

**UN MODELLO DI MARKOV PI
L'ANALISI DEL MERCATO DEL L/**

SERGIO PASTORELLO

No 140

UN MODELLO DI MARKOV PER
L'ANALISI DEL MERCATO DEL LAVORO

Sergio Pastorello *

* Affiliazione: GERMAQ, Université des Sciences Sociales
Place Anatole France
31042 TOULOUSE CEDEX

Tel. 0033/61633656

Fax. 0033/61552263

Riassunto

Le inchieste nazionali sulla situazione del mercato del lavoro forniscono di solito osservazioni in tempo discreto. Gli individui vengono intervistati in corrispondenza di date distinte, per ognuna delle quali si conosce la loro posizione sul mercato (occupati, disoccupati), ma nulla è noto di quanto si è verificato fra le date di osservazione.

Con questi dati non è possibile stimare modelli di durata multi-stato, ed è necessario ricorrere a strumenti statistici fondati sulla classe dei processi markoviani omogenei in tempo continuo. Questo articolo applica un tale approccio a dati provenienti dalla Rilevazione Trimestrale sulle Forze di Lavoro, relativi alla regione Lombardia ed al periodo gennaio 1985 - aprile 1986. Particolare attenzione viene dedicata alle restrizioni implicite imposte da questa classe di modelli, ed alle direzioni di ricerca che promettono di migliorare le proprietà e le performances empiriche.

I INTRODUZIONE

Le statistiche ufficiali del mercato del lavoro sono spesso costituite da una lista, più o meno completa, di indicatori relativi agli stocks, alle grandezze dei flussi di entrata ed uscita, e alle durate di soggiorno in alcuni stati occupazionali.

Se la descrizione del mercato del lavoro fornita da questi indicatori statistici è da un lato esauriente, dall'altro essa è tutt'altro che unificata. Per il ricercatore, così come per i responsabili delle decisioni di politica economica, il problema centrale è quello di capire a fondo i rapporti che collegano fra loro gli indicatori dell'evoluzione degli stocks, dei flussi e delle durate di soggiorno. Per la soluzione di tale problema, è necessario disporre di una modellizzazione, che da un lato sia semplice, ma che dall'altro permetta di descrivere queste interazioni, ritrovando gli ordini di grandezza osservati.

Una possibile via di ricerca in questo campo è rappresentata dall'applicazione dei processi stocastici allo studio delle transizioni individuali sul mercato del lavoro. Lo scopo di tale modellizzazione dinamica è di descrivere, per quanto possibile, la transizione degli individui fra gli stati che caratterizzano ad ogni istante la loro situazione sul mercato.

In Italia, la principale fonte di dati relativi alle forze di lavoro su scala nazionale è costituita dalla Rilevazione Trimestrale delle Forze di Lavoro, condotta dall'ISTAT con cadenza trimestrale. Essa permette di ricostruire la posizione dell'individuo sul mercato al momento dell'intervista, e contiene anche un certo numero di informazioni specifiche (sesso, età, titolo di studio, ecc.) utili nel caratterizzare il suo comportamento. Manca purtroppo qualsiasi informazione sul reddito o sulla ricchezza, così come sulla durata di soggiorno nello stato correntemente occupato - con l'eccezione della disoccupazione.

L'assenza di informazioni sulle durate, peraltro tipico nelle rilevazioni su scala nazionale delle forze di lavoro, riduce drasticamente le possibilità di scelta nell'insieme dei modelli statistici in grado di rappresentare le traiettorie degli individui. In pratica, gran parte dell'informazione disponibile è riassunta nella frequenza con cui le transizioni fra stati occupazionali vengono osservate. La maggior parte delle applicazioni empiriche fondate su questo tipo di dati hanno fatto ricorso a modelli statistici omogenei rispetto al tempo.

Questa classe di modelli è caratterizzata dal fatto che le durate di soggiorno negli stati sono variabili aleatorie esponenziali; in altre parole, la funzione di rischio associata alla durata di soggiorno nello stato i è costante. Quest'ipotesi permette di stimare i parametri del modello a partire da dati sulla *frequenza* delle transizioni fra stati occupazionali, senza bisogno di ricorrere ad informazioni sulle durate di soggiorno.

Nell'ormai vasta letteratura esistente sui modelli di durata, quest'ipotesi viene generalmente rigettata dai dati; empiricamente, sembra molto più plausibile la presenza di quella che viene indicata come *negative duration dependance* (NDD nel seguito): all'aumentare della durata di tempo già spesa nello stato, diminuisce la probabilità istantanea di uscirne.

A sua volta, questo fenomeno può avere due cause: la prima è la presenza di NDD fra gli individui che compongono il campione; la seconda è invece l'effetto di eterogeneità fra gli individui, ciascuno dei quali presenta funzione di rischio costante (si veda Salant (1977)).

Da un punto di vista economico, l'ipotesi di NDD può avere diverse giustificazioni; per esempio può derivare da un meccanismo di selezione attuato dai datori di lavoro al momento dell'assunzione, basato sul capitale umano individuale, il quale è soggetto a deterioramento progressivo durante i periodi di non occupazione. Diversi studi econometrici hanno tentato di identificare l'importanza relativa delle due cause di NDD; se da un lato è chiara la presenza di eterogeneità individuale nella gran parte dei campioni utilizzati, la presenza di NDD a livello individuale sembra essere più controversa (per applicazioni su dati *cross-section*, si vedano, fra gli altri, Flinn e Heckman (1983), Heckman e Borjas (1980), Narendranathan, Nickell e Stern (1985); per uno studio più recente, basato su serie storiche, Jackman e Layard (1991)).

In questo lavoro, l'ipotesi di omogeneità rispetto al tempo è legata direttamente alla natura dei dati. E' perciò evidente l'importanza di un adeguato trattamento dell'eterogeneità individuale, per gli effetti che questa ha sulla validità delle stime dei parametri del processo adottato.

Nella letteratura econometrica, i problemi legati alla stima di questo tipo di modelli, a partire da osservazioni fatte in tempo discreto (cioè dati di *panel*), sono stati trattati in due modi diversi. Una prima generazione di lavori ha considerato la stima delle probabilità di transizione fra due date per un processo di Markov in tempo discreto (catena di Markov); ricordiamo fra questi Hall (1972), Toikka (1976), Clark e Summers (1979), Akerlof e Main (1980, 1981), Poterba e Summers (1986), e, per un'applicazione italiana, Bernardi e Zaccarin (1984). La seconda generazione inizia con il lavoro di Kalbfleisch e Lawless (1985), in cui è proposta una procedura iterativa di stima per le intensità di transizione di un processo di Markov omogeneo in tempo continuo. Altri lavori in questa direzione sono quelli di Geweke, Marshall e Zarkin (1986a e 1986b), e le applicazioni a dati francesi di Fougère e Kamionka (1990) e di Magnac e Robin (1990).

Alla luce della maggiore complessità che una modellazione in tempo continuo comporta rispetto ad una in tempo discreto, vale la pena di domandarsi perchè la prima sia da preferire. A questa domanda è possibile rispondere in diversi modi. Per cominciare, è intuitivamente chiaro che il fenomeno sottostante si evolve in tempo continuo; più importante ancora è osservare che i modelli in tempo discreto dipendono in maniera critica dall'unità di misura del tempo prescelta, problema che ovviamente non si presenta per modelli in tempo continuo.

Si supponga, per esempio, di voler studiare la mobilità intragenerazionale sul mercato del lavoro, e di scegliere come unità di misura 2 generazioni (studiare cioè la mobilità sul lavoro da nonno a nipote). La stima dei parametri del modello fornisce una matrice di probabilità di transizione indicata con p_2 . Si supponga ora di voler utilizzare questi risultati per dedurre p_2 , la matrice delle probabilità di transizione *su una sola generazione*. La relazione fra le due è evidentemente data da $p_2 = p_1 \cdot p_1$.

Il punto fondamentale di questo esempio è il seguente: data p_2 , può non esistere alcuna matrice stocastica p_1 che verifichi questa relazione. In altre parole, non esiste un meccanismo di mobilità su una generazione in grado di replicare quello osservato su due generazioni.

Al contrario, un modello in tempo continuo è per definizione indipendente dall'unità di misura del tempo prescelta; una volta stimati i parametri che lo caratterizzano, è possibile calcolare le metriche di probabilità di transizione su qualsiasi intervallo. Inoltre, come si vedrà nel seguito, i suoi parametri possono essere facilmente stimati anche da osservazioni non equispaziate nel tempo, cosa molto più impegnativa per una specificazione in tempo discreto.

Questo lavoro, che presenta un carattere preliminare ed esplorativo, sia per quanto riguarda le caratteristiche dei dati, sia per ciò che concerne le possibilità del modello, si compone di diverse parti. Una prima descrive brevemente il processo stocastico utilizzato, e le sue principali proprietà. La seconda introduce le procedure di stima dei suoi parametri a partire da osservazioni discrete. La terza contiene le specificazioni econometriche adottate, i cui risultati vengono illustrati nella sezione seguente. Nella conclusione vengono delineate le principali linee di sviluppo di questo lavoro.

II INTRODUZIONE AI PROCESSI DI MARKOV IN TEMPO CONTINUO¹

II.1 Definizioni

Si indichi con $\{X_t | t \geq 0\}$ un processo stocastico in tempo continuo, a valori in uno spazio di stati S discreto e di dimensione finita. Si dice che il processo gode della *proprietà di Markov* quando la traiettoria futura del processo è indipendente da quella passata condizionatamente allo stato occupato al momento presente. In termini più formali, la proprietà di Markov richiede che per degli istanti temporali arbitrari $0 \leq s_0 < s_1 < \dots < s_k < s < t_1 < \dots < t_n$, e gli stati $i_0, i_1, \dots, i_k, i, j, j_1, \dots, j_n$, appartenenti a S , si abbia l'eguaglianza:

$$(2.1) \quad \text{Prob} \left\{ X_t = j, X_{t_1} = j_1, \dots, X_{t_n} = j_n \mid X_{s_0} = i_0, X_{s_1} = i_1, \dots, X_{s_k} = i_k, X_s = i \right\} = \\ = \text{Prob} \left\{ X_t = j, X_{t_1} = j_1, \dots, X_{t_n} = j_n \mid X_s = i \right\}.$$

Le probabilità condizionali:

¹Per brevità, questa sezione si limita ai risultati fondamentali concernenti questi processi; per una trattazione più rigorosa e completa si rimanda a Bhattacharya e Waymire (1990).

$$(2.2) \quad p_{ij}(s,t) = \text{Prob}\left\{X_t=j \mid X_s=i\right\}$$

sono dette *probabilità di transizione*. Si indichi inoltre con $\mathbf{p}(s,t)$ la matrice di elemento generico $p_{ij}(s,t)$. Quando $\mathbf{p}(s,t)$ è funzione solo della differenza $t-s$ indipendentemente dal valore di s e di t , il processo è detto omogeneo rispetto al tempo. In questo caso si ha:

$$(2.3) \quad \begin{aligned} \forall t > 0, \quad p_{ij}(0,t) &= p_{ij}(t) & i,j \in S \\ \forall t \geq s, \quad p_{ij}(s,t) &= p_{ij}(t-s) & i,j \in S. \end{aligned}$$

Le traiettorie di un processo di questo tipo danno luogo a funzioni costanti a tratti e continue a destra; se si suppone che sia possibile ordinare gli stati sulla base di un qualche criterio, esse possono essere rappresentate graficamente nel modo seguente:

