

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

**Modelli Statistici per l'Analisi
dei Comportamenti
di Transizione Scuola Lavoro**

Antonella D'Agostino, Giulio Ghellini, Laura Neri
Dip. di Metodi Quantitativi, Univ. di Siena

Working Paper n. 22 marzo 2000

Unità locali del progetto:

Dip. di Economia Politica, Univ. di Modena	(coord. Michele Lalla)
Dip. di Economia 'S. Cagnetti De Martiis', Univ. di Torino	(coord. Bruno Contini)
Dip. di Statistica, Univ. di Venezia	(coord. Tommaso Di Fonzo)
Dip. di Metodi Quantitativi, Univ. di Siena	(coord. Achille Lemmi)
Dip. di Scienze Statistiche, Univ. di Padova	(coord. Ugo Trivellato)

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

1 Premessa¹

La misura dinamica dei comportamenti individuali sul mercato del lavoro rappresenta indubbiamente una delle frontiere di analisi statistica di maggiore interesse di questi ultimi anni. Tale rilevante attualità può essere considerata da due distinti punti di vista: i) da un lato i profondi cambiamenti dei processi industriali di produzione di beni e servizi hanno sensibilmente modificato la natura stessa della domanda di lavoro, sempre più attenta alla flessibilità della forza lavoro, alla sua capacità di adattamento a contenuti lavorativi in costante evoluzione; ii) dall'altro i crescenti livelli di formazione scolastica, le accresciute possibilità di rimandare nel tempo l'inizio della vita lavorativa, e conseguentemente il momento di distacco dal proprio *milieu* familiare, hanno indubbiamente innalzato le aspettative lavorative delle giovani generazioni. La conseguenza più evidente di tale situazione è il perdurante fenomeno di elevati tassi di disoccupazione giovanile, fenomeno che caratterizza (in modo più o meno accentuato) tutti i paesi dell'Unione Europea in questa delicata fase del suo sviluppo economico e sociale. Indubbiamente, nel nostro Paese, la disoccupazione giovanile si presenta particolarmente accentuata, seppur con rilevanti differenze territoriali, ed appare quindi di estremo interesse individuare pertinenti basi di dati di tipo longitudinale su tali fenomeni, le uniche in grado di consentire analisi statistiche dei comportamenti delle giovani generazioni a livello micro, ad esempio analisi per valutare le componenti di volontarietà/involontarietà dello stato di disoccupazione.

Si deve inoltre tenere presente che la difficile fase di transizione dal mondo scolastico alla vita attiva, oggetto specifico di questa ricerca, rappresenta una delle fasi più delicate dello sviluppo degli individui e del loro percorso verso la vita adulta, in quanto successi e/o insuccessi in tale periodo sembrano rappresentare eventi in grado di condizionare in modo rilevante la loro futura vita, non solo sotto il profilo lavorativo.

In questo lavoro, le dinamiche di transizione scuola lavoro sono studiate utilizzando la base di dati longitudinali sui giovani residenti in Lombardia creata nell'ambito del progetto LEVA [progetto Longitudinale sull'Entrata nella Vita Attiva] (vedi successivo par. 2 per dettagli in merito) e applicando i modelli per dati longitudinali che vanno sotto il nome di "logit sequenziali" (vedi par. 3), modelli che permettono di analizzare come le scelte individuali

¹Questo articolo è il risultato degli sforzi comuni di tutti gli autori, A. D'Agostino ha scritto le sezioni 2, 4 e 6, L. Neri le sezioni 3 e 5.1, e G. Ghellini le sezioni 1 e 5.2.

attraverso il tempo possano essere influenzate anche dalle scelte effettuate nei tempi precedenti². L'analisi empirica, condotta sulla base di dati retrospettivi raccolti con il progetto LEVA, affronta sia le problematiche connesse alla scelta di continuare o meno il proprio iter scolastico negli anni successivi al termine del ciclo dell'obbligo (par. 4), sia le diverse opportunità di lavoro che caratterizzano diversi segmenti giovanili identificati sia sulla base di variabili individuali sia relativamente al momento in cui hanno lasciato la scuola (par. 5), con risultati a nostro avviso utili sia sotto un profilo conoscitivo, sia sotto quello maggiormente connesso alla definizione di efficaci politiche atte a contrastare la disoccupazione giovanile.

2 L'indagine LEVA

In Italia, uno dei primi disegni di indagine longitudinale a livello locale è stato avviato in via sperimentale dal Servizio Statistica della Regione Lombardia nel 1986.

Si tratta del progetto "Longitudinale sull'Entrata nella Vita Attiva", detto sinteticamente progetto LEVA, volto a seguire i percorsi scolastici lavorativi di giovani nei sei anni e mezzo che seguono la frequenza dell'ultimo anno della scuola media inferiore.

L'obiettivo centrale di questo articolato progetto è di mettere a punto una rilevazione *panel* su coorti di iscritti alla terza media da seguire per circa sette anni, allo scopo di analizzare i comportamenti delle giovani leve lombarde a partire dalla loro uscita dal ciclo formativo dell'obbligo.

A partire dal 1986 fino al 1992, ogni tre anni, viene sottoposta un'intervista in classe ad una nuova coorte di studenti che stanno per terminare gli studi dell'obbligo (più o meno all'età di quattordici anni). Durante l'intervista si raccolgono informazioni riguardo la situazione attuale dello studente, i suoi obiettivi futuri e la famiglia in cui essi vivono.

Le sei successive occasioni di indagine sono pianificate in modo tale da fornire informazioni anno per anno. Le prime cinque interviste sono postali e caratterizzate per la loro uniformità e simmetria. Il loro principale obiettivo è quello di catturare lo stato scolastico-lavorativo in cui gli intervistati si trovano, con sintetiche informazioni sui risultati ottenuti a scuola, il tipo

²Si coglie l'occasione per ringraziare il Servizio Statistica della Giunta Regionale della Lombardia per la messa a disposizione dei dati di base del progetto LEVA, necessari a questo lavoro.

di lavoro svolto ecc.. L'intervista finale è invece un'intervista faccia a faccia ed è organizzata in modo da avere sia informazioni sulla situazione attuale, sia sui cambiamenti avvenuti durante i precedenti sette anni di indagine, ricostruendo retrospettivamente i percorsi intrapresi dopo l'uscita dalla scuola dell'obbligo.

