

ANNALI DI STATISTICA

Anno 127

Serie X - vol. 15

ISTITUTO
NAZIONALE
DI STATISTICA

CENTRO
INTERUNIVERSITARIO
DI ECONOMETRIA

LA MISURAZIONE DELLE VARIABILI ECONOMICHE E I SUOI RIFLESSI SULLA MODELLISTICA ECONOMETRICA

Atti del Convegno
Roma, 13-14 gennaio 1997

SISTEMA STATISTICO NAZIONALE

ISTITUTO NAZIONALE DI STATISTICA

Roma 1998

Coordinamento scientifico: Enrico Giovannini

La raccolta del materiale e l'organizzazione del volume sono state curate da Francesca Pronesti

**La misurazione delle variabili
economiche e i suoi riflessi
sulla modellistica econometrica**

**Atti del Convegno
Roma, 13-14 gennaio 1997**

Annali di statistica • n. 15 - 1998

Istituto Nazionale di Statistica
Via Cesare Balbo, 16 - Roma

Coordinamento editoriale:
Servizio Sviluppo di prodotti per la diffusione
Via Tuscolana, 1788 - Roma

Realizzazione:
Sonia Vittozzi

Fotocomposizione e stampa:
Albagraf S.p.A.
Via Busto Arsizio, 9 - Pomezia (RM)

Si autorizza la riproduzione a fini non
commerciali e con citazione della fonte.

INDICE

Presentazione	<i>Pag.</i> 7
Aggregazione e disaggregazione temporale di processi ARMA (<i>Massimiliano Marcellino</i>)	9
La stima indiretta di serie storiche e le sue implicazioni per la modellistica econometrica (<i>Giulio Cainelli, Claudio Lupi</i>)	25
The Use of Chain Indices in the Netherlands (<i>Sake de Boer, Jan van Dalen, Piet Verbiest</i>)	39
Macro Modeling with Chain-Type GDP (<i>Chris Varvares, Joel Prakken, Lisa Guirl</i>)	59
I conti nazionali calcolati mediante indici a catena: alcuni primi risultati per il caso italiano (<i>Sandra Maresca</i>)	77
Labor Market Structure and Welfare: A Comparison of Italy and the U.S. (<i>Christopher J. Flinn</i>)	93
La misura della disoccupazione e la modellazione dell'offerta di lavoro: definizioni a priori e stime dipendenti da modelli a confronto (<i>Enrico Rettore, Ugo Trivellato</i>)	127
Selection and Measurement Error Problems in the European Community Household Panel Survey (<i>Franco Peracchi</i>)	147
La stima dei flussi rispetto al lavoro in presenza di indicatori affetti da errore (<i>Nicola Torelli, Francesca Bassi</i>)	171

PRESENTAZIONE

I lavori contenuti in questo volume sono stati presentati nel corso del Convegno "Problemi di misurazione e riflessi sulla modellistica econometrica" organizzato nel gennaio del 1997 dall'Istituto Nazionale di Statistica e dal Centro Interuniversitario di Econometria (CIDE), di cui l'Istat è membro. L'idea di riunire statistici ed econometrici su un tema così delicato fu proposta da chi scrive nella consapevolezza della separazione che troppo spesso esiste tra chi produce i dati statistici e chi usa questi ultimi. Da un lato, infatti, i produttori di dati concentrano i loro sforzi nell'impiego di tecniche statistiche in grado di fornire informazioni sempre più dettagliate, tempestive ed accurate nel tentativo di soddisfare le esigenze, estremamente variegata, del complesso degli utenti, trascurando spesso le problematiche di utilizzo dei dati stessi; dall'altro, gli econometrici si concentrano prevalentemente sulle problematiche metodologiche o contenutistiche della ricerca, tendendo a considerare "dati" i dati statistici, senza approfondirne la natura, le caratteristiche e la qualità, nonostante il rilievo che tali fattori possono assumere anche nella selezione dei metodi di stima da impiegare.