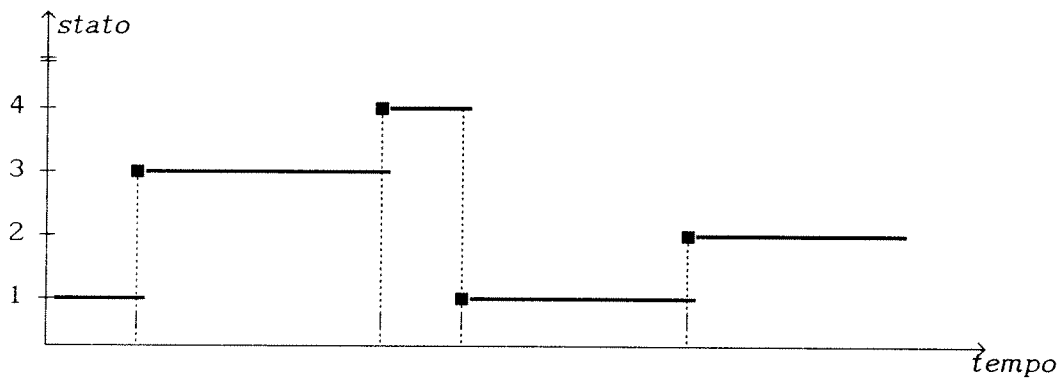


Figura 1
Traiettoria-tipo di un processo stocastico a stati discreti

La caratterizzazione della legge del processo, cioè di tutte le traiettorie che esso può descrivere, viene completata definendo le probabilità iniziali:

$$(2.4) \quad \pi_i(0) = \text{Prob}\left\{X_0=i\right\}$$

e le probabilità di transizione $\mathbf{p}(t-s)$, $\forall t \geq s$, $\forall i,j \in S$, poichè sulla base di queste è allora possibile calcolare la probabilità di ogni insieme finito di osservazioni del processo. Infatti, per ogni sequenza di istanti di osservazione $0=s_0 \leq s_1 \leq \dots \leq s_n$ e di stati occupati dal processo in queste date i_0, i_1, \dots, i_n , si ha:

$$(2.5) \quad \text{Prob}\left\{X_{s_0}=i_0, X_{s_1}=i_1, \dots, X_{s_n}=i_n\right\} = \pi_{i_0}(0) \cdot \prod_{h=1}^n p_{i_{h-1}i_h}(s_h - s_{h-1})$$

I parametri che caratterizzano la legge di questa classe di processi possono essere introdotti nella maniera seguente. Si supponga che:

$$(2.6) \quad \lim_{t \rightarrow 0^+} \mathbf{p}(t) = \mathbf{p}(0) = \mathbf{I}_K$$

dove con \mathbf{I}_K si indica la matrice identità ($K \times K$), e K rappresenta la cardinalità di S . L'espressione (2.6) implica la continuità, elemento per elemento, di $\mathbf{p}(t)$ in $t=0$. E' possibile dimostrare che la continuità a sua volta implica la derivabilità da destra in t di $\mathbf{p}(t)$. Si definisce perciò *intensità di transizione* la quantità:

$$(2.7) \quad q_{ij} = \lim_{t \rightarrow 0^+} \frac{p_{ij}(t) - p_{ij}(0)}{t}$$

mentre la matrice costituita di tali elementi, cioè:

$$(2.8) \quad \mathbf{Q} = \left\{ q_{ij} \right\}_{ij \in S}$$

viene definita *generatore istantaneo* del processo.

A partire dalle proprietà della matrice delle probabilità condizionali di transizione² è possibile derivare le seguenti restrizioni su \mathbf{Q} :

$$(2.9) \quad \left\{ \begin{array}{ll} q_{ij} \geq 0 & \forall i \neq j, \\ q_{ii} = - \sum_{\substack{j \in S \\ j \neq i}} q_{ij} \leq 0 & \forall i \end{array} \right.$$

Un punto estremamente importante è costituito dalla relazione fra intensità di transizione e probabilità condizionali di transizione. Si può dimostrare che questa relazione è data da³:

$$(2.10) \quad \mathbf{p}(t) = \exp(\mathbf{Q}t), \quad t \geq 0,$$

dove la matrice $\exp(\mathbf{Q}t)$ è definita nel modo seguente:

² $\mathbf{p}(t)$ è una matrice stocastica $\forall t$; in altre parole $p_{ij}(t) \geq 0 \quad \forall i, j \in S$, e $\sum_j p_{ij}(t) = 1, \quad \forall i, \forall t$.

³(2.10) è la soluzione comune unica delle due equazioni differenziali matriciali $\dot{\mathbf{p}}(t) = \mathbf{p}(t) \cdot \mathbf{Q}$ e $\dot{\mathbf{p}}(t) = \mathbf{Q} \cdot \mathbf{p}(t)$ (si veda Bhattacharya e Waymire (1990)).

$$(2.11) \quad \exp(\mathbf{Q}t) = \sum_{n=0}^{\infty} \frac{(\mathbf{tQ})^n}{n!} = \mathbf{I}_K + \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbf{Q}^n.$$

Si dimostra (si veda Doob (53), pag. 240-241) che se \mathbf{Q} verifica la condizione (2.9), $\mathbf{p}(t)$ è in effetti una matrice di probabilità condizionali di transizione.

La (2.10) rappresenta una relazione fondamentale in questo tipo di modelli. Attraverso di essa, infatti, è possibile risalire alle probabilità di transizione che, come si è visto, caratterizzano completamente la traiettoria percorsa dal processo nel tempo. In altre parole, la matrice \mathbf{Q} contiene i parametri del processo, che dovranno essere stimati a partire dalle traiettorie osservate.

II.2. Proprietà

Bhattacharya e Waymire (90, pag. 275 e segg.) propongono un'importante caratterizzazione probabilistica per i processi introdotti nella sezione II.1. Prima di presentare questo importante risultato, tuttavia, è necessario introdurre due concetti che saranno utili nel seguito.

Definizione

Per un processo stocastico $\{X_t\}$, la successione di v. a. $\{\tau_n | n=0,1,2,\dots\}$ definita nel modo seguente:

$$(2.12) \quad \tau_0 = 0; \tau_1 = \inf\left\{t>0 | X_t \neq X_0\right\}; \dots; \tau_n = \inf\left\{t>\tau_{n-1} | X_t \neq X_{\tau_{n-1}}\right\}$$

è composta dai *tempi di salto* del processo. A partire da $\{\tau_n\}$ si costruisce la successione di v. a. $\{T_n | n=0,1,2,\dots\}$ dei *tempi di soggiorno*, definiti da:

$$(2.13) \quad T_0 = \tau_1; T_{n-1} = \tau_n - \tau_{n-1} \quad \text{per } n \geq 1. \quad \square$$

In generale, T_n rappresenta il tempo di soggiorno nello stato nel quale il processo salta alla sua n -esima transizione.

Si ponga:

$$(2.14) \quad \lambda_i = -q_{ii}; k_{ii} = 0; k_{ij} = \frac{q_{ij}}{-q_{ii}} \quad \text{per } i \neq j$$

se $q_{ij} \neq 0$; e:

$$(2.15) \quad k_{ii} = 1; k_{ij} = 0 \quad \text{per } i \neq j$$

se $q_{ii} = 0$.

Il teorema seguente rappresenta uno strumento indispensabile

nell'interpretazione dei modelli statistici considerati in questo lavoro.

Teorema

- (a) Sia $\{X_t\}$ una catena di Markov di generatore istantaneo Q e stato iniziale i_0 . Allora $\{Y_n \equiv X_{\tau_n} \mid n=0, 1, \dots\}$ rappresenta una catena di Markov in tempo discreto di matrice di probabilità di transizione K , di elemento generico k_{ij} definito sopra. Data $\{Y_n\}$, i tempi di soggiorno negli stati T_0, T_1, T_2, \dots sono v. a. esponenziali indipendenti, di parametri $\lambda_{i_0}, \lambda_{i_1}, \lambda_{i_2}, \dots$ rispettivamente.
- (b) Supponendo, viceversa, che $\{Y_n\}$ sia una catena di Markov in tempo discreto di matrice di probabilità di transizione K e stato iniziale i_0 , e che per ogni sequenza di stati i_0, i_1, i_2, \dots , sia T_0, T_1, T_2, \dots una sequenza di v. a. esponenziali indipendenti di parametro rispettivamente $\lambda_{i_0}, \lambda_{i_1}, \lambda_{i_2}, \dots$; supponendo inoltre che le sequenze $\{T_0, T_1, T_2, \dots\}$, corrispondenti alle diverse sequenze di stati, siano indipendenti, definiamo:

$$(2.16) \quad \begin{cases} X_t = Y_0, & \text{per } 0 \leq t < T_0 \\ X_t = Y_{k+1} & \text{per } T_0 + T_1 + \dots + T_k \leq t < T_0 + T_1 + \dots + T_{k+1}. \end{cases}$$

Allora il processo $\{X_t \mid t \geq 0\}$ così costruito è markoviano di stato iniziale i_0 e generatore istantaneo Q .

Un'altra proprietà estremamente importante dei processi stocastici sin qui considerati è che essi ammettono, sotto un'ulteriore ipotesi, l'esistenza di una *distribuzione d'equilibrio*, o *distribuzione stazionaria*, o ancora *distribuzione invariante*.

Proposizione

- (a) Per un processo di Markov omogeneo in tempo continuo, qualunque sia la condizione iniziale, la probabilità $\pi_i(t)$ che lo stato i sia occupato al tempo t tende, quando $t \rightarrow \infty$, ad un limite ρ_i , e $\{\rho_i\}_{i \in S}$ è una distribuzione di probabilità; tale cioè che:

$$(2.17) \quad \rho_i > 0 \quad \forall i \in S; \quad \sum_{i \in S} \rho_i = 1.$$

- (b) Se la probabilità iniziale d'occupazione degli stati è $\rho = \{\rho_i\}_{i \in S}$, allora $\pi_i(t) = \rho_i \quad \forall t \geq 0, \forall i \in S$.
- (c) Le probabilità d'equilibrio si ottengono risolvendo il sistema di

equazioni lineari:

$$(2.18) \quad \mathbf{r}'\mathbf{Q} = 0$$

nelle variabili \mathbf{r} .

III PROCEDURE DI STIMA

Una procedura di stima generale per questi modelli è stata proposta da Kalbfleisch e Lawless (1985). Tuttavia, nel caso in cui le osservazioni siano equidistanti fra loro, è possibile stimare \mathbf{Q} in maniera diretta secondo la tecnica proposta, per esempio, in Singer e Spilerman (1976). Anche se questo modo di procedere presenta dei difetti importanti, esso presenta tuttavia il vantaggio di introdurre l'importante concetto di *compatibilità* delle osservazioni con la struttura statistica del modello.

III.1 Osservazioni equispaziate nel tempo

Si supponga di osservare n individui in corrispondenza delle $(L+1)$ date distinte e fra loro equidistanti t_0, t_1, \dots, t_L , tali che

$$(3.1) \quad t_l - t_{l-1} = T \quad l=1,2,\dots,L$$

Si definisca inoltre:

- $n_{ij}(l)$ il numero di individui che transitano dallo stato i allo stato j fra le date t_{l-1} e t_l ;
- $n_i(l-1)$ il numero di individui che occupavano lo stato i al tempo t_{l-1} .