Quest'ultima informazione permette di ottenere una completa panoramica delle transizioni individuali dalla scuola alla vita attiva durante i sette anni, circa, di riferimento dell'indagine. Al contrario, la rilevazione panel non fornisce una così dettagliata informazione, sia per problemi dovuti all'*attrition*, sia per disegno, in quanto sono previste, in alcune particolari occasioni di indagine, interviste a specifici sottocampioni (per informazioni più dettagliate sull'indagine si veda Ghellini, 1987, 1989, 1990, 1992; Bernardi e Ghellini, 1992).

Al fine di studiare i percorsi dalla scuola al mondo del lavoro sembra, pertanto, estremamente interessante utilizzare l'informazione retrospettiva. E' comunque opportuno sottolineare che anche i dati retrospettivi possono presentare alcuni problemi, principalmente legati al ricordo. Da precedenti analisi condotte sugli stessi dati si è però mostrato che l'effetto memoria è limitato. In particolare, è stato messo in evidenza (D'Agostino e Ghellini, 1998) che i principali problemi sono dovuti ad episodi molto brevi trascorsi in un particolare stato, che usualmente non implicano errori sulla condizione occupata alla data di indagine corrispondente all'intervista panel. Tali errori possono influenzare analisi svolte facendo riferimento al tempo trascorso in un particolare stato (si pensi ad esempio a modelli di durata in tempo continuo prendendo come riferimento temporale il mese piuttosto che l'anno), ma influenzano in modo trascurabile analisi condotte prendendo come riferimento l'anno. Per questo motivo, tale base di dati rappresenta un utile ed adeguato strumento per iniziare a studiare alcune dinamiche proprie della transizione scuola-lavoro.

3 Un modello per dati di panel binari

3.1 Struttura di base

La scelta di un individuo al tempo t può essere rappresentata mediante una variabile qualitativa a due o più categorie, si pensi ad esempio alla scelta se abbandonare o non abbandonare la scuola, accettare o non accettare un

lavoro ecc.. L'analisi di variabili di questo tipo prevede la specificazione di modelli per variabili di risposta qualitative (Amemiya, 1985) e quando le possibilità di scelta sono due, si definiscono modelli per variabili di risposta qualitative binarie.

Si assuma che l'utilità associata ad ogni possibile scelta sia funzione di alcune caratteristiche del soggetto \mathbf{x} , osservabili ed invarianti nel tempo³, e di un termine di errore additivo $\boldsymbol{\varepsilon}$.

Siano U_{i1}^t e U_{i0}^t le utilità indirette associate rispettivamente alle scelte 1/0 al tempo t relative all' i -esimo individuo; si assuma inoltre che le utilità possano essere espresse come funzioni lineare delle caratteristiche \mathbf{x} . Siano le seguenti le funzioni di utilità:

$$U_{i0}^t = \alpha_{0,t} + \mathbf{x}'_i \beta_{0,t} + \boldsymbol{\varepsilon}_{i0}^t \quad (1)$$

$$U_{i1}^t = \alpha_{1,t} + \mathbf{x}'_i \beta_{1,t} + \boldsymbol{\varepsilon}_{i1}^t. \quad (2)$$

L'assunzione di base, sulle funzioni di utilità è che l'individuo i decide per la scelta 1 al tempo t se $U_{i1}^t > U_{i0}^t$ e 0 se $U_{i1}^t < U_{i0}^t$ (nel caso in cui $U_{i1}^t = U_{i0}^t$, si ha uno stato di indecisione, ma assumendo che $\boldsymbol{\varepsilon}_{i0}^t$ e $\boldsymbol{\varepsilon}_{i1}^t$ siano variabili casuali continue, la probabilità associato allo stato di indecisione è nulla).

La probabilità che l'individuo i , al tempo t , decida per la scelta 1, può essere scritta come segue:

$$\begin{aligned} P(y_i^t = 1 | \mathbf{x}_i) &= P(U_{i1}^t > U_{i0}^t | \mathbf{x}_i) \quad (3) \\ &= P[\boldsymbol{\varepsilon}_{i0}^t - \boldsymbol{\varepsilon}_{i1}^t < \alpha_{1,t} - \alpha_{0,t} + \mathbf{x}'_i (\beta_{1,t} - \beta_{0,t})] \\ &= F[(\alpha_{1,t} - \alpha_{0,t}) + \mathbf{x}'_i (\beta_{1,t} - \beta_{0,t})], \end{aligned}$$

dove F è la funzione di ripartizione della variabile casuale $\boldsymbol{\varepsilon}_{i0}^t - \boldsymbol{\varepsilon}_{i1}^t$, i parametri α_t riflettono il livello di utilità associato ad ogni singola scelta ed i parametri β_t l'effetto dei regressori sull'utilità. E' da notare che non si giunge all'identificazione del livello di utilità associato ad ogni scelta, ma solo all'identificazione della differenza tra i due livelli.

Assumendo che la variabile casuale $\boldsymbol{\varepsilon}_{i0}^t - \boldsymbol{\varepsilon}_{i1}^t$ abbia distribuzione logistica e normalizzando i parametri $\alpha_{0,t}$ e $\beta_{0,t}$ a zero, la (3) diventa:

$$P(y_i^t = 1 | \mathbf{x}_i) = (1 + \exp(-\alpha_t - \mathbf{x}'_i \beta_t))^{-1}, \quad (4)$$

³L'ipotesi di invarianza nel tempo di certe caratteristiche o informazioni sugli individui non è necessaria, ma è utile in questo contesto per semplificare la notazione.

dove $\alpha_t = \alpha_{1,t}$ e $\beta_t = \beta_{1,t}$.

Quando le informazioni sono rilevate sulle stesse unità nel tempo la specificazione di un modello del tipo (4) non è plausibile in quanto le osservazioni non sono indipendenti; in tal caso è opportuno specificare un modello che tenga conto della correlazione tra gli stessi individui nel tempo.