I lavori di questa raccolta coprono diverse aree d'interesse: il primo gruppo di lavori è dedicato a problematiche connesse al tema dell'aggregazione. In particolare, essi trattano gli effetti che l'omissione di variabili rilevanti esercita all'interno di modelli dinamici e le conseguenze che le metodologie di costruzione dei dati di contabilità nazionale a cadenza infrannuale possono indurre nella stima di equazioni di comportamento ¹.

Il secondo gruppo di lavori esamina le metodologie di calcolo degli aggregati di contabilità nazionale a prezzi costanti e le loro ripercussioni sulla metodologia econometrica: in particolare, sono valutate le conseguenze del passaggio da un sistema (come quello attualmente adottato in Italia) basato su un unico anno base ad un sistema di indici a catena. Tale modifica, infatti, impone ai costruttori di modelli econometrici di mutare significativamente l'approccio abitualmente seguito nella definizione della struttura di questi ultimi, in quanto, come noto, le serie costruite sulla base di indici a catena non rispettano la proprietà dell'additività. I lavori qui raccolti illustrano l'esperienza statunitense e quella olandese, presentando anche per la prima volta una prima ricostruzione dei conti nazionali italiani calcolati utilizzando indici a catena.

Il terzo gruppo di articoli si concentra, infine, sulle problematiche che errori di misura possono comportare per analisi microeconomiche. I lavori si concentrano sia su tematiche di carattere teorico, sia su alcuni casi di studio, utilizzando prevalentemente modelli rivolti all'analisi del mercato del lavoro e dei comportamenti delle famiglie, utilizzando dati relativi al caso italiano ed al nuovo panel europeo costruito sulle famiglie.

Il dibattito sollevato nel corso del convegno dai lavori qui raccolti è stato molto ampio ed ha effettivamente coinvolto moltissimi esperti di diverse discipline, favorendo quello scambio di punti di vista che l'iniziativa voleva stimolare e fornendo lo stimolo per futuri approfondimenti di tematiche econometriche e statistiche d'interesse della

¹ Nel convegno era stato presentato un ulteriore lavoro, a cura di M. Forni, dedicato al tema degli effetti delle variabili omesse nei modelli dinamici, il quale è stato pubblicato, in inglese, su una rivista internazionale.

comunità scientifica. A tale proposito, un sincero ringraziamento va a M. Lippi, G. Gambetta (all'epoca presidente del CIDE), T. Di Fonzo, C. Giannini, I. Visco, M. Grassini, B. Contini e R. Orsi per aver presieduto le sessioni ed avere svolto con grande efficacia il ruolo di *discussant*.

L'augurio è che incontri come quello dei quali ora vengono pubblicati gli atti vedano ancora l'Istat ed il CIDE protagonisti insieme nello stimolare una ricerca interdisciplinare che consenta di fornire informazioni statistiche di migliore qualità e risultati econometrici sempre più in grado di offrire adeguati supporti all'analisi ed alla politica economica.

