corso al momento della rilevazione, essendo iniziato in momenti diversi, può essere influenzato da differenti condizioni iniziali. Nell'analisi della durata della disoccupazione, l'ipotesi di un processo generatore delle durate invariante nel tempo è certamente restrittiva dal punto di vista della teoria economica, ma consente una trattazione semplificata dei notevoli problemi derivanti dall'utilizzo di dati da *stock*.

Una descrizione completa delle funzioni di verosimiglianza derivanti dai differenti tipi di campionamento si trova, ad esempio, in Pudney (1989). In particolare, una possibile correzione per la distorsione da lunghezza si ottiene tenendo conto esplicitamente delle medie condizionate alle variabili esplicative; in questo modo si riequilibra il contributo alla verosimiglianza delle diverse parti della popolazione, penalizzando gli individui sovrarappresentati nel campione per il fatto di possedere caratteristiche legate mediamente a durate superiori.

Nel caso della mistura di due differenti popolazioni, il valore atteso della durata dell'episodio è differente per i due gruppi: $E_1(t; x, \beta) = E(t|y=1; x, \beta)$ è infatti legato ai parametri del modello di durata, e varia pertanto all'interno dei procedimenti di stima, mentre $E_0(t) = E(t|y=0) = t_{LT} + \lambda_0^{-1}$ è costante e dipende unicamente dalla soglia t_{LT} prescelta e dal rischio λ_0 assunto per le durate superiori alla soglia.

Definite le durate attese, è necessario ricordare che il contributo del gruppo con $y_i = 0$ alla verosimiglianza marginale (8) deriva dalla (2) con l'assunzione $S_0(t|t < t_{LT}) = 1$; definendo con LB un indicatore che vale 1 quando la durata è distorta da lunghezza e 0 altrimenti, si ottiene pertanto la seguente verosimiglianza:

$$L(\beta, \gamma) = \prod_{i=1}^n \left\{ \left[\frac{p(\gamma) f_1(\beta)}{E_1(t; \beta)^{LB_i}} \right]^{\delta_i} \left[\frac{p(\gamma) S_1(\beta)}{E_1(t; \beta)^{LB_i}} + \frac{1-p(\gamma)}{E_0(t)^{LB_i}} \right]^{1-\delta_i} \right\}^{1-LT_i} [1-p(\gamma)]^{LT_i}.$$

Si noti che $E_0(t)$ ha effetto solo nel contributo alla verosimiglianza degli individui per i quali non si conosce Y , mentre per quelli con $y_i = 0$ entra in una costante moltiplicativa che non dipende dai parametri di interesse e non ne influenza la stima.

Gli effetti della correzione per distorsione da lunghezza nel procedimento di stima sono più evidenti se si introduce la variabile latente Y attraverso l'algoritmo EM. Al passo M , condizionalmente a Y la verosimiglianza fattorizza come in (10): poiché γ non è presente in nessuna delle due durate attese, la stima della regressione logistica non subisce variazioni, mentre l'effetto di $E_1(t)$ sulla stima di β è quello tipico dei modelli di durata:

$$L(\beta|y) = \prod_{i=1}^n \left[\frac{\lambda_1(t_i, x_i; \beta)^{\delta_i} S_1(t_i, x_i; \beta)}{E_1(t; x, \beta)^{LB_i}} \right]^{y_i}.$$

L'effetto di $E_0(t)$ è invece evidente al passo E dove, con alcuni semplici passaggi, la (12) diventa:

$$E(y_i|\beta, \gamma) = \frac{p(z_i; \gamma) S_1(t_i, x_i; \beta)}{p(z_i; \gamma) S_1(t_i, x_i; \beta) + [1 - p(z_i; \gamma)] [E_1(t; x, \beta) / E_0(t)]^{LB_i}}. \quad (13)$$

Si nota immediatamente che l'unica differenza fra (12) e (13) è nel peso da dare alla componente $1-p$ nell'imputazione di Y . Tale peso, per gli episodi con distorsione da lunghezza, dipende dal rapporto fra le durate attese per i due gruppi, che per le assunzioni fatte è minore di 1; per questo motivo, il valore imputato per Y risulta superiore a quello ottenuto senza correzione, e la differenza è maggiore all'aumentare della discrepanza fra le distribuzioni delle durate nei due gruppi.