Per semplificare la notazione indichiamo con $H_i^t = \{y_i^l, l = 1 \dots t - 1\}$ la storia relativa al soggetto i fino al tempo t . Al fine di tenere conto della correlazione seriale, un modo generale per esprimere il modello (4), è quello di definire la distribuzione condizionata di ogni risposta y_i^t come funzione esplicita della storia passata H_i^t ; in questo modo, la correlazione tra y_i^1, \dots, y_i^T esiste perchè i valori passati y_i^1, \dots, y_i^{t-1} influenzano esplicitamente le osservazioni presenti y_i^t . In questo caso, viene specificato un modello di transizione per la scelta al tempo t .

I modelli di transizione più utilizzati sono le catene di Markov, nella specificazione di tali modelli le distribuzioni condizionali di y_i^t dato H_i^t dipendono solo da q precedenti osservazioni; in particolare per $q = 1$ si specifica una catena di Markov del primo ordine.

Assumendo quindi che ciò che accade al tempo t sia funzione solo di quanto è avvenuto al tempo $t - 1$, cioè un processo di Markov del primo ordine (Cox, 1970; Korn and Whittemore, 1979; Zeger *et al.*, 1985), la (4) può essere scritta come:

$$P(y_i^t = 1 | \mathbf{x}_i, H_i^t) = \left(1 + \exp(-\alpha_t - \mathbf{x}_i' \beta_t - \gamma_t y_i^{t-1})\right)^{-1}. \quad (5)$$

Secondo la specificazione espressa dalla (5), il contributo alla verosimiglianza del soggetto i è dato da

$$L_i(y_i^1, \dots, y_i^T) = P(y_i^1 | \mathbf{x}_i) \prod_{t=2}^T P(y_i^t | \mathbf{x}_i, H_i^t), \quad (6)$$

dove $H_i^t = y_i^{t-1}$.

In econometria, il modello (5) è noto come modello logit sequenziale del primo ordine (Clogg *et al.*, 1990).

3.2 Test di specificazione

Una delle fasi fondamentali di un processo inferenziale è la specificazione di un modello statistico sui dati. L'obiettivo principale in questa fase, è quello

di specificare un modello statistico abbastanza generale e plausibile per la realtà che deve rappresentare, ma allo stesso tempo trattabile da un punto di vista computazionale. In contesti dinamici, talvolta i modelli possono essere anche molto complessi, e possono creare serie difficoltà nella fase di stima, per questo ogni possibile semplificazione del modello stesso è sempre auspicabile. Un'ipotesi che semplifica notevolmente la stima dei modelli dinamici, è quella di stazionarietà; secondo tale ipotesi, nel tempo, non cambia nè l'effetto delle covariate nè l'effetto della variabile ritardata sulla variabile dipendente.

Dato il modello espresso dalla (5), l'ipotesi di stazionarietà si esplicita specificando il seguente sistema di ipotesi:

$$H_0 : \alpha_t = \alpha, \beta_t = \beta, \gamma_t = \gamma \quad (7)$$

$$H_1 : \alpha_t \neq \alpha, \beta_t \neq \beta, \gamma_t \neq \gamma$$

per ogni istante $t = 2 \dots T$. Se k è il numero di variabili indipendenti nel modello e w rappresenta il numero di occasioni di indagine, il numero di parametri da stimare sotto l'ipotesi H_1 è pari a $\{(k+1) \times w\} + w - 1$; sotto ipotesi nulla, il modello risulta vincolato, ed in particolare l'ipotesi di stazionarietà stabilisce che l'effetto dei regressori del modello sia invariante nel tempo, quindi il numero di parametri da stimare si riduce a $[2 \times (k+1) + 1]$.

L'ipotesi H_0 espressa dalla (7) può essere verificata specificando un test del rapporto di verosimiglianze (LR test). In questo caso, la statistica test LR, sotto H_0 , avrebbe una distribuzione χ^2_ν con $\nu = \{(k+1) \times w\} + w - 1 - [2 \times (k+1) + 1]$ gradi di libertà. Se l'ipotesi di stazionarietà più restrittiva (7), cioè quella di invarianza per tutti i parametri del modello, viene rifiutata, si possono specificare ipotesi nulle meno restrittive; si può infatti ipotizzare invariante nel tempo ogni singolo effetto del modello o combinazioni di effetti, in questo caso l'idea è quella di individuare il modello più semplice, che nel contempo offra una opportuna rappresentazione della realtà. Per esempio, potrebbe essere interessante verificare l'ipotesi in cui l'effetto temporale sia completamente spiegato dal termine intercetta, in questo caso il sistema di ipotesi sarebbe:

$$H_0 : \beta_t = \beta, \gamma_t = \gamma \quad (8)$$

$$H_1 : \beta_t \neq \beta, \gamma_t \neq \gamma$$

per ogni $t = 2...T$. In questo caso accettare tale ipotesi implicherebbe l'invarianza temporale di tutte le covariate incluse nel modello.

4 Scuola/non scuola: un modello a scelta binaria

Come sottolineato precedentemente, lo studio della transizione scuola lavoro è basata sui dati retrospettivi dell'indagine LEVA. In particolare, l'analisi è stata condotta sulle coorti del 1986 e 1989 ($N = 3828$) che rispettivamente coprono un arco temporale di circa 7 anni; fino al 1992 per la prima coorte e fino al 1995 per la seconda.

Seguendo le linee descritte nella sezione 3, si indichi con y_i^t la scelta di non abbandonare ($y_i^t = 0$) /abbandonare ($y_i^t = 1$) la scuola per l'individuo i al tempo t . Considerando che un diploma di scuola secondaria si ottiene, in condizioni di regolarità, dopo cinque anni, è ragionevole analizzare l'uscita dalla scuola facendo riferimento alle prime cinque occasioni di indagine.

Alcune analisi preliminari (Ghellini *et al.*, 1993) hanno suggerito di considerare le seguenti variabili come fattori rilevanti nella scelta dei soggetti sotto esame -tutte le variabili sono state trasformate in variabili di tipo dummy 0/1 -: il sesso, SEX (1 maschio); la regolarità negli studi nella scuola dell'obbligo, REGRIT (1 regolare); il livello di istruzione della madre TISTUM (1 basso); il livello di istruzione del padre TISTUP (1 basso); la situazione occupazionale della madre, COPROM (1 occupata); il settore di attività del padre, POPROP (1 operaio); la presenza di entrambi i genitori, FA2 (1 no); il livello di scolarità dell'area, AREA (1 alto). Tutte queste variabili sono state considerate invarianti nel tempo e fissate al tempo $t = 0$ ⁴.