Enrico Giovannini

Direttore centrale dell'Istat e
membro del consiglio direttivo del CIDE

AGGREGAZIONE E DISAGGREGAZIONE TEMPORALE DI PROCESSI ARMA

Massimiliano Marcellino

IGIER, Università Bocconi e Istituto Universitario Europeo

1. Introduzione

In econometria, come in altre discipline, la frequenza di raccolta dei dati è spesso più elevata di quella del processo che si suppone li generi, così che le osservazioni disponibili possono essere pensate come realizzazioni di un processo diverso, aggregato temporalmente. Il caso limite è quello di un processo originario in tempo continuo, ipotesi spesso adottata in finanza, ma questo fenomeno si verifica anche per processi in tempo discreto dato che, ad esempio, per molte variabili macroeconomiche solo dati trimestrali o mensili sono disponibili. L'interesse si concentra però di solito nel verificare ipotesi su particolari caratteristiche del processo originario, non di quello aggregato, quali l'esogeneità di alcune variabili, l'esistenza di relazioni di causalità, la presenza di cointegrazione o l'uguaglianza di alcuni parametri a specifici valori. Perché questo sia possibile bisogna o dimostrare che le caratteristiche di interesse sono invarianti all'aggregazione temporale, o determinare quali sono le loro implicazioni per le proprietà del processo aggregato e verificare se queste ultime sono o meno soddisfatte, o cercare di ottenere delle stime dei dati disaggregati non osservabili e usarle per testare le ipotesi di interesse. Questo articolo propone una breve rassegna dei principali risultati disponibili in letteratura per implementare queste alternative.

Nella sezione 2 viene studiata la relazione tra il processo aggregato e quello disaggregato nell'ipotesi che quest'ultimo appartenga alla classe ARMA univariata in tempo discreto, problema interessante anche dal punto di vista meramente statistico. Si mostra che il processo aggregato è ancora un ARMA ma sia l'ordine delle componenti AR ed MA che i loro coefficienti sono in generale molto diversi da quelli disaggregati. Le conseguenze di una frequenza di aggregazione molto elevata e della diminuzione del numero di osservazioni disponibili per la stima dei parametri sono anche considerate.

Nella sezione 3 l'analisi viene estesa al caso multivariato, anche in tempo continuo. Questo permette poi di analizzare gli effetti dell'aggregazione temporale su caratteristiche quali l'integrazione, la cointegrazione, l'esogeneità, la causalità o la risposta agli shock. Mentre le proprietà di lungo periodo sono invarianti, quelle di breve subiscono notevoli cambiamenti e non dovrebbero perciò essere verificate direttamente con dati aggregati. I cambiamenti possono però essere esattamente determinati e quindi verificati quando il campione aggregato è sufficientemente ampio.

Questo articolo è una versione rivista della relazione presentata al Convegno CIDE-ISTAT "Problemi di misurazione e riflessi sulla modellistica econometrica", Roma, 13-14 Gennaio 1997. Desidero ringraziare i due discussants, Tommy Di Fonzo e Carlo Giannini, per i loro validi commenti. Ringrazio inoltre Giampiero Gallo, Clive Granger, Marco Lippi e Grayham Mizon per molte utili discussioni sugli argomenti trattati in questo articolo. Rimango comunque l'unico responsabile di eventuali errori o imprecisioni

Nella sezione 4 si mostra invece come non sia possibile identificare un solo processo disaggregato a partire da quello aggregato, a meno che non siano verificate particolari condizioni. Un metodo semplice ma alquanto generale per la determinazione di stimatori dei dati disaggregati è poi discusso.

La sezione 5 presenta alcuni commenti conclusivi sugli effetti dell'aggregazione temporale.

2. Aggregazione

Per determinare la relazione tra il meccanismo generatore del processo originario, $x = \{x_t\}_{t=0}^{\infty}$, e quello aggregato temporalmente, $\mathbf{x} = \{\mathbf{x}_T\}_{T=0}^{\infty}$, è necessario analizzare più in dettaglio quella tra gli elementi aggregati e disaggregati. Quest'ultima dipende dal tipo di variabili in esame. Per variabili di tipo stock come la ricchezza o il capitale, il processo aggregato si ottiene selezionando i k -esimi elementi di x , dove k è la frequenza di aggregazione, ovvero $\{x_0, x_1, \dots, x_T, \dots\} = \{x_0, x_k, \dots, x_{Tk}, \dots\}$. Per variabili di tipo flusso come il reddito o gli investimenti, il processo aggregato è invece $\{x_0, x_1, \dots, x_T, \dots\} = \{\sum_{i=0}^{k-1} x_i, \sum_{i=0}^{k-1} x_{k+i}, \dots, \sum_{i=0}^{k-1} x_{Tk+i}, \dots\}$. Il primo tipo di aggregazione viene definito usualmente *point-in-time sampling* o semplicemente *sampling*, mentre il secondo *average sampling* o *temporal aggregation*. Per altre variabili come tassi di interesse o di cambio, gli elementi aggregati possono essere delle medie parziali, possibilmente ponderate, degli elementi originari. Una formulazione generale è quindi $\mathbf{x} = \{\mathbf{x}_T\}_{T=0}^{\infty} = \{\omega(L)x_{ik+k-1}\}_{i=0}^{\infty}$, dove L indica l'operatore ritardo e $\omega(L) = \omega_0 + \omega_1 L + \omega_{k-1} L^{k-1}$ rappresenta lo schema di aggregazione.