La verosimiglianza delle traiettorie osservate degli individui, in questo caso, è data dalla formula seguente:

$$(3.2) \quad L(\mathbf{p}(T)) = \prod_{l=1}^L \prod_{j=1}^K \left[p_{ij}(T) \right]^{n_{ij}(l)} = \prod_{j=1}^K \left[p_{ij}(T) \right]^{\sum_{l=1}^L n_{ij}(l)}$$

Lo stimatore di massima verosimiglianza degli elementi di $\mathbf{p}(T)$ si ottiene massimizzando L sotto il vincolo:

$$(3.3) \quad \sum_{j=1}^K p_{ij}(T) = 1 \quad \forall i \in S$$

La massimizzazione della funzione di verosimiglianza è data da: (si veda Anderson e Goodman (1957)):

$$(3.4) \quad \hat{p}_{ij}(T) = \frac{\sum_{l=1}^L n_{ij}(l)}{\sum_{l=1}^L n_i(l-1)}$$

Una volta ottenuta la stima di massima verosimiglianza di $\mathbf{p}(T)$, è possibile risalire direttamente allo stimatore, ugualmente di massima verosimiglianza, di \mathbf{Q} invertendo la relazione fondamentale:

$$(3.5) \quad \hat{\mathbf{p}}(T) = \exp(\hat{\mathbf{Q}}T) \quad T > 0.$$

La relazione inversa è data da (si vedano Singer e Spilerman (1976)):

$$(3.6) \quad \hat{\mathbf{Q}} = T^{-1} \log(\hat{\mathbf{p}}(T)) = T^{-1} \sum_{h=1}^{\infty} \frac{(-1)^h (\hat{\mathbf{p}}(T) - \mathbf{I})^h}{h!}.$$

Si definisca:

$$(3.7) \quad \mathbf{Q} = \left\{ \mathbf{Q} \mid q_{ij} \geq 0, \forall i \neq j; \sum_{j=1}^K q_{ij} = 0, \forall i \right\}$$

l'insieme dei generatori istantanei ammissibili. Se $\hat{\mathbf{Q}}$, ottenuta invertendo la relazione fondamentale, appartiene a \mathbf{Q} , allora $\hat{\mathbf{Q}}$ è stima di massima verosimiglianza di \mathbf{Q} . Questa procedura può dar luogo a due tipi di problemi:

- 1) compatibilità: data la matrice $\hat{\mathbf{p}}$, $\hat{\mathbf{Q}} \notin \mathbf{Q}$. In questo caso $\hat{\mathbf{p}}$ è detta *non compatibile*, nel senso che essa non può essere stata generata da un processo di Markov omogeneo in tempo continuo;
- 2) identificazione: data la matrice $\hat{\mathbf{p}}$, esistono diverse $\hat{\mathbf{Q}} \in \mathbf{Q}$, ottenute invertendo la relazione fondamentale.

Questa molteplicità di soluzioni è dovuta al fatto che la funzione logaritmo è a valori multipli se definita sull'insieme dei numeri complessi (al quale appartengono gli autovalori di \mathbf{p}). Il problema dell'identificazione può essere risolto ricorrendo ad informazioni a priori sulla natura della matrice \mathbf{Q} , oppure utilizzando osservazioni supplementari.

La caratterizzazione della compatibilità mediante condizioni necessarie e/o sufficienti da applicare direttamente alla matrice \mathbf{p} è un problema ancora aperto in statistica; tuttavia le implicazioni della compatibilità sull'analisi di dati di panel sono abbastanza limitate. Ciò è dovuto a ragioni diverse:

- a) se $\hat{\mathbf{p}}$ non è compatibile, ciò non significa che il processo di Markov si adatti male ai dati; inoltre se si suppone che \mathbf{p} sia osservata con errore, è possibile che la vera matrice $\tilde{\mathbf{p}}$ sia compatibile e allo stesso tempo molto "vicina" a \mathbf{p} ;
- b) anche se $\hat{\mathbf{p}}$ è compatibile è possibile che il modello si adatti male alle osservazioni, o che i parametri siano stimati in maniera

imprecisa; questo problema si presenta in particolare quando la lunghezza T dell'intervallo è grande.

Infine, è importante far osservare che la procedura di stima proposta in questo paragrafo si adatta solo ad osservazioni equispaziate, e ad una matrice Q "saturata", nella quale cioè lo spazio parametrico sia composto dagli elementi fuori della diagonale principale in Q . Al contrario, la procedura presentata nella sezione seguente si adatta a parametrizzazioni diverse ed in particolare allo studio di osservazioni non equispaziate.

III.2 Osservazioni non equispaziate

Si definisca:

$$(3.8) \quad w_l = t_l - t_{l-1} \quad l=1,2,\dots,L$$

Il problema di stima di massima verosimiglianza è formalmente descritto da:

$$(3.9) \quad \hat{\vartheta} = \operatorname{argmax}_{\vartheta \in \Theta} \left\{ \log L(\vartheta) = \sum_{l=1}^L \sum_{i=1}^K n_{ij}(1) (\log [p_{ij}(w_l)]) \right\}$$

dove: $p(t) = \exp(Qt)$, $Q = Q(\vartheta)$, $\vartheta = (\vartheta_1, \vartheta_2, \dots, \vartheta_b)'$, e $\Theta = \{\vartheta \in \mathbb{R}^b \mid Q(\vartheta) \in Q\}$.

L'idea centrale della procedura, proposta da Kalbfleisch e Lawless (1985), è di utilizzare un algoritmo efficiente per il calcolo di $\partial p(t; \vartheta) / \partial \vartheta$. Questo è fornito dalla formula seguente:

$$(3.10) \quad \frac{\partial p(t; \vartheta)}{\partial \vartheta_u} = A V_u A^{-1} \quad u=1,2,\dots,b$$

dove V_u è una matrice ($K \times K$), di elemento (i,j) dato da:

$$(3.11) \quad \begin{cases} g_{ij}^{(u)} (\exp(d_i t) - \exp(d_j t)) / (d_i - d_j) & \text{se } i \neq j \\ g_{ii}^{(u)} t \cdot \exp(d_i t) & \text{se } i = j \end{cases}$$

e:

$$(3.12) \quad g_{ij}^{(u)} = \text{elemento } (i,j) \text{ in } G^{(u)} = A^{-1} \left[\frac{\partial Q(\vartheta)}{\partial \vartheta_u} \right] A.$$

Dietro l'apparente complessità di questa formula si nasconde il grande vantaggio di poter calcolare $\partial p(t; \vartheta) / \partial \vartheta$ senza ricorrere ad una rappresentazione esplicita (ed estremamente complessa) di $p(t; \vartheta)$ in funzione di ϑ .

A questo punto la procedura iterativa di stima è facilmente descritta: il vettore dello "score" è dato da:

$$(3.13) \quad S_u(\vartheta) = \frac{\partial \log L(\vartheta)}{\partial \vartheta_u} = \sum_{i=1}^L \sum_{j=1}^K n_{ij}(1) \cdot \frac{\partial \log(p_{ij}(w_1))/\partial \vartheta_u}{p_{ij}(w_1)}$$

per $u = 1, 2, \dots, b$. La matrice d'informazione si scrive:

$$(3.14) \quad \frac{\partial^2 \log L(\vartheta)}{\partial \vartheta_u \partial \vartheta_v} = \sum_{i=1}^L \sum_{j=1}^K n_{ij}(1) \cdot \left\{ \frac{\partial^2 p_{ij}(w_1)/\partial \vartheta_u \partial \vartheta_v}{p_{ij}(w_1)} - \frac{(\partial p_{ij}(w_1)/\partial \vartheta_u) \cdot (\partial p_{ij}(w_1)/\partial \vartheta_v)}{(p_{ij}(w_1))^2} \right\}$$

Ricorrendo al metodo dello score è possibile evitare il calcolo delle derivate seconde, perchè:

$$(3.15) \quad E \left\{ - \frac{\partial^2 \log L(\vartheta)}{\partial \vartheta_u \partial \vartheta_v} \right\} = \sum_{i=1}^L \sum_{j=1}^K \frac{E(n_i(1-1))}{p_{ij}(w_1)} \cdot \frac{\partial p_{ij}(w_1)}{\partial \vartheta_u} \cdot \frac{\partial p_{ij}(w_1)}{\partial \vartheta_v}$$

e questa quantità può essere stimata da:

$$(3.16) \quad M_{uv}(\vartheta) = \sum_{i=1}^L \sum_{j=1}^K \frac{n_i(1-1)}{p_{ij}(w_1)} \cdot \frac{\partial p_{ij}(w_1)}{\partial \vartheta_u} \cdot \frac{\partial p_{ij}(w_1)}{\partial \vartheta_v}$$

Data una stima iniziale $\vartheta^{(0)}$, l'algoritmo fornisce un valore aggiornato $\vartheta^{(1)}$ mediante la formula:

$$(3.17) \quad \vartheta^{(1)} = \vartheta^{(0)} + \mathbf{M}^{-1}(\vartheta^{(0)}) \cdot S(\vartheta^{(0)})$$

e le iterazioni continuano fino alla convergenza, che si suppone raggiunta quando la distanza fra due stime successive di ϑ è inferiore ad una certa soglia. Lo stimatore ϑ così ottenuto gode delle proprietà seguenti:

- a) $\mathbf{M}^{-1}(\hat{\vartheta})$ fornisce una stima della matrice di covarianza asintotica di $\hat{\vartheta}$;
- b) se il vero valore di ϑ giace all'interno della regione ammissibile, $n^{-1/2}(\hat{\vartheta} - \vartheta)$ tende in distribuzione ad una normale multivariata.

IV LE SPECIFICAZIONI ADOTTATE

Come si è visto, la procedura proposta da Kalbfleisch e Lawless rappresenta uno strumento flessibile per stimare i parametri indipendenti in \mathbf{Q} . Il caso di una matrice \mathbf{Q} saturata (un parametro per ogni intensità di transizione) costituisce un primo ed utile tentativo di adattare il modello ai dati, ma non fornisce risultati interpretabili in termini di quantità economicamente significative. A tal fine, è necessaria la formulazione di un modello economico che descriva la decisione, da parte degli individui, di transitare da uno stato ad un altro. Questo modello impone dei vincoli sui parametri della forma ridotta, rendendone possibile l'interpretazione in

termini comportamentali. Questi vincoli possono inoltre essere sottoposti a verifica empirica studiando l'adattamento ai dati del modello vincolato (Q saturata) e non vincolato.

In questa sezione vengono presentate le due diverse parametrizzazioni della matrice Q che sono state adottate. La prima rappresenta il caso di Q saturata, e verrà illustrato brevemente, poichè non contiene concetti innovativi rispetto a quanto visto sinora. La seconda contiene un tentativo di modellazione strutturale di Q , utilizzando il modello di utilità stocastica dinamica. Nell'ultima parte viene illustrata la tecnica utilizzata per ottenere la stima iniziale di ϑ necessaria per l'applicazione dell'algoritmo iterativo.