L'analisi si è sviluppata specificando un modello logit sequenziale includendo un effetto coorte. I risultati ottenuti hanno mostrato la non significatività dell'effetto coorte, per questo motivo le successive analisi sono state condotte senza considerare quest'ultimo, e mettendo insieme gli individui di entrambe le coorti. In seguito, poiché solo il primo ritardo è risultato significativo è stata fatta un'assunzione di Markov del primo ordine. Utilizzando i test di specificazione, sottolineati nella sezione 3.2, sono state sottoposte a verifica alcune ipotesi alternative. La tabella 1 mostra i risultati.

⁴Tale ipotesi può apparire piuttosto forte per alcune delle variabili, ma tale assunzione è comunque imposta dall'informazione disponibile.

Tabella 1: Risultati del test di specificazione

Ipotesi H0	χ^2	d.f
Stazionarietà	114.00	30
tutti i par. fissi tranne intercetta	32.46	27

Tabella 2: Stime del modello logit sequenziale

Variables	<i>stato iniziale</i> ($t = 0$) <i>Stime (std)</i>	$t = 1, 2, 3, 4$ <i>Stime (std)</i>
Intercetta t=0	-1.845(0.29)	-
Intercetta t=1	-	-2.848(0.17)
Intercetta t=2	-	-2.402(0.17)
Intercetta t=3	-	-2.097(0.17)
Intercetta t=4	-	-2.114(0.17)
Var. ritard.	-	6.196(0.14)
SESSO	0.132(0.10)	0.318(0.06)
REGRIT	-2.048(0.11)	-1.578(0.09)
COPROM	-0.335(0.11)	-0.164(0.06)
TISTUM	1.198(0.22)	0.955(0.11)
TISTUP	1.058(0.18)	0.780(0.09)
AREA	-0.653(0.10)	-0.282(0.06)
POPPOP	0.230(0.11)	0.353(0.07)
FA2	0.654(0.20)	0.290(0.14)

Si osserva che l'ipotesi di stazionarietà (7) è troppo forte per i dati a disposizione, ma un semplice modello con differenti effetti nel termine intercetta (8) è abbastanza generale per rappresentare il processo osservato tanto quanto un modello complesso dove tutti i parametri variano nel tempo.

Le stime ottenute con quest'ultimo modello sono riportate nella tabella 2.

Come ci si attendeva l'effetto della variabile ritardata (essere usciti dal sistema scolastico al tempo $t - 1$) risulta molto consistente; ciò nonostante le altre variabili esplicative presenti nel modello risultano essere tutte ancora significative, anche se si può notare un generale abbassamento del valore dei coefficienti passando dal modello relativo allo stato iniziale del processo - per definizione privo della variabile ritardata - al modello generalizzato per le transizioni ai quattro istanti di tempo successivi. Possiamo quindi dire

Tabella 3: Caratteristiche individuali

Individui	Caratteristiche specifiche
<i>Caso base</i>	maschio, regolare nella scuola dell'obbligo, la madre non è occupata, il livello di istruzione della madre e del padre è basso, vive in una zona ad alto livello di scolarità, il padre non è operaio, ha entrambi i genitori
<i>leduar</i>	area con basso livello di scolarità
<i>empl_m</i>	la madre è occupata
<i>par0-1</i>	non ha entrambi i genitori
<i>bluecol_f</i>	il padre è operaio
<i>noreg</i>	non è stato regolare nella scuola dell'obbligo
<i>fem</i>	femmina
<i>hedu_m</i>	il livello di istruzione della madre è alto
<i>hedu_f</i>	il livello di istruzione del padre è alto

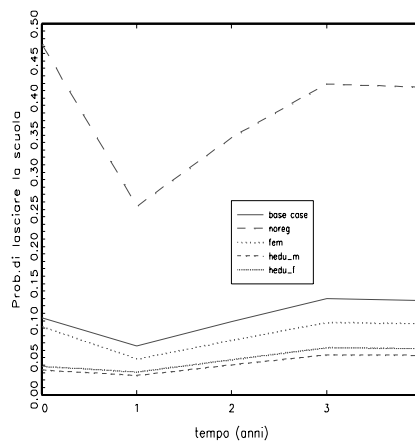
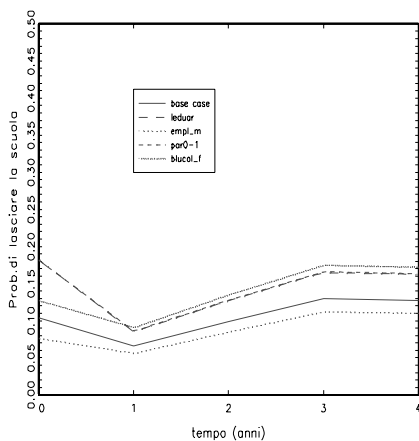


Tabella 4: Prob. di lasciare la scuola 0, 1, 2, 3, 4 anni dopo la scuola dell'obbligo

che il processo di uscita dal sistema scolastico nei primi quattro anni di scuola media superiore risulta essere, seppur diverso in termini quantitativi (le differenze si registrano nei valori delle intercette) fortemente condizionato da un insieme costante di variabili esplicative sia di tipo individuale che di tipo familiare.

Per interpretare meglio i risultati di tale modello sono state calcolate le probabilità di uscita dal sistema scolastico per individui diversi ai vari istanti temporali considerati (vedi tabella 4). Si è innanzitutto definito un individuo base, ovvero quello caratterizzato dai valori modali delle variabili esplicative presenti nel modello (vedi tabella 3), definendo poi altri 8 individui diversi per una sola caratteristica dall'individuo base. Dall'esplorazione del grafico appare evidente come la discriminazione maggiore sia determinata dalla variabile relativa alla storia scolastica pregressa, nel senso che aver perso uno o più anni comporta probabilità di abbandono notevolmente più elevate (circa quattro volte tanto) in tutto il periodo di osservazione. Meno accentuate risultano le altre differenze anche se si deve sottolineare come le femmine e i giovani aventi uno dei genitori con elevato titolo di studio presentano costantemente probabilità di lasciare la scuola meno elevate, mentre l'assenza di un genitore o l'aver il padre operaio tendono ad innalzare la probabilità di non concludere il quinquennio medio superiore.