Dal punto di vista teorico, determinare il meccanismo generatore aggregato è quindi molto semplice. Partendo da una funzione di distribuzione di dimensione finita per il processo disaggregato, quella aggregata si può ottenere applicando i metodi standard per la trasformazione (lineare) di variabili, ad esempio Mood *et al.* [1974 cap. 5]. In pratica però è estremamente difficile ricavare analiticamente la funzione di distribuzione di dimensione finita aggregata, ed il tipo di trasformazione indotto dall'aggregazione dipende dall'ipotesi sul processo disaggregato, per cui non si possono ottenere risultati generali. In letteratura l'attenzione si è quindi ristretta all'aggregazione di processi lineari appartenenti alla classe ARMA, cosa che, come vedremo, permette di superare queste difficoltà.¹

Assumiamo quindi che $x = \{x_t\}_{t=0}^{\infty}$ evolva secondo l'equazione alle differenze stocastica

$$g(L)x_t = s(L)\varepsilon_t \quad (2.1)$$

dove $g(L)$ e $s(L)$ sono polinomi di ordine g e s in L , $g(L) = 1 - g_1 L - g_2 L^2 - \dots - g_g L^g$, $s(L) = 1 - s_1 L - s_2 L^2 - \dots - s_s L^s$, e ε_t è un errore white noise (WN), $\varepsilon_t \sim WN(0, \sigma_\varepsilon)$.

Proposizione 1. Se x è generato da [2.1], $\mathbf{x} = \{\mathbf{x}_T\}_{T=0}^{\infty}$ soddisfa

$$c(Z)\mathbf{x}_T = h(Z)e_T, \quad e_T \sim WN(0, \sigma_e), \quad (2.2)$$

dove $Z = L^k$ è l'operatore ritardo nella nuova frequenza temporale, ovvero $Z\mathbf{x}_T = \mathbf{x}_{T-k}$. $c(Z)$, $h(Z)$ e σ_e sono determinati nella dimostrazione.

¹ Un'eccezione è l'articolo di Granger e Lee (1993) che studia gli effetti dell'aggregazione lineare e temporale per particolari processi autoregressivi non lineari tramite simulazioni. La conclusione più interessante è che l'aggregazione sembra diminuire la non linearità nel senso che test per questa ipotesi tendono a rifiutarla quando applicati a dati aggregati.

Dimostrazione. Seguendo il metodo introdotto da Brewer (1973) e successivamente perfezionato da Wei (1981), Weiss (1984) e Engel (1984) per i casi di point-in-time e average sampling,² iniziamo con il premoltiplicare entrambi i membri di [2.1] per $\omega(L)$, così che

$$g(L)\omega(L)x_t = s(L)\omega(L)\varepsilon_t \quad [2.3]$$

Il polinomio $g(L)$ può essere decomposto in

$$g(L) = (1-\gamma_1 L)(1-\gamma_2 L)\dots(1-\gamma_g L), \quad [2.4]$$

per cui è possibile individuare un polinomio di ordine $gk - g$ in L , $b(L)$, tale che nel prodotto $c(L) = b(L)g(L)$ tutti i coefficienti dei ritardi che non sono multipli di k sono uguali a zero. In particolare, deve essere