IV.1 Forma ridotta non vincolata

Da un punto di vista formale, il problema di stima del modello nel caso di una matrice Q saturata costituisce un caso di massimizzazione vincolata della verosimiglianza:

$$(4.1) \quad \begin{aligned} & \max \log L(\mathbf{p}(t)) \\ & \text{s.v. } \mathbf{p}(t) = \exp(Q\mathbf{t}) \\ & \text{e } Q \in \mathcal{Q}. \end{aligned}$$

Il vincolo di positività degli elementi liberi di Q è stato introdotto nella procedura iterativa ponendo:

$$(4.2) \quad q_{ij} = \begin{cases} x_{ij}^2 & \text{per } i \neq j \\ - \sum_{k \neq i} q_{ik} & \text{per } i = j \end{cases} .$$

E' interessante notare che dalla stima \hat{Q} sono immediatamente dedotte alcune quantità di grande interesse nell'analisi delle mobilità individuali sul mercato del lavoro, quali:

- a) *durate medie di soggiorno negli stati*, pari a $-q_{ii}^{-1}$ (si ricordi infatti che le durate di soggiorno nei vari stati seguono leggi esponenziali di parametro q_{ii});
- b) *probabilità d'occupazione d'equilibrio*.

Inoltre, per entrambi i casi, è possibile ottenere facilmente una stima della rispettiva matrice di covarianza asintotica applicando la versione multivariata del δ -Theorem (si veda Kalbfleisch e Lawless (1985)).

IV.2 Forma strutturale

Il modello di utilità stocastica dinamica, presentato in questa sezione, sta alla base di molti studi empirici dei fenomeni di transizione fra stati; alcuni esempi possono essere Burdett, Kiefer, Mortensen e Neumann (1984), Mortensen e Neumann (1984), Magnac e Robin (1990), Olsen, Smith e Farkas (1986), e con un punto di vista leggermente diverso, Flinn e Heckman (1982).

L'idea centrale di questo modello è costituita dall'ipotesi che il

lavoratore transiti da uno stato all'altro in risposta a shocks aleatori nelle possibilità di occupazione e nel costo opportunità del tempo libero.

Supponiamo che la posizione di un individuo sul mercato del lavoro possa essere perfettamente descritta dall'appartenenza ad uno dei tre stati disgiunti indicati con O (occupazione), D (disoccupazione) e I (non forza lavoro, o inattività). Ogni individuo riceve dal suo soggiorno nello stato i , l'utilità istantanea $u_i(w, \varepsilon)$, dove w rappresenta il valore corrente del salario, ed ε costituisce un indicatore del valore del tempo libero.

L'arrivo di shocks al sistema è rappresentato dall'arrivo simultaneo di nuovi valori di w ed ε . Questi arrivi si verificano secondo un processo di Poisson (quindi i periodi fra un arrivo e l'altro sono delle variabili esponenziali) di parametro λ_i , $i=O, D, I$, indipendenti fra loro e dai vecchi valori di w ed ε . I nuovi valori di w ed ε , qualunque sia lo stato occupato al momento dello shock, rappresentano delle realizzazioni indipendenti ed identicamente distribuite di una variabile aleatoria di funzione di ripartizione $F(w, \varepsilon)$.

Al verificarsi di uno shock, l'individuo fronteggia un problema di scelta fra alternative discrete, la cui soluzione consiste nello scegliere lo stato che gli garantisce il più elevato flusso atteso di utilità, al tempo presente.

Utilizzando il principio di programmazione dinamica di Bellman, è possibile definire il valore presente atteso dell'utilità associata allo stato i :

$$(4.3) \quad V_i(w, \varepsilon) = \mathbb{E} \left\{ \int_0^{\tilde{t}_i} u_i(w, \varepsilon) \cdot e^{-\rho s} ds + e^{-\rho \tilde{t}_i} \cdot \max_k V_k(\tilde{w}, \tilde{\varepsilon}) \right\}$$

L'interpretazione di questa formula è immediata: durante il soggiorno nello stato i , di durata aleatoria \tilde{t}_i , rispetto alla quale è calcolato il valore atteso, l'individuo percepisce un flusso di utilità costante $u_i(w, \varepsilon)$, opportunamente scontato ad un tasso ρ ; il secondo addendo rappresenta l'utilità attesa (rispetto ai nuovi valori di w ed ε) che deriva da una scelta ottimale da parte dell'individuo al momento di effettuare la transizione. Le ipotesi fatte sinora permettono di esplicitare ulteriormente $V_i(w, \varepsilon)$ in funzione dei parametri strutturali:

$$(4.4) \quad V_i(w, \varepsilon) = \frac{u_i(w, \varepsilon)}{\rho + \lambda_i} + \frac{\lambda_i}{\rho + \lambda_i} \int \max_k V_k(\tilde{w}, \tilde{\varepsilon}) dF(\tilde{w}, \tilde{\varepsilon})$$

dove l'integrale è calcolato sul dominio di w ed ε .

Conviene introdurre ora un'utile partizione del dominio di (w, ε) , definita da:

$$(4.5) \quad A_j = \left\{ (w, \varepsilon) \mid V_j(w, \varepsilon) = \max_k V_k(w, \varepsilon) \right\}$$

Gli A_j vengono definiti *insiemi di scelta*. A partire da questi è facile definire la probabilità di scegliere lo stato j al momento di effettuare la transizione; dette *probabilità di transizione* vengono definite dall'espressione:

$$(4.6) \quad \pi_j = \int_{A_j} dF(\tilde{w}, \tilde{\varepsilon}).$$

Si ricordi ora che i parametri della forma ridotta non vincolata, cioè gli elementi liberi di Q , rappresentano la probabilità istantanea di transitare verso lo stato j partendo dallo stato i ; di conseguenza la relazione fra parametri di forma strutturale e parametri di forma ridotta è la seguente:

$$(4.7) \quad q_{ij} = \lambda_i \cdot \pi_j \quad \text{per } i \neq j.$$

La probabilità istantanea di transitare dallo stato i allo stato j è data dal prodotto fra il tasso istantaneo di arrivo degli shocks nello stato i e la probabilità di effettuare la transizione verso lo stato j . Inoltre, per l'ipotesi fatta sulla legge di arrivo degli shocks, la funzione di rischio delle durate nello stato i è data dall'equazione:

$$(4.8) \quad -q_{ii} = \sum_{j \neq i} \lambda_i \pi_j = \lambda_i \cdot (1 - \pi_i).$$

La forma ridotta vincolata è stata così stimata ricorrendo alla parametrizzazione seguente:

$$(4.12) \quad \left\{ \begin{array}{ll} q_{ij} = \lambda_i \pi_j & i, j = O, D, I; i \neq j \\ \lambda_i = x_i^2 & i = O, D, I \\ \pi_j = \frac{e^{y_j}}{\sum_{h=1}^k e^{y_h}} & j = O, D, I; y_0 = 0, \end{array} \right.$$

che presenta il vantaggio di soddisfare, per costruzione, i vari vincoli su λ_i e π_j .

L'aver scelto per i π_i una forma di tipo logit equivale ad assumere una $F(w, \varepsilon)$ di tipo *extreme value*. Questa ipotesi non è certamente di poco conto, ma la sua importanza scompare rispetto a quelle introdotte in precedenza. Intuitivamente sembra molto plausibile ritenere che valori successivi di w ed ε siano correlati tra loro, ed anche che la loro distribuzione dipenda dallo stato occupato all'arrivo dello shock; ancora meno plausibile è l'implicazione che ogni transizione da uno stato all'altro sia assolutamente volontaria. Tuttavia, ogni tentativo di rimediare a questi inconvenienti impedirebbe di

interpretare il modello teorico come una restrizione sulla classe dei processi di Markov in tempo continuo omogenei. Per esempio, l'ipotesi che gli shocks siano intervallati da durate che evolvono esponenzialmente garantisce che le intensità di transizione vincolate siano costanti nel tempo. In altre parole, la scelta di un modello teorico molto semplificato di questo tipo è imposta dalla struttura statistica nella quale esso va inserito.

Questa formulazione, peraltro, presenta alcuni vantaggi non trascurabili. I parametri strutturali λ_i e π_j sono definiti nella letteratura anglossassone rispettivamente *chance* e *choice*, cioè componente aleatoria e componente di scelta. Essi permettono, in particolare, di distinguere all'interno del fenomeno di transizione l'importanza di due effetti distinti: quello di preferenza individuale per uno stato o per l'altro (scelta discreta fra alternative distinte), e quello di frequenza delle opportunità di transizione (scelta di quando effettuare il passaggio da uno stato all'altro). In generale, l'importanza relativa di questi due effetti è legata alle caratteristiche osservate e non osservate del soggetto.

IV.3 I valori iniziali

E' chiaro che l'applicazione dell'algoritmo iterativo di Kalbfleisch e Lawless, descritto nella sezione II, richiede una stima iniziale del vettore dei parametri ϑ ; nelle applicazioni considerate in questo lavoro, ϑ rappresenta il vettore dei parametri indipendenti in \mathbf{Q} (cioè fuori dalla diagonale principale), nel caso della "forma ridotta"; e i due vettori λ e π per la "forma strutturale". Come sottolineato da Kalbfleisch e Lawless, è importante che questa stima iniziale non sia troppo lontana da quella di massima verosimiglianza; empiricamente, infatti, si è riscontrato che in questo caso la convergenza può essere estremamente laboriosa, oppure non verificarsi affatto. In questo lavoro sono state considerate due possibili soluzioni al problema, per avere una parziale garanzia sull'unicità globale del massimo ottenuto iterativamente. In tutti i casi considerati, in effetti, i valori verso i quali converge l'algoritmo sono stati gli stessi, indipendentemente dai valori iniziali.

Una prima procedura per ottenere $\hat{\vartheta}_0$ è data da una tecnica analoga a quella illustrata nella sezione III.1. Anche se le osservazioni non sono equispaziate nel tempo, è sempre possibile applicare la relazione fondamentale inversa (3.6) alle matrici \mathbf{p} stimate su diversi intervalli. Per evitare di scontrarsi con il problema della compatibilità ($\mathbf{Q} \neq \mathbf{Q}$) è preferibile applicare la procedura sull'intervallo per il quale la matrice \mathbf{p} non contenga celle vuote; infatti, la più semplice delle condizioni necessarie per la compatibilità di una matrice di transizione con un processo di Markov omogeneo in tempo continuo è che essa non contenga celle a frequenza nulla (si veda Singer e Spilerman (1976)). Anche così facendo, tuttavia, è possibile che la presenza di non compatibilità impedisca il ricorso a questa procedura.