L'intuizione alla base di questa correzione è semplice, poiché in un campionamento da *stock* un individuo appartenente al gruppo di *long-term survivors* ha una probabilità maggiore di entrare nel campione, ed il mancato bilanciamento dei pesi porterebbe ad una sovrastima degli individui con $y_i = 0$ nella popolazione. La correzione nella (13) serve pertanto a dare alle due popolazioni un peso inversamente proporzionale alla loro probabilità di entrare nel campione, con l'obiettivo di ristabilire il giusto equilibrio fra durate brevi e lunghe.

Si noti che, se si estremizzasse la situazione portando la soglia all'infinito o ipotizzando un rischio λ_0 nullo, la durata attesa $E_0(t)$ tenderebbe all'infinito ed una parte della popolazione avrebbe una probabilità tendente ad 1 di entrare nel campione; di conseguenza, una correzione per distorsione da lunghezza risulta impossibile nel modello con immuni, se non si è disposti ad ammettere una soglia al di là della quale vi sia un seppur minimo rischio di transizione.

Un'ulteriore importante conseguenza della (13) è che, in presenza di distorsione da lunghezza, sia il rischio λ_0 che il valore della soglia t_{LT} influiscono nel procedimento inferenziale. È pertanto necessario definire tali valori sulla base di una qualche giustificazione teorica o di considerazioni empiriche legate ad analisi di sensibilità delle stime ottenute.

5. Un'applicazione

5.1 I dati

I dati utilizzati per un'applicazione del modello proposto provengono dalla Rilevazione Trimestrale delle Forze di Lavoro (RTFL) dell'Istat, e ne sfruttando la struttura longitudinale implicita nel sistema di rotazione del campione. In particolare, a seguito di una procedura di abbinamento probabilistico di record, è possibile seguire per un periodo di 15 mesi gli individui relativi, teoricamente, fino ad un quarto del campione totale. Per quanto riguarda la condizione lavorativa, sono disponibili informazioni utili

per la ricostruzione dei percorsi lavorativi degli individui nel periodo di osservazione, oltre ad un quesito retrospettivo sulla durata dell'eventuale ricerca di un'occupazione precedente alla rilevazione; sono inoltre presenti le tipiche variabili socio-demografiche, oltre ad alcune informazioni sulle caratteristiche della ricerca del lavoro. Per ulteriori dettagli sulla generazione del campione longitudinale e sulle specifiche caratteristiche della procedura di abbinamento si veda Paggiaro (1998), mentre per una descrizione più dettagliata della RTFL e dei suoi possibili utilizzi si vedano Trivellato (1991), e Casavola e Sestito (1994) per le variazioni più recenti.

Le singole indagini utilizzate sono quelle relative al periodo da gennaio 95 ad aprile 96 per la regione Campania, con circa 18000 record per trimestre. Gli individui abbinati sull'intero periodo sono 2956, dei quali 419 presentano almeno un episodio di disoccupazione osservato nel periodo. Tralasciando i pochi episodi multipli osservati, si considerano esclusivamente i primi episodi osservati per ogni individuo, le cui caratteristiche principali sono raccolte nella tabella 1.

Per quanto riguarda la distribuzione delle durate, si noti che la struttura della rilevazione consente di osservare solo in alcuni casi un episodio completo misurato puntualmente, mentre le durate degli episodi sono in gran parte rilevate all'interno di intervalli. L'analisi della distribuzione delle durate minime (estremi inferiori degli intervalli osservati) permette comunque di osservare che il 50% degli episodi superano certamente la definizione ufficiale di disoccupazione di lunga durata; se si considerano inoltre altre informazioni relative agli intervalli (non riportate nella tabella), ne deriva che tale percentuale è certamente una sottostima di quella reale.