Alla luce dei risultati dei modelli stimati, l'uscita dal sistema scolastico sembra essere quindi un processo relativamente semplice, prevalentemente determinato dai perduranti effetti delle stratificazioni sociali esistenti. Inoltre i risultati ottenuti nulla ci dicono sui successivi destini di chi lascia la scuola, sulle diverse opportunità di lavoro che caratterizzano i diversi segmenti di popolazione giovanile che si trova ad entrare nel mercato del lavoro una volta lasciato il mondo della scuola. Processi certamente meno lineari e semplici di quelli testé descritti, ma che cercheremo di affrontare e interpretare nel successivo paragrafo.

5 Lavoro/non lavoro: modelli a scelta binaria

5.1 Specificazione dei modelli

Lo studio condotto nella sezione precedente ha cercato di spiegare le dinamiche del processo legate all'uscita dalla scuola senza considerare che gli individui che lasciano la scuola possono trovare un lavoro o meno. Nell'ana-

lisi condotta in questa sezione il principale scopo è quello di analizzare più in particolare il mercato del lavoro di questi giovani, cercando di vedere se la permanenza più o meno lunga nella scuola influenza l'opportunità di trovare un lavoro.

Il modello specificato per questa analisi è un modello logit sequenziale stimato sugli individui che lasciano la scuola in un dato tempo t per i tre anni seguenti l'uscita dalla scuola. Poiché l'interesse è ora quello di spiegare le dinamiche del mercato del lavoro, e non quello che riguarda la scelta di lasciare o meno la scuola, si considerano nell'analisi anche quegli individui che hanno ottenuto un diploma di scuola secondaria.

Sia z_i^t una variabile dicotomica che assume valore 1 se l'individuo i al tempo t è occupato e 0 se si trova in un'altra condizione. L'interesse è quello di stimare la seguente probabilità:

$$P(z_i^t = 1 | \mathbf{x}_i, Hz_i^t, y_i^t = 1) = \left(1 + \exp\left(-\alpha_t^z - \mathbf{x}_i' \beta_t^z - \gamma_t^z z_i^{t-1}\right)\right)^{-1}, \quad (9)$$

dove Hz_i^t indica la storia passata relativa a z e $\alpha_t^z, \beta_t^z, \gamma_t^z$ sono parametri incogniti.

Per ogni t sono stati stimati tre modelli logit sequenziali ed è stata sottoposta a verifica l'ipotesi di stazionarietà. In tabella 5 sono riportati i risultati del test di specificazione; per brevità si riportano solo i risultati delle prime cinque ipotesi sottoposte a verifica⁵. Sembra che per gli individui che lasciano la scuola all'inizio del processo e dopo un anno l'ipotesi di stazionarietà possa essere accettata, mentre per $t = 2, 3$ un semplice modello stazionario non è in grado di rappresentare la dinamica del processo sottostante il mercato del lavoro e un modello più complesso deve essere specificato. In particolare, un modello con un termine intercetta che varia nel tempo è sufficiente per spiegare la variabilità del processo stesso. In $t = 4$, nè l'ipotesi di stazionarietà, nè un modello con un'intercetta variabile nel tempo sono abbastanza generali per rappresentare adeguatamente la realtà, così che deve essere specificato un modello ancora più complesso. I risultati mostrano che un modello ridotto non è capace di rappresentare il processo e l'ipotesi di non stazionarietà deve essere accettata per $t = 4$.

Tale situazione suggerisce una semplice riflessione: via via che l'abbandono del mondo della scuola si sposta in avanti, e quindi i giovani escono con

⁵Si noti che il test non è stato effettuato per $t = 5$ in quanto si dispone solo dell'informazione per due occasioni.

Tabella 5: Risultati del test di specificazione

Ipotesi	χ^2	d.f
stazionarietà (t=0)	12.15	10
stazionarietà (t=1)	3.70	10
stazionarietà (t=2)	20.69	10
tutti par. fissi tranne intercetta (t=2)	13.79	9
stazionarietà (t=3)	34.99	10
tutti par. fissi tranne intercetta (time 3)	17.05	9
stazionarietà (t=4)	33.14	10
tutti par. fissi tranne intercetta (t=4)	23.12	9
tutti par. fissi tranne intercetta e ritardo (t=4)	21.31	8
tutti par. fissi tranne intercetta, ritardo, sex (t=4)	16.74	7
tutti par. fissi tranne intercetta, ritardo, sex, regrit (t=4)	16.7	6
tutti par. fissi tranne intercetta, ritardo, sex, regrit, coprom (t=4)	16.32	5

livelli di scolarità più elevati, diventa più complessa la possibilità di riprodurre, tramite modelli, il comportamento sul mercato del lavoro dei giovani. Ovvero più aumenta il grado di istruzione più articolato si presenta il mix di opportunità e/o scelte che i giovani si trovano ad affrontare.

5.2 Risultati empirici

Per dare un'interpretazione più immediata dei risultati ottenuti, si è preferito non riportare le stime ottenute con la specificazione dei diversi logit sequenziali, ma piuttosto calcolare le probabilità attese di trovare un lavoro per diversi individui, come fatto nel par. 4.

Ogni grafico presentato in tabella 6 rappresenta le probabilità di trovare un lavoro immediatamente dopo aver lasciato la scuola ($t = 0$), dopo un anno ($t = 1$) e dopo due ($t = 2$), per singola coorte di abbandono⁶, evidenziando così le diverse opportunità attribuibili alle diverse caratteristiche degli individui; le diverse linee presenti indicano infatti i diversi individui considerati che differiscono per determinate caratteristiche secondo quanto riportato nella precedente tabella 3. Si deve sottolineare come le coorti 1 e

⁶Con il termine coorte, in questo paragrafo, si intendono individui che condividono la stessa distanza intercorsa tra la fine della scuola dell'obbligo e l'abbandono del sistema scolastico. Si noti inoltre che per l'ultima coorte (coorte 5) l'osservazione disponibile è relativa solo ad un anno dopo l'abbandono.