Una seconda tecnica è, a differenza della precedente, *sempre* applicabile (non presenta difficoltà legate alla compatibilità), ma si basa su un'ipotesi piuttosto restrittiva, e cioè che nell'intervallo di tempo considerato gli individui effettuino o 0 o 1 transizione. Si ricordi che il Teorema presentato nella sezione II permette di interpretare $-q_{ii}$ come il parametro della legge

esponenziale che governa la durata di soggiorno nello stato i , mentre $-q_{ij}/q_{ii}$ rappresenta la probabilità di transitare dallo stato i allo stato j . Se il numero massimo di transizioni sul periodo è 1, la stima dell'intensità q_{ij} è data da⁴

$$(4.13) \quad \hat{q}_{ij} = \frac{n_{ij}}{n_{i.}} \cdot \left[-\frac{1}{t} \log \left(\frac{n_{ii}}{n_{i.}} \right) \right] \cdot \left[1 - \frac{n_{ii}}{n_{i.}} \right]^{-1}$$

La stima iniziale dei parametri della specificazione strutturale è più complessa. Si noti, infatti, che per $K=2$ il modello non è identificato, mentre per $K \geq 3$ esso è sovraidentificato. In particolare per $K=3$ è possibile esplicitare il seguente vincolo di sovraidentificazione:

$$(4.14) \quad \frac{q_{OI}}{q_{ID}} \cdot \frac{q_{ID}}{q_{IO}} = \frac{q_{DI}}{q_{DO}}$$

Se nella procedura iterativa la sovraidentificazione può essere facilmente imposta attraverso una opportuna parametrizzazione della matrice Q , il problema viceversa si pone nel calcolo dei valori iniziali. La procedura seguita è stata di scegliere arbitrariamente una delle due distinte funzioni di Q che permette di passare da una stima di questa alle stime dei parametri della specificazione strutturale. Per passare dalla stima iniziale \tilde{Q} a $\tilde{\lambda}$ e $\tilde{\pi}$ si è perciò proceduto nella maniera seguente:

$$(4.15) \quad \tilde{\pi}_D = \left[1 + \frac{\tilde{q}_{DI}}{\tilde{q}_{DO}} + \frac{\tilde{q}_{ID}}{\tilde{q}_{IO}} \right]^{-1}$$

$$(4.16) \quad \tilde{\pi}_I = \frac{\tilde{q}_{OI}}{\tilde{q}_{OD} \tilde{\pi}_D}$$

$$(4.17) \quad \tilde{\pi}_O = 1 - \tilde{\pi}_D - \tilde{\pi}_I$$

$$(4.18) \quad \tilde{\lambda}_i = -\frac{\tilde{q}_{i1}}{1 - \tilde{\pi}_1} \quad i=O, D, I$$

Si noti che la presenza di un vincolo di sovraidentificazione può essere sottoposto a verifica empirica mediante una opportuna procedura di test. In effetti la forma ridotta vincolata è un caso particolare di quella non vincolata. Questo approccio permette di studiare, in maniera molto semplice, la plausibilità del modello teorico alla luce dei comportamenti individuali osservati.

⁴Si veda Pastorello (1992), appendice 4.3.

V APPLICAZIONE AI DATI ITALIANI RELATIVI AL PERIODO 1985/1988

Questa sezione si divide in due parti. La prima descrive brevemente i dati utilizzati nell'applicazione empirica. Le loro caratteristiche sembrano infatti essere importanti nell'interpretare la validità dei risultati ottenuti, così come nel delineare alcune vie di sviluppo e di miglioramento del lavoro.

La seconda parte commenta brevemente le tabelle riassuntive dei risultati.

V.1 I dati

In Italia la principale fonte di informazioni statistiche sul mercato del lavoro è la *Rilevazione Trimestrale sulle Forze di Lavoro* (RTFL nel seguito), condotta dall'Istat su base trimestrale. Una presentazione dettagliata e approfondita delle sue caratteristiche esula dagli obiettivi di questo lavoro; per questo, si rinvia a Fabbris e Bernardi (1986), e alle varie pubblicazioni dell'ISTAT citate nella loro bibliografia.

La struttura del campione su scala nazionale è a rotazione, con cadenza di osservazione trimestrale, e suddiviso in quarti. La rilevazione avviene nei mesi di gennaio, aprile, luglio e ottobre di ogni anno. Ogni quarto campionario viene osservato in due diverse date, separate fra loro da un intervallo di tre mesi; esce dalla rilevazione per le due date successive, e rientra in seguito per essere osservato altre due volte. Ogni famiglia intervistabile viene dunque osservata 4 volte, separate fra loro da intervalli di 3, 9 e 3 mesi rispettivamente.

I dati vengono forniti dall'Istat sotto forma di osservazioni puntuali; in altre parole, per applicazioni del tipo descritto in questo lavoro, è innanzitutto necessario procedere al collegamento delle risposte (*linkage*) fornite dallo stesso individuo in corrispondenza di date diverse. Quest'operazione è stata effettuata da Giusti, Marliani e Torelli (1987 e 1988) utilizzando una procedura di *matching* probabilistico.

Il risultato di questa procedura è la costruzione di un *file* nel quale ogni riga rappresenta le risposte fornite da un individuo in corrispondenza delle quattro date d'inchiesta. Per una serie di ragioni (errori di risposta, errori di codificazione, sostituzione delle unità non rispondenti), è abbastanza difficile operare un *linkage* perfetto - quattro risposte effettive su una riga. Gli individui in questo *file* sono perciò fittizi, e non corrispondono a quelli effettivamente campionati. In un certo numero di casi, la riga contiene una sola risposta effettiva, e si rivela perciò inutile nella ricostruzione di una transizione; di conseguenza tutte le osservazioni di questo tipo sono state abbandonate. In tutti gli altri casi, è possibile ricostruire almeno una transizione; questa può essere a tre mesi (l'individuo risponde in corrispondenza dei mesi di gennaio - aprile), a nove mesi (risposte utili nei mesi di aprile e gennaio successivo), a dodici (da gennaio a gennaio, o da aprile ad aprile), e infine a quindici (da gennaio ad aprile dell'anno successivo); quest'ultimo caso si presenta, tuttavia, con frequenza abbastanza ridotta. Le osservazioni longitudinali a nostra disposizione si riferiscono ai quarti osservati per la prima volta nel mese di gennaio 1985 e 1988 nella regione Lombardia.

Il punto di partenza è consistito nel classificare gli individui ad ogni

istante secondo la loro posizione sul mercato del lavoro. Le condizioni di occupazione, disoccupazione e inattività sono mutualmente esclusive e possono essere dedotta dalle risposte fornite al questionario della RTFL. In realtà perchè un soggetto sia classificato come disoccupato è necessario che sia senza lavoro, disposto a lavorare entro breve tempo nel caso riceva una proposta d'impiego, e che in un periodo di tempo precedente la rilevazione abbia cercato attivamente lavoro. La lunghezza di questo periodo determina le differenze fra gli schemi di classificazione possibili. In questo lavoro si sono seguite le raccomandazioni dall'ILO (1983), scegliendo un periodo di riferimento di 4 settimane. Questa scelta non è stata senza conseguenze; la brevità del periodo di riferimento, unitamente al fatto che i dati si riferiscono alla Lombardia, e che la procedura di *linkage* riduce nei dati longitudinali la percentuale di individui disoccupati rispetto alle osservazioni fornite dall'Istat (si veda Giusti, Marliani e Torelli (1987)), implica che il numero di individui osservati nel passaggio dall'occupazione verso la disoccupazione sia abbastanza ridotto, in particolar modo alle età adulte (oltre i 30 anni circa). Questo problema si riflette in particolare nella stima della matrice di covarianza delle stime in corrispondenza di queste età; inoltre, a partire dall'età di 56 anni, gli stati *D* e *I* sono stati aggregati.

E' indubbio che ogni campione contenga individui anche molto diversi fra loro; la mancata presa in considerazione di questa eterogeneità può avere conseguenze molto pesanti sulla validità delle stime dei parametri, e sulla capacità del modello di rappresentare adeguatamente la realtà osservata. Per tentare di ovviare, almeno parzialmente, a questo fenomeno è opportuno far dipendere le intensità di transizione da un certo insieme di variabili esplicative, che si suppone caratterizzino il comportamento dell'individuo. La strategia adottata in questo lavoro è stata di stratificare il campione secondo sesso ed età dei soggetti. Si tratta senz'altro di una maniera rudimentale di trattamento dell'eterogeneità; una parametrizzazione diretta delle intensità di transizione in funzione di sesso, età, e possibilmente titolo di studio (ci si limita qui alle informazioni disponibili dai dati RTFL) sarebbe probabilmente più convincente. La stratificazione, tuttavia, mantiene un certo numero di vantaggi, come quello di poter rapportare fra loro risultati relativi a classi sociali diverse (confronto donne-uomini, età giovani-età adulte), contenendo al minimo la complessità computazionale della procedura. Un'estensione diretta della procedura di Kalbfleisch e Lawless che permetta di parametrizzare le intensità di transizione in funzione delle variabili esplicative rappresenta una linea di sviluppo di questa ricerca.

La stratificazione del campione è stata operata rispetto al sesso e all'età. L'appartenenza ad una certa classe d'età è stata decisa osservando l'età raggiunta alla prima occasione d'inchiesta. Per chiarire questo punto, può essere utile ricordare che il processo di mobilità si evolve rispetto a due scale temporali che interagiscono fra loro: il tempo di calendario, che Arjas (1985) chiama il tempo reale, e l'età individuale. Si è fatta perciò l'ipotesi che ogni intensità di transizione, per un individuo di un dato sesso, sia costante all'interno di una classe d'età, ma variabile con la classe. Se si indica con *t* il tempo di calendario e con *e* l'età raggiunta, le intensità di transizione presentano la forma seguente:

$$(5.1) \quad q_{ij}(t,e) = q_{ij}(e) = q_{ijm} \quad \text{se } e \in [e_{m-1}, e_m)$$

dove $m=1,2,\dots,8$; $e_0=14$, $e_1=21$, $e_2=26$, $e_3=36$, $e_4=46$, $e_5=56$, $e_6=61$, $e_7=66$, $e_8=99$; e ogni q_{ijm} è una costante positiva o nulla.

La tabella 1 contiene le numerosità campionarie osservate in ogni strato; le cifre si riferiscono al *numero di transizioni*, e non al numero di individui.

V.2 I risultati

Le due specificazioni proposte - forma ridotta e forma strutturale - sono state stimate separatamente sui campioni del 1985 e 1988 descritti in precedenza; data la rassomiglianza di buona parte dei risultati, le tabelle raccolte in appendice si limitano in buona parte a quelli ottenuti sul primo insieme di dati. Eventuali punti di discordanza sono evidenziati nel testo.

La tabella 2 contiene le intensità di transizione relative alla forma ridotta non vincolata stimate sui differenti strati. Prima di analizzare i risultati di stima, tuttavia, è opportuno fare una considerazione preliminare. I due casi (donne 1985, classe 46-55; e uomini 1988, classe 36-45) la stima dell'intensità ($O \rightarrow D$, e $D \rightarrow I$ rispettivamente) è sulla frontiera della regione ammissibile, cioè zero. Questo è dovuto al fatto che queste transizioni sono osservate con frequenza nulla in tutte le matrici di transizione - a 3, 9, 12 e 15 mesi. In tutti gli altri casi, è stato sufficiente osservare anche una sola transizione, qualunque sia l'intervallo di riferimento, perchè la stima di massima verosimiglianza sia all'interno della regione ammissibile, anche se non sempre significativa.

In assenza di un'interpretazione strutturale, le uniche osservazioni che è lecito formulare a partire dai risultati ottenuti per la forma ridotta sono di carattere qualitativo, confrontando fra loro i valori stimati su strati diversi, e rapportandoli a quanto ci si sarebbe atteso a priori sulla base della teoria economica. Per cominciare, è possibile notare che alcune intensità di transizione sembrano evolvere con l'età in maniera abbastanza regolare. Per esempio, le intensità in uscita dall'inattività (sia $I \rightarrow O$ che $I \rightarrow D$) sono dapprima crescenti (fino alla terza classe d'età per gli uomini, alla seconda per le donne), e successivamente decrescono, fino alle ultime classi, per le quali l'uscita dalla forza lavoro è uno stato pressochè assorbente (dal quale, cioè, è impossibile l'uscita).