Nella tabella 2 si riporta la distribuzione nel campione delle variabili che potenzialmente possono essere inserite nel modello:

- *età*, suddivisa in classi rappresentate da variabili dummy, 15-24 come classe di base;
- *sex*, variabile dicotomica, 1 se maschio;
- *relazione con il capofamiglia*, 1 se capofamiglia;
- *stato civile*, 1 se non sposato;
- *istruzione*, 1 se possiede almeno un diploma che permette l'accesso all'università;
- *esperienze lavorative*, 1 se da un qualunque quesito risulta un precedente lavorativo;
- *indennità di disoccupazione*, 1 se viene percepita indennità o sussidio.

Tabella 1: Descrizione degli episodi di disoccupazione iniziali nel campione per la regione Campania, gennaio 1995 - aprile 1996

Num.	Campion.		Transizione verso			Durata minima dell'episodio			
	Flusso	Stock	O	NFL	Cens.	0-6	6-12	12-24	24+
419	192	227	64	198	157	146	64	66	143

Tabella 2: Valore medio delle variabili esplicative nel campione

Età	Sex	Rel.cap.	St.civ.	Istruz.	Esp.lav.	Ind.dis.
29,78	0,51	0,17	0,65	0,37	0,34	0,07

5.2 Specificazione del modello

Alcune stime preliminari mostrano per la funzione di rischio di base, sia nel modello di durata semplice che nel modello di mistura, una chiara forma non monotona, per cui non appare opportuno l'utilizzo della distribuzione Weibull, molto diffusa nelle applicazioni economiche. Il modello viene pertanto specificato in forma semiparametrica, con una funzione di rischio di base esponenziale a tratti, assumendo un rischio costante all'interno di predefiniti intervalli di tempo; le classi prescelte sono relative a 6, 12 e 24 mesi, sia per un'opportuna distribuzione degli episodi che per considerazioni di carattere teorico ed interpretativo.

La funzione di verosimiglianza utilizzata è la (8), con alcune modifiche relative all'introduzione esplicita della presenza di episodi non osservati puntualmente, con la transizione o la censura registrate in un intervallo (t_{\min}, t_{\max}) . Per quanto riguarda gli episodi censurati, il contributo è dato come sempre dalla funzione di sopravvivenza calcolata nella durata minima osservata dell'episodio, $S(t_{\min})$; il contributo degli episodi completi è invece la probabilità che un episodio si concluda all'interno dell'intervallo osservato, ottenibile semplicemente come $S(t_{\min}) - S(t_{\max})$. La distinzione fra i due tipi di osservazione non altera comunque la struttura di base della verosimiglianza (8), con la sola attenzione nel considerare rispettivamente un rischio istantaneo o intervallare.

Per la definizione delle due popolazioni, non essendoci per il momento una qualche giustificazione teorica, la soglia viene posta a 60 mesi: È stata comunque effettuata un'analisi di sensibilità delle stime, facendo muovere la soglia da un minimo di 36 mesi ad un massimo di 98; tale valore è dovuto all'arrotondamento a due cifre effettuato nel questionario della RTFL, per cui il valore massimo osservabile è 99 mesi ed oltre a questo si ricadrebbe nell'ambito del modello con immuni. Al variare della soglia, si osservano alcune variazioni nella significatività dei parametri, ma mai nei segni; in particolare, come ci si poteva attendere, al diminuire della soglia tende ad aumentare il rischio di transizione per le classi di durata superiori, in quanto gli individui con $Y=1$ hanno meno tempo utile per la transizione.