2 rappresentino in buona approssimazione la situazione dei cosiddetti *drop-out* dal sistema scolastico medio superiore, ovvero coloro che hanno iniziato tale ciclo lo lasciano nel primo biennio senza conseguire alcun titolo ulteriore rispetto alla licenza media, mentre le coorti 3 e 4 fanno riferimento prevalentemente a uscite dal sistema dopo il conseguimento di una con qualifica⁷. Tale fatto ci permette di valutare comparativamente le probabilità di successo nella ricerca di un lavoro tra coorti con percorsi scolastici difformi e di cogliere inoltre come tali difformità possano essere modificate dalle diverse caratteristiche individuali considerate. Si deve inoltre far presente che pur disponendo di dati relativi a un processo a stati (occupato/non-occupato) e a tempi ($t = 0, 1, 2$) discreti, è stato supposto un trend lineare per le probabilità di essere occupati tra i vari istanti considerati e, pertanto, il processo è stato rappresentato con linee continue.

L'esplorazione dei grafici relativi ai diversi individui considerati mette in luce innanzitutto le diverse opportunità occupazionali, sia in termini di probabilità sia nell'andamento temporale, di coloro che hanno già un'occupazione al tempo precedente (ad esclusione della probabilità al tempo 0 che è ovviamente uguale a quella dell'individuo base - ovvero compresa tra 0.7 e 0.8 per tutte le coorti -, essendo, per definizione, impossibile che l'individuo risultasse occupato prima di lasciare la scuola). In generale possiamo osservare che al tempo 1 e 2 la relativa probabilità raggiunge valori prossimi all'unità, con valori leggermente più bassi per la coorte 3 (gli usciti con qualifica professionale). Relativamente agli altri individui considerati è possibile fare ulteriori annotazioni.

Appare evidente come man mano che ci si allontana dal momento di uscita dal sistema scolastico le opportunità di occupazione tendano progressivamente a decrescere, in particolare dopo il primo anno, con la già ricordata eccezione di chi aveva già un'occupazione al tempo $t - 1$ e della coorte 4. Si può quindi affermare che per chi esce precocemente dal sistema scolastico è fondamentale avere un lavoro subito; se ciò non accade le successive possibilità di trovarlo si riducono drasticamente. Inoltre le opportunità di occupazione risultano generalmente più basse per le femmine e per chi ha

⁷Purtroppo i dati retrospettivi utilizzati per queste analisi non ci hanno permesso di avere informazioni precise sul titolo di studio posseduto al momento dell'abbandono scolastico (tale dato è rilevabile solo dai dati panel, affetti come noto da maggiori problemi di mancata risposta). In futuro ci si ripromette di valutare la possibilità di meglio qualificare i dati retrospettivi integrandoli, ove possibile, con le informazioni panel disponibili o con stime da esse desumibili.

avuto un percorso scolastico dell'obbligo non regolare, ma anche per chi ha genitori con titoli di studio medio-alti. Inoltre appare interessante osservare come nel caso di famiglie monogenitore si registrino probabilità di occupazione mediamente più elevate, soprattutto al tempo $t = 1$, momento in cui in generale si riscontrano le maggiori difficoltà.

Si potrebbe quindi supporre che le difficoltà occupazionali delle giovani generazioni con livelli di istruzione medio bassi, oltre che motivate dalle rigidità attuali del mercato del lavoro, trovi origine anche dalle diverse motivazioni nel ricercarlo: ad esempio, chi vive in famiglie mono-genitore, usualmente con meno disponibilità economiche, sente con maggior pressione la necessità di conseguire un reddito di chi invece vive in famiglie con livelli di istruzione medio-alta, usualmente in grado di supportare economicamente lunghi periodi di inoccupazione dei propri figli.

Possiamo inoltre notare come le prime tre coorti presentino generalmente andamenti piuttosto simili, anche se per quanto riguarda la coorte 2, si nota un certo peggioramento della situazione occupazionale per le femmine e per coloro che hanno terminato il ciclo dell'obbligo in ritardo.

Per quanto riguarda la coorte 3, dove ricordiamo essere presente una quota maggioritaria di uscite con qualifica professionale (circa i due terzi), si nota invece una probabilità di impiego iniziale piuttosto elevata (superiore a 0.70), probabilità che però decresce sensibilmente nei due anni successivi, ad eccezione delle femmine che però iniziano il processo da livelli di occupazione molto più bassi.

Leggermente diversa la situazione per la coorte 4 che presenta un andamento a "V" con buone opportunità di occupazione iniziali, un drastico calo a distanza di un anno ed un marcato recupero nell'anno successivo. Si deve inoltre sottolineare come proprio la coorte 4, quella "più istruita", presenti anche la situazione occupazionale migliore per le femmine. Tale fatto sembra quindi confermare la maggiore importanza del livello di istruzione per la componente femminile nella ricerca di lavoro; importanza peraltro ben percepita dalle donne che come è noto in questi ultimi anni hanno notevolmente innalzato i loro tassi di scolarità, in particolare nell'istruzione superiore ed universitaria.