Andamento opposto sembra invece evidenziare l'intensità da O verso I , decrescente fino alla terza classe d'età, e in seguito crescente. L'intensità di transizione da O verso D decresce regolarmente, stabilizzandosi a partire dalla terza classe d'età su valori molto bassi (quasi sempre non significativamente diversi da zero). Queste evoluzioni con l'età sembrano adattarsi abbastanza bene con quanto era lecito attendersi a priori.

Il caso delle intensità di transizione in uscita dalla disoccupazione si presenta purtroppo di difficile interpretazione. Le stime dei parametri si evolvono in maniera irregolare rispetto alle età, ed in alcuni casi i valori stimati sembrano implausibili (per esempio, per gli uomini nel 1985, l'intensità $D \rightarrow O$ nella quinta classe d'età è di 104,458, che significherebbe una fortissima propensione, per gli individui disoccupati a quell'età, a transitare verso l'occupazione).

In realtà è abbastanza facile trovare una giustificazione a questo fenomeno. Con l'esclusione delle prime due classi di età, le transizioni osservate in uscita dalla disoccupazione sono in numero molto ridotto; di conseguenza i parametri relativi a questi passaggi non possono essere stimati con precisione; si noti che in generale la significatività di queste stime è piuttosto ridotta.

Nonostante queste precauzioni, un'ulteriore conferma della plausibilità dei valori stimati si deduce confrontando fra loro, a parità di classe d'età, i risultati relativi alle donne ed agli uomini. Le prime evidenziano ad ogni età una maggiore tendenza ad uscire dallo stato occupazione, soprattutto uscendo dalla forza lavoro (con l'eccezione della prima classe, sia nel 1985 che nel 1988). E' inoltre possibile notare come la spinta verso uno degli stati di attività sia molto più forte per gli uomini che per le donne. Anche le intensità in uscita da D , nonostante tutte le cautele dovute alle osservazioni fatte precedentemente, sembrano evidenziare una più forte tendenza all'uscita dalla forza lavoro per le donne che per gli uomini.

Se da un lato questi risultati si adattano bene alla conoscenza qualitativa che si ha dei meccanismi che operano sul mercato del lavoro (ed in effetti possono essere ottenuti anche con una formulazione in tempo discreto del modello di Markov), è comunque chiaro che un certo numero di importanti questioni interpretative resta comunque aperto. Per esempio, fino a che punto il divario uomo-donna è giustificabile sulla base di semplici considerazioni di preferenza individuale per uno stato anziché per l'altro? In che misura invece incidono le maggiori opportunità d'impiego che intuitivamente ci si aspetta siano aperte agli uomini, rispetto alle donne? Una possibile risposta a questi problemi può essere ottenuta analizzando i risultati della stima della forma strutturale. Per il momento, tuttavia, si considera il problema della verifica dell'adattamento del modello markoviano alle osservazioni.

Si consideri l'ipotesi nulla:

H_0 : Il modello è markoviano omogeneo sulla classe d'età considerata.

Dal punto di vista del comportamento individuale, quest'ipotesi significa che, nello strato campionario considerato, gli individui, nello scegliere lo stato occupazionale verso il quale transitare, si basano solo sullo stato attualmente occupato, e su nessun altro aspetto della loro precedente esperienza lavorativa (per esempio da quanto tempo essi sono nello stato attuale); inoltre che la probabilità condizionale di effettuare una transizione dipende dall'intervallo preso in considerazione solo attraverso la sua lunghezza (e non, per esempio, dalle date che delimitano l'intervallo).

Il test è basato sul confronto del numero di transizioni osservate e quelle previste dal modello; in termini più formali:

$$(5.3) \quad \hat{n}_{ij}(l) = n_i(l-1) \cdot \hat{p}_{ij}(w_l)$$

rappresentano le transizioni previste ($\hat{p}(w_l) = \exp(Qw_l)$), raccolte nella matrice $\hat{N}(l)$, mentre quelle osservate sono contenute nella matrice $N(l)$, dove $l=1,2,\dots,L$. Le metriche utilizzate per confrontare $N(l)$ e $\hat{N}(l)$ sono di due tipi:

a) un test basato sul rapporto di verosimiglianza

$$(5.4) \quad \Lambda = 2 \sum_{l=1}^L \sum_{j=1}^K n_{lj}(1) \cdot \log \left[\frac{n_{lj}(1)}{\hat{n}_{lj}(1)} \right]$$

b) e la statistica test di Pearson

$$(5.5) \quad \chi^2 = \sum_{l=1}^L \sum_{j=1}^K \frac{(n_{lj}(1) - \hat{n}_{lj}(1))^2}{n_{lj}(1)}$$

Sotto H_0 (L fisso, $n \rightarrow \infty$), entrambe sono asintoticamente distribuite secondo un χ^2_ν , dove:

$$(5.6) \quad \nu = \sum_{l=1}^L \left\{ K(K-1) - S_l \right\} - b$$

e S_l rappresenta il numero di caselle vuote nella matrice di transizione l , mentre b è il numero di parametri indipendenti in \mathbf{Q} . Si noti che, sotto H_0 , le due statistiche test sono equivalenti.

I risultati relativi a questa verifica d'ipotesi sono contenuti in tabella 3. Le conclusioni che se ne deducono sono parzialmente confortanti: ad un livello di significatività del 5%, l'ipotesi nulla è rigettata in 9 casi su 32 complessivi, e in un caso la decisione è diversa a seconda della statistica test prescelta. Tuttavia, in molti casi il valore della statistica test è elevato, benchè non più grande della soglia critica, e ciò non depone a favore del modello markoviano.

Non esistono classi d'età in cui l'ipotesi di Markov sembra adattarsi particolarmente bene, qualunque sia l'anno o il sesso. Per le donne, nel 1985 H_0 è accettata sulle prime tre classi d'età, mentre essa viene rifiutata nel 1988. Nel complesso, i risultati sono buoni solo per gli uomini nel 1988, a tutte le età.

Le cause di questo mancato adattamento possono essere diverse, alcune legate alle ipotesi sottostanti il modello e non verificate dai dati, ed altre più direttamente collegate alla natura stessa dei dati; alcuni possibili motivi di errata specificazione sono:

i) sebbene le matrici di transizione osservate contengano un numero abbastanza elevato di osservazioni, queste, in corrispondenza di alcune classi d'età, sono mal distribuite fra le celle delle matrici, e non permettono una stima precisa di una parte importante dei parametri (quelli relativi allo stato D), dando luogo a dei problemi di natura essenzialmente numerica;

ii) i dati possono non verificare la proprietà di Markov, cioè l'indipendenza della storia futura dal passato condizionatamente allo stato

occupato nel presente.

iii) il processo di mobilità fra stati può non essere omogeneo nel tempo, il che implicherebbe intensità di transizione funzione del tempo di calendario t (ciò non significa tuttavia che esista una rappresentazione generale, per questo tipo di processi, del tipo $p(s,t) = \exp(Q(s,t))$; l'ipotesi di omogeneità nel tempo era infatti determinante nella sua derivazione, ottenuta nella prima parte di questo lavoro). Un chiaro motivo di non omogeneità è la frequenza trimestrale dei dati, e la presenza di stagionalità che ne deriva. E' infatti chiaro che, in presenza di stagionalità, la struttura delle matrici di transizione a tre mesi è diversa da quelle a nove e a dodici mesi; in altre parole, non tenendone conto si cerca di riassumere con un solo modello strutture che, in linea di principio, potrebbero essere anche molto diverse fra loro. Come parziale attenuante segnaliamo che la presenza di stagionalità può non essere molto forte in regioni, come la Lombardia, in cui il peso dell'agricoltura ed altri settori a forte stagionalità non è preponderante nell'occupazione complessiva. Inoltre studi precedenti (si veda Bernardi e Zaccarin (1984)) hanno evidenziato che riducendo il numero di stati considerati, viene ridotta l'importanza di oscillazione stagionali nei flussi fra gli stati stessi. E' tuttavia indubbio che, nel caso fossero disponibili dati relativi a transizioni su un anno e in numero adeguatamente elevato da permettere di tener conto dell'eterogeneità individuale, questi sarebbero da preferire a quelli utilizzati in questo contesto.

iv) il campione può contenere dell'eterogeneità individuale, non adeguatamente sintetizzata dalle variabili sesso ed classe d'età. Si tratta di un'eventualità estremamente probabile, quando si pensi che non è stata considerata una variabile indubbiamente critica come il livello di studi. Tuttavia, esiste una estensione della procedura di Kalbfleisch e Lawless che permette di tener conto in maniera migliore dell'eterogeneità individuale. Senza entrare in grande dettaglio, osserviamo che perchè essa possa essere applicata è necessario che il vettore di variabili esplicative, che si suppone caratterizzare completamente l'individuo, assuma valori in un insieme discreto e con un numero di modalità finito e possibilmente limitato. Un tentativo in questa direzione costituisce una delle linee di sviluppo naturali di questo lavoro.

Le tabelle 4, 5 e 6 contengono statistiche ausiliarie, immediatamente derivate dalla stima della matrice Q . La tabella 5 presenta le durate medie di soggiorno negli stati secondo sesso, età ed anno di rilevazione; la tabella 6 le distribuzioni stazionarie di occupazione degli stati, secondo sesso, età e anno di rilevazione. A partire da queste sono calcolati i tassi di disoccupazione limite previsti dal modello, raccolti in tabella 4; la formula per il calcolo è data dal rapporto:

$$(5.7) \quad \frac{\hat{\rho}_{Dm}}{\hat{\rho}_{Om} + \hat{\rho}_{Dm}}$$

dove m rappresenta l'indicatore dello stato, e $\hat{\rho}$ è il vettore di probabilità limite stimato.

Le durate attese di soggiorno confermano le considerazioni qualitative formulate a proposito alle intensità di transizione stimate, con in più un'intuizione supplementare: pressochè ovunque, le durate di soggiorno in O aumentano passando dal 1985 al 1988, e viceversa per le durate in D e in I . Questo risultato concorda con l'opinione diffusa che il 1985 rappresentasse per la Lombardia un anno di crisi, che nel 1988 era stata ormai riassorbita.

Queste osservazioni restano valide per i risultati di tabella 4 e 6. In mancanza di dati esatti sulla frequenza osservata di occupazione degli stati, considerazioni ulteriori sulle previsioni del modello di Markov sono comunque rimandate ad un lavoro successivo.

I risultati di stima della forma strutturale sono riassunti nella tabella 7. Per quel che riguarda i λ_i , si noti che in generale λ_D è più elevato per gli uomini che per le donne, così come λ_I ; il contrario è vero per λ_O . In altre parole, la differenza di queste stime secondo il sesso rappresenta una misura, per quanto limitata, del divario di opportunità d'impiego per donne ed uomini.

Anche i π_j sembrano concordi con quanto atteso. In generale, ed in particolare alle età più giovani, gli uomini preferiscono occupazione e disoccupazione all'uscita dalla forza lavoro. Le differenze sembrano indubbiamente significative, anche se è opportuno ricordare che alle età adulte il numero limitato di osservazioni in uscita dalla disoccupazione impedisce di fatto una stima precisa dei parametri relativi a questo stato.

Si osservi che i π_j non rappresentano le probabilità di scelta statiche; per costruzione, infatti, essi non sono esenti da considerazioni dinamiche, legate al differenziale esistente fra i tassi di arrivo dell'informazione nei vari stati. Olsen, Smith e Farkas (1986) definiscono le probabilità di scelta statica degli stati in un contesto del tutto analogo a questo.