5.3 Stima del modello

I risultati della stima del modello di mistura sui dati relativi alla Campania sono riportati nella tabella 3. L'aspetto più evidente è che tutti gli effetti osservati sul modello standard di durata hanno un corrispettivo diretto nella regressione logistica, se si eccettua quello poco significativo relativo all'indennità di disoccupazione, mentre l'unico effetto che rimane anche sulle durate è quello dell'esperienza lavorativa.

La presenza di eterogeneità nella popolazione appare evidente, con una maggiore probabilità di appartenere al gruppo svantaggiato legata ad alcune specifiche caratteristiche come un ridotto livello di esperienza lavorativa ed istruzione, il fatto di non essere sposato, l'appartenere ad una fascia di età intermedia.

Guardando nel particolare alcuni fra gli effetti più significativi, si osserva come in molti casi trovino conferma le esperienze empiriche e le previsioni della teoria economica. Un chiaro esempio si ha per l'esperienza lavorativa, che ha un effetto positivo sia sulla probabilità di appartenere al gruppo avvantaggiato che, all'interno di questo, sulla

durata del periodo di ricerca. Inoltre, confermando i risultati empirici di Ricciardi (1991), si può osservare che, una volta tenuto conto dell'esperienza, relativamente all'età si ha un effetto positivo sui più giovani; potrebbe pertanto valere anche per l'Italia l'evidenza per cui la disoccupazione di lunga durata, *ceteris paribus*, pare pesare soprattutto sulle età intermedie. L'elevata incidenza della disoccupazione di lunga durata osservata per i giovani della Campania pare dunque spiegabile in gran parte attraverso le ridotte esperienze lavorative degli individui in età inferiore, confermando la presenza di forti barriere all'entrata. Una possibile alternativa all'esperienza lavorativa appare comunque l'istruzione, che presenta un effetto positivo sulla probabilità di appartenere al gruppo meno svantaggiato. Effetti più incerti, e pertanto difficilmente interpretabili, si hanno invece per le restanti variabili esplicative.

Tabella 3: Stima dei parametri del modello con soglia a 60 mesi per la regione Campania (errori standard in parentesi). Parametri significativamente non nulli al livello del 10% (), 5%(**) e 1%(***)*.

	Modello	Modello di mistura	
	Di durata	Durate	Logistica
Durata 6-12	1,26 (0,41)***	1,25 (0,39)***	
Durata 12-24	0,84 (0,41)**	0,93 (0,41)**	
Durata 24+	0,67 (0,34)**	1,64 (0,36)***	
Intercetta	-4,02 (0,37)***	-4,26 (0,49)***	2,88 (0,55)***
Età 25-34	-0,53 (0,16)***	0,09 (0,31)	-1,63 (0,38)***
Età 35-44	-0,72 (0,19)***	-0,22 (0,49)	-2,62 (0,54)***
Età 45+	-0,23 (0,24)	0,07 (0,62)	-1,05 (0,72)*
Sesso M	-0,17 (0,14)	-0,06 (0,23)	-0,40 (0,35)
Capofamiglia	-0,14 (0,22)	-0,22 (0,38)	-0,64 (0,59)
Non sposato	-0,37 (0,17)**	-0,21 (0,32)	-1,36 (0,47)***
Istr. Elevata	0,18 (0,13)*	-0,19 (0,20)	0,75 (0,33)**
Esp. lavor.	0,95 (0,16)***	1,02 (0,26)***	1,97 (0,46)***
Ind. disocc.	0,34 (0,21)*	-0,06 (0,30)	0,15 (0,60)

Pur essendo un risultato presente solo nelle stime distorte per lunghezza, peraltro qui non presentate, appare interessante una possibile interpretazione di eventuali effetti delle esplicative con segno diverso sulle due parti del modello, che possono portare a risultati fuorvianti se l'analisi avviene con i modelli di durata standard. Si pensi ad esempio all'istruzione, con un effetto positivo sulla probabilità di appartenere al gruppo avvantaggiato ma negativo sulle durate. La stima di un modello standard potrebbe portare, ad esempio, alla conclusione che un'istruzione più elevata tenda ad accelerare la transizione verso un'occupazione, mentre in realtà tale effetto è dovuto alla combinazione di due fattori opposti: l'istruzione è fortemente correlata in senso positivo con l'appartenenza al gruppo avvantaggiato, ma condizionatamente a tale vantaggio l'individuo si può permettere una ricerca più selettiva per trovare un'occupazione adatta alle sue esigenze, in un'ottica di teoria di *job search*.