Dall'ultimo grafico della tabella 6 possiamo inoltre osservare il comportamento della coorte 5 di cui disponiamo però solo delle probabilità di trovare lavoro al tempo 0 e 1. La maggior parte di loro sono giovani che hanno concluso il ciclo medio superiore e conseguito un diploma di maturità. Risulta di un certo interesse osservare tutti gli individui presentino probabilità

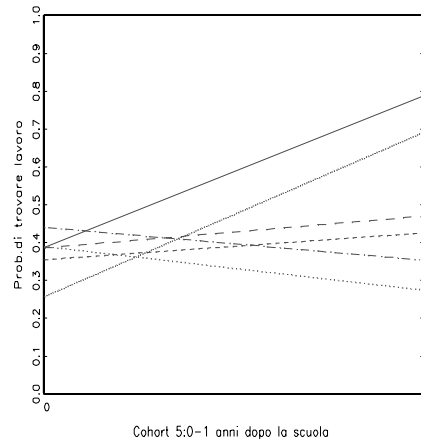
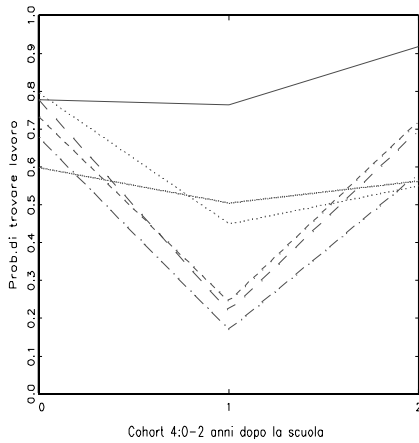
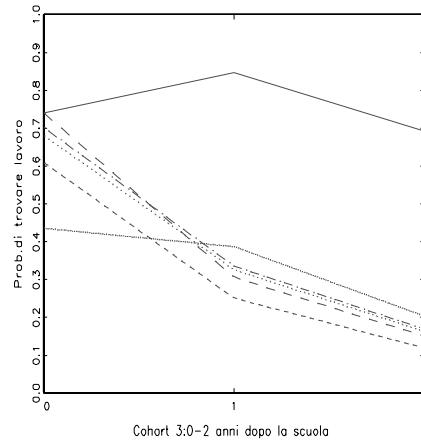
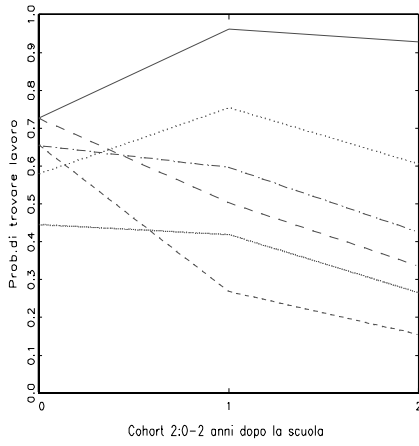
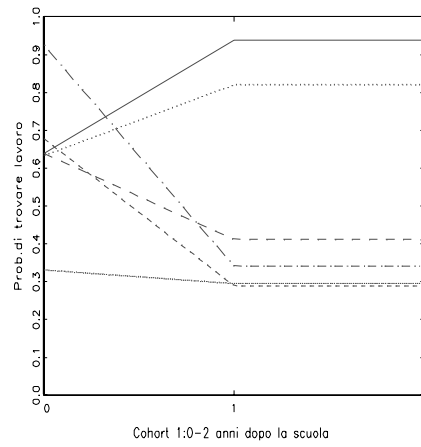
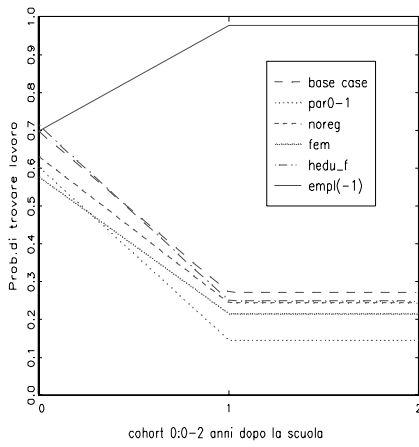


Tabella 6: Prob. di trovare un lavoro 0,1,2, anni dopo l'uscita dalla scuola (la spezzata relativa all'individuo empl(-1) va intesa come la prob. di permanere nello stato di occupazione ad un anno di distanza)

di occupazione piuttosto basse subito dopo il termine degli studi (tra 0.25 e 0.45), con una tendenza alla stabilità anche nell'anno successivo (tale dato, visto anche il *trend* crescente delle femmine, risulta probabilmente influenzato anche dalla presenza di un certo numero di ragazzi impegnati nel servizio militare di leva).

In conclusione possiamo quindi sottolineare alcune evidenze empiriche emerse dall'analisi svolta:

a) la probabilità di trovare lavoro per le femmine è generalmente meno elevata di quella dei colleghi maschi, anche se sono emerse interessanti differenze per le coorti più istruite, dove tale *gap* tende a ridursi considerevolmente. In qualche misura, la componente femminile sembra aver introiettato tale situazione, come dimostra il costante aumento dei livelli di scolarità femminili di questi ultimi anni;

b) chi vive in famiglie con livelli di istruzione più elevati, presenta invece una maggiore facilità di accesso immediato al lavoro - in questo giocano probabilmente un ruolo importante le amicizie e gli aiuti familiari -, anche se tale vantaggio viene drasticamente ridotto nei periodi seguenti. Si potrebbe quindi supporre che in tale segmento siano presenti fenomeni di inoccupazione volontaria, di una ricerca del lavoro più selettiva, resa possibile dalle discrete possibilità economiche delle relative famiglie.

c) Infine, l'analisi svolta individua come il primo anno dopo l'uscita rappresenti indubbiamente il momento più critico del difficile passaggio dal mondo scolastico a quello lavorativo. A questo riguardo sembra importante sottolineare ancora una volta l'importanza di creare opportunità lavorative, anche di breve durata, soprattutto per quei segmenti giovanili scarsamente qualificati, per i quali, in assenza di adeguati interventi e facilitazioni all'accesso, si prefigura un futuro di crescente emarginazione.

6 Considerazioni conclusive

Lo scopo principale di questo lavoro è stato quello di analizzare in un'ottica dinamica i comportamenti individuali del mercato del lavoro giovanile. La fonte di dati utilizzata è una base di dati molto ricca di informazioni in quanto copre un intervallo temporale di circa 7 anni e fornisce informazioni sulle dinamiche di due coorti distanziate di tre anni; di conseguenza i risultati ottenuti, pur se circoscritti alla realtà lombarda, hanno fornito spunti di discussione circa il fenomeno oggetto di studio.

Piuttosto che rimarcare ulteriormente i principali risultati ottenuti, sembra più interessante riflettere su due punti, di natura prettamente metodologica. I modelli specificati nell'analisi sono modelli piuttosto semplici alla cui base vi sono due ipotesi fondamentali: l'esogeneità dello stato iniziale del processo di transizione e l'ipotesi di assenza di fattori non osservati o non osservabili che influenzano i comportamenti individuali osservati.