Un confronto fra i risultati relativi al 1985 e al 1988 suggerisce la possibilità di uno studio dell'evoluzione dei flussi fra gli stati in funzione, per esempio, di un indicatore della congiuntura, utilizzato per parametrizzare le osservazioni provenienti da anni diversi. Per questa via sarebbero inoltre possibili esercizi di statica comparata, con risultati analoghi al "moltiplicatore" nel modello di regressione univariato o multivariato classico.

Come ricordato in precedenza, con 3 stati la forma strutturale implica la presenza di un vincolo di sovraidentificazione, che è possibile sottoporre a verifica empirica mediante un semplice test del rapporto di verosimiglianza. I risultati del test sono presentati nella tabella 8, e forniscono una evidenza empirica a favore della forma strutturale. Complessivamente, ad un livello di significatività del 5%, l'ipotesi nulla $H_0: q_{ij} = \lambda_i \cdot \pi_j$ è rigettata solo in due casi su venti, e la statistica test ha molto spesso valore prossimo a zero.

Un ulteriore esercizio consiste nel simulare delle traiettorie individuali generate aleatoriamente sulla base dei parametri stimati. All'età di 14 anni ogni individuo si trova fuori dalla forza lavoro; la sua

traiettorie (la successione di passaggi da uno stato all'altro) viene costruita ricordando le proprietà dei processi di Markov omogenei ricordate nel Teorema della sezione II. Per ogni sesso, 10000 traiettorie indipendenti vengono simulate, troncandole all'età di 70 anni. E' così possibile calcolare una serie di statistiche che rappresentano la situazione del mercato del lavoro in Lombardia nel 1985, come per esempio la durata media cumulata di soggiorno in ognuno degli stati, il numero medio di episodi di soggiorno, la durata media di un singolo episodio, ed i rispettivi scarti quadratici medi empirici. I risultati sono contenuti nella tabella 9.

VI CONSIDERAZIONI CONCLUSIVE

Come sottolineato in più punti, la natura di questo lavoro è essenzialmente di esplorazione preliminare dei dati, e delle possibilità di specificazione alternative che la struttura statistica dei processi di Markov omogenei consente. Anziché ricordare una volta ancora i risultati conseguiti, perciò, è forse preferibile concludere il lavoro delineando le linee di sviluppo futuro di questa ricerca.

Uno dei problemi principali dello studio dei dati di *panel* è senza dubbio quello di una adeguata presa in conto dell'eterogeneità presente fra gli individui. Le complesse procedure di stima di questo tipo di modelli rendono difficile l'introduzione diretta delle variabili esplicative nelle intensità di transizione; ed è per questa ragione che, per questo lavoro, si è scelta la strada più semplice, di stratificare il campione secondo due delle possibili variabili caratterizzanti il comportamento dell'individuo.

Accanto alla sua estrema semplicità, questa procedura contiene alcuni difetti rilevanti. Per esempio, è intuitivamente chiaro che sesso ed età non esauriscono la lista delle variabili che caratterizzano ogni individuo. Sarebbe perciò importante tentare di stimare il modello parametrizzando le intensità direttamente in funzione delle variabili esplicative, per esempio ponendo $q_{ij}(x) = \exp(x'\beta_{ij})$ dove il vettore x assume valori in uno spazio di modalità discrete e di numerosità finita (possibilmente anche limitata). In assenza di informazioni di qualsiasi tipo sul reddito o sulla ricchezza degli individui, sembra in particolare importante includere in x , oltre all'età e al sesso, il livello di studi e variabili demografiche o di stato civile (numero di figli, sposato o no, ecc...). Si è inoltre fatto un breve accenno all'eventuale presenza di stagionalità nei dati, e di conseguenza di non omogeneità del processo di mobilità rispetto al tempo.

Sembra importante osservare inoltre che le caratteristiche stesse dei dati pongano alcuni vincoli importanti all'applicazione di questi modelli. Si è ricordati che, per una serie di cause concomitanti, la frequenza di osservazioni in entrata ed in uscita dalla disoccupazione è estremamente ridotta oltre i 30 anni; aldilà di quest'età le stime dei parametri relativi alla disoccupazione sono inaffidabili, e l'aggregazione degli stati di disoccupazione e inattività è praticamente obbligatoria, per quanto dolorosa da un punto di vista economico.

L'adattamento ai dati evidenziato dal modello di Markov, sia nella sua forma strutturale che in quella ridotta, ha mostrato un certo numero di imperfezioni, talvolta abbastanza gravi. Una conoscenza più approfondita delle

cause di questo comportamento non può essere acquisita senza analizzare più in profondità la plausibilità di ciascuna delle ipotesi sottostanti il modello statistico (si veda la sezione V).

BIBLIOGRAFIA

- Akerlof, G. e B. Main (1980): "Unemployment Spells and Unemployment Experience", *American Economic Review*, n. 70, pag. 885-893.
- Akerlof, G. e B. Main (1981): "An Experience-Weighted Measure of Unemployment and Unemployment Durations", *American Economic Review*, n. 71, pag. 1003-1011.
- Anderson, T. W. e L. A. Goodman (1957): "Statistical Inference about Markov Chains", *Annals of Mathematical Statistics*, vol. 28, pag. 89-109.
- Arjas, E. (1985): "Stanford Heart Transplantation Data Revisited: a Real Time Approach", in *Modern Statistical Methods in Chronic Disease Epidemiology* (a cura di S. H. Moolgavkar e R. L. Prentice), New York, Wiley.
- Bernardi, L. e Zaccarin, S. (1984): "Indicatori di mobilità: applicazione di un modello markoviano ai dati della rilevazione trimestrale delle FL", *Economia e Lavoro*, XVIII, n. 4, pag. 91-103.
- Bhattacharya, R. N. e E. C. Waymire (1990): *Stochastic Processes with Applications*, New York, Wiley.
- Burdett, K., N. M. Kiefer, D. T. Mortensen e G. R. Neumann (1984): "Earnings, Unemployment and the Allocation of Time over Time", *Review of Economic Studies*, vol. 51, pagg. 559-578.
- Clark, K. e L. Summers (1979): "Labor Market Dynamics and Unemployment: a Reconsideration", *Brookings Papers on Economic Activity*, n. 1, pag. 13-60.
- Cox, D. R. e H. D. Miller (1965): *The Theory of Stochastic Processes*, Londra, Methuen & Co. .
- Doob, J. L. (1953): *Stochastic Processes*, New York, Wiley.
- Fabbris, L. e L. Bernardi (1986): "Forze di lavoro: disegno dell'indagine e analisi strutturali", Nota interna, Dipartimento di Scienze Statistiche, Università di Padova.
- Flinn, C. e J. Heckman (1982): "New Methods for Analysing Structural Models of Labor Force Dynamics", *Journal of Econometrics*, vol. 18, pag. 115-168.
- Flinn, C. e J. Heckman (1983): "Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?", *Journal of Labor Economics*, n. 1, pag. 28-42.
- Fougère, D. e T. Kamionka (1990): "Un modèle markovien du marché du travail", Cahier n. 9022, GREMAQ, Université de Toulouse.
- Geweke, J., R. C. Marshall e G. A. Zarkin (1986a): "Exact Inference for Continuous Time Markov Chains Models", *Review of Economic Studies*, LIII, Pag. 653-669.
- Geweke, J., R. C. Marshall e G. A. Zarkin (1986b): "Mobility Indices in Continuous Time Markov Chains", *Econometrica*, vol. 54, n. 6, pag. 1407-1423.
- Giusti, A., G. Marliani e N. Torelli (1987): "L'abbinamento dei dati individuali di successive rilevazioni nell'indagine trimestrale delle forze di lavoro", Rapporto di ricerca n. 2, Dipartimento di Scienze

Statistiche, Università di Padova.

- Giusti, A., G. Marliani e N. Torelli** (1988): "L'uso di valutazioni probabilistiche per il collegamento di informazioni provenienti da indagini diverse", Rapporto di ricerca n. 10, Dipartimento di Scienze Statistiche, Università di Padova.
- Hall, R. E.** (1972): "Turnover in the Labor Force", *Brookings Papers on Economic Activity*, n. 3, pag. 709-756.
- Heckman, J. e G. J. Borjas** (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers for a Continuous-Time Model of Heterogeneity and State Dependence", *Economica*, n. 47, pag. 247-283.
- ILO** (1983): "Resolution Concerning Statistics of The Economically Active Population, Employment, Unemployment and Underemployment", (Resolution I, 13th International Conference of Labour Statisticians, Geneva, 18-29 October 1982), *Bullettin of Labour Statistics*, 3, pagg. IX-XV.
- Jackman, R. e R. Layard** (1991): "Does Long-term Unemployment Reduce a Person's Chance of a Job? A time-series Test", *Economica*, vol. 58, pag. 93-106.
- Kalbfleisch, J. D. e J. F. Lawless** (1985): "The Analysis of Panel Data under a Markov Assumption", *Journal of the American Statistical Association*, vol. 80, n. 392, pag. 863-871.
- Magnac, T. e J.-M. Robin** (1990): "Econometric Analysis of The Transition Between Self-Employment and Wage-Work", Document n. 90-09, Delta, Parigi.
- Mortensen, D. T. e G. R. Neumann** (1984): "Choice or Chance? A Structural Interpretation of Individual Labor Market Histories", in *Studies in Labor Market Dynamics*, a cura di G. R. Neumann e N. C. Westegard-Nielsen, Springer-Verlag, Berlino.
- Narendranathan, W., S. Nickell e J. Stern** (1985): "Unemployment Benefit Revisited", *Economic Journal*, n. 95, pag. 307-329.
- Olsen, R. J., D. A. Smith e G. Farkas** (1986): "Structural and Reduced-Form Models of Choice Among Alternatives in Continuous Time: Youth Employment Under a Guaranteed Jobs Program", *Econometrica*, vol. 54, n. 2, pagg. 375-394.
- Pastorello, S.** (1992): "La mobilità nel mercato del lavoro. Un'analisi econometrica a partire da osservazioni in tempo discreto", Tesi per il conseguimento del Dottorato di Ricerca in Economia Politica, Università di Bologna.
- Poterba, J. M. e L. H. Summers** (1986): "Reporting Errors and Labor Market Dynamics", *Econometrica*, vol. 54, n. 6, pag. 1319-1338.
- Salant, S.** (1977): "Search Theory and Duration Data: A theory of Sorts", *Quarterly Journal of Economics*, vol. 91 (feb.), pag. 39-57.
- Singer, B. e S. Spilerman** (1976): "'The Representation of Social Processes by Markov Models", *American Journal of Sociology*, vol. 82, n. 1, pagg. 1-54.
- Toikka, R. S.** (1976): "A Markovian Model of Labor Market Decisions by Workers", *American Economic Review*, n. 66, December, pagg. 821-834.