6. Conclusioni

La presenza nella popolazione di un gruppo di persone particolarmente svantaggiate nella ricerca di un'occupazione presenta dei problemi, oltre dal punto di vista strettamente socio-economico, anche nella specificazione di un modello idoneo all'analisi della probabilità individuale di trovare un'occupazione.

Dai risultati presentati appare evidente che la specificazione di un modello di mistura che tenga conto della segmentazione nella popolazione, anche con ipotesi fortemente semplificate, permette di cogliere vari aspetti non percepibili attraverso un modello di durata stimato con i metodi tradizionali. In particolare, la distinzione degli effetti delle variabili esplicative sull'appartenenza ai due gruppi dagli effetti sulle durate consente di conseguire due obiettivi: (a) la stima di un modello di durata per gli individui di effettivo interesse e (b) la definizione delle caratteristiche degli individui al margine del mercato del lavoro.

I risultati del procedimento inferenziale sembrano mostrare, in particolare, che gli effetti osservati sulle durate in un modello standard sono in realtà collegati alla segmentazione della popolazione, e sono dovuti ad una talvolta piccola parte del campione con durate estremamente elevate più che a reali effetti sulle durate nel breve termine. Vi è, infatti, quasi sempre una diretta corrispondenza nella regressione logistica, mentre per quanto riguarda gli effetti sul rischio condizionato i risultati sono spesso non significativi o addirittura di segno opposto. In particolare, per la Campania, solo l'*esperienza lavorativa* mostra un forte effetto dello stesso segno sia sull'appartenenza ai gruppi che sulle durate.

Le assunzioni introdotte nei modelli, in particolare per l'applicazione empirica, sono indubbiamente restrittive. È tuttavia ragionevole ritenere che non tutte siano necessarie per l'identificabilità del modello, e che alcune di esse possano essere allentate; tale verifica è lasciata al lavoro futuro. Si è inoltre mostrato come, seppure all'interno di una modellazione in forma ridotta, alcune assunzioni possano essere collegate alle teorie economiche che vedono nella segmentazione del mercato la principale causa delle nuove forme di disoccupazione osservate; si noti come tale corrispondenza con la teoria sia invece più problematica nel caso si mantenga la classica ipotesi di rischi proporzionali, la cui giustificazione risiede spesso esclusivamente nella maggiore semplicità computazionale.