Se la prima delle due assunzioni sembra piuttosto plausibile in tale contesto in quanto la storia passata del processo, prima dello stato iniziale, è costituita dalla permanenza nello stato di studente (permanenza nella scuola dell'obbligo); la seconda può apparire più forte, in quanto fattori inosservabili quali predisposizione agli studi o risultati ottenuti a scuola possono certamente influire sia sull'abbandono prematuro dalla scuola sia successivamente sulle opportunità di trovare un lavoro.

In tal senso sembra opportuno approfondire successivamente lo studio cercando di specificare modelli più complessi che tengano conto della presenza di eterogeneità non osservabile per studiare l'influenza di quest'ultima sulle dinamiche studiate.

Riferimenti bibliografici

AMEMIYA T., (1985), *Advanced Econometrics*, Basil Blackwell.

CLOGG C.C., EALISON S.R., GREGO J.M., (1990), Models for the Analysis of change in discrete variables, in A. Von Eye (ed.), *Statistical Methods in Longitudinal Research*, vol. II, Academic Press.

BERNARDI L., GHELLINI G., (1992), Dopo l'obbligo: percorsi scolastici brevi e prospettive lavorative nell'esperienza di una coorte, relazione presentata al *Convegno Internazionale di Metodi Quantitativi per le Scienze Applicate*, Siena 11-13 Giugno 1992 (mimeo).

COX D.R., (1970), *Analysis of Binary Data*, London:Methuen.

D'AGOSTINO A., GHELLINI G., (1998), Errori di memoria -un confronto tra dati panel e dati retrospettivi - in un'indagine sulla transizione scuola lavoro, relazione presentata al *Workshop su Lavoro e disoccupazione: questioni di misura e di analisi*, Torino Settembre 1998 (mimeo).

- GHELLINI G., (1987), Progetto LEVA: scuola e lavoro nelle scelte dei giovani studenti lombardi dopo la licenza media, *Collana di Documentazione Statistica*, 26, Regione Lombardia.
- GHELLINI G., (1989), Progetto LEVA- un anno dopo la licenza media: scelte e percorsi dei giovani in Lombardia, *Collana di Documentazione Statistica*, 52, Regione Lombardia.
- GHELLINI G. (1992), Progetto LEVA: Scuola e lavoro nelle scelte degli studenti lombardi dopo la licenza media. Coorte '89, *Collana di Documentazione Statistica*, **66**, Regione Lombardia.
- GHELLINI G. (1990), Progetto LEVA: Successi e insuccessi dei giovani in Lombardia nel triennio post-obbligo- Coorte '86, *Collana di Documentazione Statistica*, **87**, Regione Lombardia.
- GHELLINI G., LAURO S., LEMMI A., REGOLI A., (1993), A transition from school to working life : a longitudinal approach, *Statistica*, anno LIII, **3**, 487-499.
- GHELLINI G. ET ALL. (1993), The Transition from Scholl to Working Life: a Longitudianl Approach, *Statistica*, **3** , anno LIII.
- KORN E.L., WHITTEMORE A.S, (1979), Method for analyzing panel studies of acute health effect of air pollulation., *Biometrics*, **35**, 795-802.
- ZEGER S.L., LIANG K.-Y., SELF S.G., (1985)The analysis of binary longitudinal data with time independent covariates, *Biometrika*, **72**, 31-38.

Statistical Models for School to Work Transition Behaviour

Summary

School to work transition is studied on a longitudinal data base created in the framework of LEVA Project (project Longitudinale sull'Entrata nella Vita Attiva of young people living in Lombardia). The study is conducted using Time Sequence Logit Models; such class of models considers individual choices, not only influenced by a proper set of covariates, but also by the choices made at the previous time. The study is interested in analysing the relationships between individual characteristics and the decision of continuing or not school-life, on one side, and job opportunities after the exit of school, on the other. In order to evaluate these relationships two sets of models have been estimated. Results suggest that the most critical period in the labour market is the first year out of school.

Keywords: Longitudinal data, School to work transition, Time sequence logit models.

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.
13. T. Di Fonzo e P. Gennari, *Le serie storiche delle forze di lavoro per il periodo 1984.1-92.3: prospettive e problemi di ricostruzione*, Giugno 1999.
14. S. Campostrini, A. Giraldo, N. Parise e U. Trivellato, *La misura della partecipazione al lavoro in Italia: presupposti e problemi metodologici di un approccio "time use"*, Ottobre 1999.
15. A. Paggiaro e N. Torelli, *Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro*, Ottobre 1999.
16. A. D'Agostino, G. Ghellini e L. Neri, *A Multiple Imputation Method for School to Work Panel Data*, Ottobre 1999.
17. G. Betti, B. Cheli e A. Lemmi, *Occupazione e condizioni di vita su uno pseudo panel italiano: primi risultati, avanzamenti e proposte metodologiche*, Ottobre 1999.
18. B. Anastasia, M. Gambuzza e M. Rasera, *La durata dei rapporti di lavoro: evidenze da alcuni mercati locali del lavoro veneti*, Marzo 2000.
19. F. Bassi, M. Gambuzza e M. Rasera, *Struttura e qualità delle informazioni del sistema NETLABOR. Una verifica sui dati delle Scica delle province di Belluno e Treviso*, Marzo 2000.
20. N. Rosati, *Permanent and Temporary Inequality in Italy in the 1980s and 1990s*, Marzo 2000.
21. G. Betti, B. Cheli e A. Lemmi, *Analisi delle dinamiche di povertà e disoccupazione su uno pseudo panel italiano*, Marzo 2000.
22. A. D'Agostino, G. Ghellini e L. Neri, *Modelli statistici per l'analisi dei comportamenti di transizione scuola lavoro*, Marzo 2000.

Le richieste di working papers vanno indirizzate alla Sig.ra Patrizia Piacentini, Dipartimento di Scienze Statistiche, via C. Battisti 241-243, 35121 Padova, oppure all'indirizzo e-mail patrizia.piacentini@unipd.it.