TABELLA 1

**Numerosità Campionarie
secondo sesso e classe d'età**

Anno Sesso Classe d'età	1985		1988	
	Uomini	Donne	Uomini	Donne
14-20	1604	1537	1707	1498
21-25	1485	1398	1546	1560
26-35	2217	2124	2245	2189
36-45	2466	2425	2337	2398
46-55	2284	2390	2420	2571
56-60	982	770	975	1082
61-65	840	924	738	905
> 66	1383	2308	1243	2088
Totali	13261	13876	13211	14291
	27137		27502	

TABELLA 2

Intensità di transizione degli uomini per classe d'età - anno 1985

Transizione Età	O→D	O→I	D→O	D→I	I→O	I→D
14-20	1.265 (1.754)	11.272 (7.304)	30.397 (4.102)	30.550 (4.04)	5.323 (7.080)	4.485 (5.642)
21-25	1.782 (3.380)	2.456 (4.791)	43.304 (5.384)	25.144 (3.800)	11.090 (5.972)	10.033 (4.832)
26-35	0.254 (1.871)	0.393 (2.501)	40.026 (2.553)	31.555 (2.061)	38.679 (4.953)	12.628 (2.102)
36-45	0.191 (2.054)	0.401 (3.054)	14.029 (2.031)	9.346 (1.467)	23.298 (3.584)	11.068 (2.240)
46-55	0.741 (1.924)	1.472 (4.800)	104.458 (1.711)	83.781 (1.540)	5.068 (3.579)	1.668 (1.232)
56-60	-	7.405 (7.220)	-	-	4.762 (5.155)	-
61-65	-	8.535 (4.931)	-	-	1.584 (3.581)	-
> 66	-	24.867 (5.472)	-	-	2.199 (5.470)	-

Intensità di transizione delle donne per classe d'età - anno 1985

Transizione Età	O→D	O→I	D→O	D→I	I→O	I→D
14-20	4.267 (3.290)	4.992 (4.020)	23.984 (4.999)	20.188 (4.570)	2.484 (4.850)	4.054 (6.058)
21-25	0.782 (2.003)	4.372 (6.200)	22.315 (4.362)	26.685 (4.530)	8.628 (6.248)	8.762 (5.283)
26-35	0.608 (2.482)	2.200 (5.729)	12.811 (1.868)	46.352 (3.697)	6.922 (8.148)	1.729 (3.077)
36-45	0.285 (1.355)	4.346 (7.585)	13.678 (1.203)	64.891 (2.705)	3.906 (8.231)	0.982 (2.791)
46-55	0.000 (0.000)	7.214 (8.569)	47.160 (1.617)	41.399 (1.537)	2.127 (6.746)	0.782 (0.955)
56-60	-	13.899 (4.952)	-	-	1.712 (3.941)	-
61-65	-	20.660 (3.901)	-	-	0.870 (2.966)	-
> 66	-	47.119 (4.000)	-	-	0.593 (2.431)	-

Legenda: O=occupazione; D=disoccupazione; I=non forza lavoro.

NB: I valori delle intensità sono moltiplicati per 10^4 . T-ratio fra parentesi.
Nelle ultime tre classi d'età gli stati D e I sono stati raggruppati.

TABELLA 3

Valori delle Statistiche Test per Sesso ed Età - Anno 1985

Età Test	14-20	21-25	26-35	36-45	46-55	56-60	61-65	> 66
Uomini								
Λ	15.53*	32.23	18.92*	19.96*	10.65*	6.40*	12.99	49.21
χ^2	16.51*	34.23	21.70	16.06*	9.21*	6.21*	11.08	73.00
gdl	12	16	12	16	12	4	5	5
Donne								
Λ	20.58*	15.60*	16.03*	27.94	9.58*	12.80	3.54*	6.47*
χ^2	19.57*	13.00*	15.26*	31.90	8.17*	13.59	4.45*	7.14*
gdl	12	12	14	14	10	4	4	4

Valori delle Statistiche Test per Sesso ed Età - Anno 1988

Età Test	14-20	21-25	26-35	36-45	46-55	56-60	61-65	> 66
Uomini								
Λ	21.77*	15.80*	11.34*	13.42*	14.32*	1.93*	1.90*	8.02*
χ^2	20.28*	15.53*	11.31*	11.84*	12.85*	1.95*	1.35*	6.88*
gdl	14	12	10	10	10	4	4	6
Donne								
Λ	27.86	36.91	23.36	12.76*	17.00*	5.13*	3.23*	12.67
χ^2	27.83	36.28	23.33	12.66*	15.08*	4.08*	3.02*	24.83
gdl	12	14	10	10	10	4	3	5

NB: Λ =test del rapporto di verosimiglianza; χ^2 =test chi-quadrato; gdl=gradi di libertà della distribuzione asintotica; * = H_0 non è rigettata ad un livello di significatività del 5%.

TABELLA 4

Tasso di Disoccupazione Stazionario per Sesso e Classe d'Età
Anno 1985

Sesso	Età	14-20	21-25	26-35	36-45	46-55	56-60	61-65	> 66
Uomini		12.08	4.96	0.54	1.54	0.65	-	-	-
Donne		18.87	7.74	2.05	1.72	0.30	-	-	-

TABELLA 5

Durate Medie di Soggiorno negli Stati secondo Sesso ed Età - Anno 1985

Età Stato	14-20	21-25	26-35	36-45	46-55	56-60	61-65	> 66
Occupazione Uomini	795.71 (7.90)	2352.54 (6.65)	15467.1 (3.66)	16905.6 (4.06)	4519.08 (6.08)	1350.41 (7.22)	1171.62 (4.93)	402.14 (5.47)
Donne	1080.04 (5.70)	1940.46 (7.12)	3562.17 (6.95)	2159.27 (8.06)	1385.54 (8.57)	719.47 (4.95)	484.02 (3.90)	212.23 (4.00)
Disoccupaz. Uomini	164.08 (6.08)	146.10 (6.68)	139.70 (3.58)	467.81 (2.74)	53.12 (2.17)	-	-	-
Donne	226.39 (6.92)	204.08 (6.55)	169.02 (4.32)	127.28 (3.03)	112.92 (2.22)	-	-	-
N. F. L. Uomini	1019.53 (10.36)	473.43 (9.30)	194.90 (5.62)	290.98 (4.36)	1484.56 (4.39)	2100.16 (5.16)	6312.28 (3.58)	4547.03 (5.47)
Donne	1529.61 (9.13)	575.06 (9.00)	1155.96 (8.75)	2045.86 (8.57)	4534.71 (6.89)	5840.08 (3.94)	11497.2 (2.97)	16874.9 (3.43)

NB: N.F.L.=Non Forze di Lavoro; le cifre si intendono espresse in giorni; t-ratio fra parentesi; nelle ultime tre classi d'età gli stati D e I sono stati raggruppati.

TABELLA 6

Distribuzioni Stazionarie secondo il Sesso e la Classe d'Età - Anno 1985

Età Stato	14-20	21-25	26-35	36-45	46-55	56-60	61-65	> 66
Occupazione Uomini	36.84 (10.90)	81.26 (32.20)	98.37 (213.3)	96.93 (102.6)	76.50 (17.36)	39.14 (7.32)	15.66 (3.58)	8.12 (4.82)
Donne	36.84 (7.84)	68.30 (19.06)	72.06 (20.35)	46.70 (11.56)	23.09 (7.29)	10.97 (3.73)	4.04 (2.66)	1.24 (3.16)
Disoccupaz. Uomini	5.06 (6.01)	4.24 (5.54)	0.54 (2.51)	1.52 (2.34)	0.50 (2.65)	-	-	-
Donne	8.57 (6.36)	5.73 (5.42)	1.51 (3.62)	0.82 (3.07)	0.07 (0.96)	-	-	-
N. F. L. Uomini	58.10 (15.99)	14.50 (4.80)	1.08 (1.72)	1.54 (1.00)	22.99 (5.19)	60.86 (11.38)	84.34 (19.27)	91.87 (54.54)
Donne	54.59 (10.76)	25.97 (6.19)	26.43 (7.13)	52.48 (12.82)	76.84 (24.24)	89.03 (30.30)	95.96 (63.28)	98.76 (251.2)

NB: N.F.L.= Non Forze di Lavoro; le cifre si intendono espresse in termini percentuali; t-ratio fra parentesi; nelle ultime tre classi d'età gli stati D e I sono stati raggruppati.

TABELLA 7

Stime dei Parametri di Forma Strutturale per sesso ed età anno 1985

Parametro Età	λ_0	λ_D	λ_I	π_0	π_D	π_I
14-20						
Uomini	18.078 (6.456)	75.713 (5.386)	20.906 (5.350)	29.14 (5.942)	17.99 (5.157)	52.87 (7.864)
Donne	13.315 (5.199)	72.778 (4.982)	9.474 (7.108)	29.67 (7.269)	21.05 (6.700)	30.95 (5.012)
21-25						
Uomini	8.704 (5.528)	93.337 (5.148)	32.384 (7.095)	40.22 (7.952)	26.36 (5.898)	33.42 (5.974)
Donne	7.180 (6.216)	62.303 (5.732)	35.743 (5.189)	27.57 (6.120)	21.05 (5.142)	51.38 (7.728)
26-35						
Uomini	1.262 (2.773)	86.902 (3.157)	77.672 (3.573)	48.71 (4.437)	17.59 (2.813)	33.70 (2.995)
Donne	4.021 (5.353)	65.813 (4.105)	21.801 (3.653)	30.10 (3.958)	9.78 (3.346)	60.13 (6.533)
36-45						
Uomini	1.048 (3.935)	29.226 (2.554)	54.829 (2.925)	43.70 (3.492)	19.38 (3.136)	36.93 (2.925)
Donne	5.669 (5.938)	82.539 (2.954)	21.307 (2.020)	18.25 (2.075)	4.69 (1.992)	77.06 (7.249)
46-55						
Uomini	4.229 (2.860)	226.912 (1.864)	10.431 (3.138)	47.62 (3.223)	17.09 (1.953)	35.29 (3.073)
Donne	13.386 (2.670)	89.530 (2.207)	5.116 (2.041)	42.46 (2.126)	0.72 (0.958)	56.81 (2.809)

NB: i λ sono moltiplicati per 10^4 ; i π sono moltiplicati per 100; t-ratio fra parentesi.

TABELLA 8

Test di Sovraidentificazione - Anno 1985

Età Modello	14-20	21-25	26-35	36-45	46-55
Uomini -2 Δ L	8.55	1.72*	0.14*	0.10*	0.02*
Donne -2 Δ L	1.85*	5.36	2.82*	0.03*	1.93*
gdl	1	1	1	1	1

NB: -2 Δ L=statistica test, pari a -2 volte il logaritmo del rapporto fra verosimiglianza vincolata e non vincolata; sotto H_0 , $-2\Delta L \rightarrow \chi^2_{gdl}$; * significa che H_0 è rigettata ad un livello di significatività del 5%.

TABELLA 9

Simulazione di traiettorie individuali fra 14 e 70 anni
 Forma strutturale - 10000 replicazioni
 anno 1985

Stato Statistica	Occupazione	Disoccupazione	Inattività
Uomini			
Durate ⁽¹⁾	37.828 (7.420)	0.974 (0.781)	17.196 (3.441)
Periodi ⁽²⁾	4.261 (2.621)	2.213 (1.531)	5.432 (3.139)
Episodio ⁽³⁾	8.879	0.440	3.165
Donne			
Durate	21.222 (3.025)	1.174 (1.062)	33.603 (4.846)
Periodi	4.290 (2.730)	2.155 (1.528)	5.967 (3.347)
Episodio	4.947	0.545	5.631

Legenda: (1) durata media di soggiorno negli stati in anni;
 (2) numero medio di episodi di soggiorno negli stati;
 (3) durata media di un episodio di soggiorno negli stati in anni.
 Fra parentesi gli scarti quadratici medi empirici.