Riferimenti bibliografici

- Casavola, P., Sestito, P. (1994). L'indagine ISTAT sulle forze di lavoro. *Lavoro e relazioni industriali*, 1, 179-195.
- Chesher, A., Lancaster, T.L. (1983). The estimation of models of labour market behaviour. *Review of Economic Studies*, 50, 609-624.
- Clark, K.B., Summers, L.H. (1979). Labor market dynamics and unemployment: a reconsideration. *Brookings Papers on Economic Activity*, 1, 13-60.
- Cox, D.R., Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- Dempster, A.P., Laird, N.H., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.
- Dunsmuir, W., Tweedie, R., Flack, L., Mengersen, K. (1989). Modelling of transitions between employment states for young australians. *Australian Journal of Statistics*, 31a, 165-196.
- Farewell, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38, 1041-1046.
- Han, A., Hausman, J.A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5, 1-28.
- Kuk, A.Y.C., Chen, C.H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 79, 3, 531-541.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
- Layard, N., Nickell, S. (1986). Unemployment in Britain. *Economica*, 53.
- Lindbeck, A., Snower, D. (1986). Wage setting, unemployment and insider-outsider relations. *American Economic Review, Papers and Proceedings*, 235-239.
- Lindbeck, A., Snower, D. (1988). Cooperation, harassment, and involuntary unemployment: an insider-outsider approach. *American Economic Review, Papers and Proceedings*, 167-188.
- Maller, R., Zhou, X. (1996). *Survival Analysis with Long-term Survivors*. Chichester: Wiley & Sons.
- OECD (1993). *Employment Outlook*. Paris.
- Paggiaro, A. (1998). *Modelli di mistura per l'analisi della durata della disoccupazione: aspetti metodologici e un'applicazione ai dati dell'indagine sulle forze di lavoro*. Tesi di Dottorato di Ricerca, Dipartimento di Scienze Statistiche, Università di Padova.
- Pudney, S. (1989). *Modelling Individual Choice: the Econometrics of Corners, Kinks and Holes*. Oxford: Basil Blackwell.
- Ricciardi, L. (1991). La disoccupazione di lunga durata in Italia: un'analisi dell'evidenza empirica nel periodo 1977-1989. *Economia & Lavoro*, 25, 2, 69-94.
- Salant, S.W. (1977). Search theory and duration data: a theory of sorts. *Quarterly Journal of Economics*, 91, 39-57.
- Swaim, P., Podgursky, M. (1994). Female labor supply following displacement: a split-population model of labor force participation and job search. *Journal of Labor Economics*, 12, 4, 640-656.
- Titterton, D.M., Smith, A.F.M., Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley & Sons.

- Trivellato, U. (1991) (a cura di). *Forze di lavoro: disegno dell'indagine e analisi strutturali*. Annali di Statistica, serie 9, vol. 11. Roma: Istat.
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of permanent employment in Japan. *Journal of the American Statistical Association*, 87, 284-292.

Working Papers già pubblicati

1. E. Battistin, A. Gavosto e E. Rettore, *Why do subsidized firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy*, Agosto 1998.
2. N. Rosati, E. Rettore e G. Masarotto, *A lower bound on asymptotic variance of repeated cross-sections estimators in fixed-effects models*, Agosto 1998.
3. U. Trivellato, *Il monitoraggio della povertà e della sua dinamica: questioni di misura e evidenze empiriche*, Settembre 1998.
4. F. Bassi, *Un modello per la stima di flussi nel mercato del lavoro affetti da errori di classificazione in rilevazioni retrospettive*, Ottobre 1998.
5. Ginzburg, M. Scaltriti, G. Solinas e R. Zoboli, *Un nuovo autunno caldo nel Mezzogiorno? Note in margine al dibattito sui differenziali salariali territoriali*, Ottobre 1998.
6. M. Forni e S. Paba, *Industrial districts, social environment and local growth. Evidence from Italy*, Novembre 1998.
7. B. Contini, *Wage structures in Europe and in the USA: are they rigid, are they flexible?*, Gennaio 1999.
8. B. Contini, L. Pacelli e C. Villosio, *Short employment spell in Italy, Germany and Great Britain: testing the "Port-of-entry" hypothesis*, Gennaio 1999
9. B. Contini, M. Filippi, L. Pacelli e C. Villosio, *Working careers of skilled vs. unskilled workers*, Gennaio 1999
10. F. Bassi, M. Gambuzza e M. Rasera, *Il sistema informatizzato NETLABOR. Caratteristiche di una nuova fonte sul mercato del lavoro*, Maggio 1999.
11. M. Lalla e F. Pattarin, *Alcuni modelli per l'analisi delle durate complete e incomplete della disoccupazione: il caso Emilia Romagna*, Maggio 1999.
12. A. Paggiaro, *Un modello di mistura per l'analisi della disoccupazione di lunga durata*, Maggio 1999.