$$b(L) = \left(\sum_{i=0}^{k-1} \gamma_1^i L^i\right) \left(\sum_{i=0}^{k-1} \gamma_2^i L^i\right) \dots \left(\sum_{i=0}^{k-1} \gamma_g^i L^i\right),$$

che implica

$$c(L) = (1 - \gamma_1^k L^k) (1 - \gamma_2^k L^k) \dots (1 - \gamma_g^k L^k).$$

Premoltiplichiamo quindi entrambi i membri di [2.3] per $b(L)$, ottenendo

$$c(L)\omega(L)x_t = b(L)\omega(L)s(L)\varepsilon_t = n(L)\varepsilon_t. \quad [2.5]$$

Il termine di sinistra in [2.5] può essere riscritto come $c(Z)x_T$, per cui la componente autoregressiva aggregata è determinata.

La componente a media mobile aggregata deve invece essere tale che la sua funzione di autocovarianza coincida con quella di $n(L)\varepsilon_t$ valutata a tutti i ritardi multipli di k . Pertanto, i coefficienti di $h(Z)$ e σ_ε sono le soluzioni del seguente sistema non lineare:

$$\left\{ \begin{array}{l} \sum_{i=0}^h h_i^2 \sigma_\varepsilon = \sum_{i=0}^{gk-g+s+k-1} n_i^2 \sigma_\varepsilon \\ -h_j \sigma_\varepsilon + \sum_{i=1}^{h-j} h_{i+j} \sigma_\varepsilon h_i = -n_{jk} \sigma_\varepsilon + \sum_{i=1}^{gk-g+s-jk+k-1} n_{i+jk} \sigma_\varepsilon n_i, \quad j = 1, \dots, h. \end{array} \right.$$

Una soluzione analitica, nell'ipotesi di invertibilità della componente MA, può essere determinata per mezzo del filtro di Kalman, ma questo richiede di introdurre una ulteriore notevole complicazione nella notazione; si rimanda pertanto ad Hamilton (1994, p. 391) per i dettagli.

È quindi possibile caratterizzare completamente il processo aggregato quando quello iniziale e lo schema di aggregazione sono noti. Alcuni commenti ulteriori possono essere utili.

1. L'ordine della componente AR è in generale invariante all'aggregazione temporale, non può mai aumentare ma in casi particolari può diminuire (ad esempio, quando $g(L) = 1 - gL^2$ e $k = 2$, l'ordine passa da due a uno).
2. In generale i coefficienti AR sono alquanto diversi da quelli originari ma non dipendono dallo schema di aggregazione, $\omega(L)$ (a meno che, ad esempio, per particolari scelte di $\omega(L)$ $g(L)$ non possa essere decomposto in $\tilde{g}(L)\omega(L)$).

²Engel (1984) considera anche gli effetti di trasformazione più generali di processi ARMA.

3. Le radici della componente AR passano da γ_i a γ_i^k , $i = 1, \dots, g$, per cui non possono aversi cambiamenti nella stazionarietà del processo. In particolare, processi integrati rimangono integrati, sebbene possano aversi cambiamenti nella frequenza spettrale delle radici unitarie, si vedano Granger e Siklos (1995), Marcellino (1996a) per maggiori dettagli.
4. Un processo ARMA può risultare dall'aggregazione di un processo non ARMA o di differenti processi ARMA, ad esempio $x_t = -g_1 x_{t-1} + \varepsilon_t$ e $x_t = g_1 x_{t-1} + \varepsilon_t$, con $\varepsilon_t \sim WN(0,1)$, danno entrambi luogo a $X_T = g_1 X_{T-1} + e_T$, con $e_T \sim WN(0, 1+g_1^2)$, nel caso di point in time sampling con $k=2$. Questa osservazione ha importanti conseguenze per la disaggregazione temporale, come vedremo in dettaglio nella sezione 4.
5. Una componente MA è in generale introdotta come conseguenza dell'aggregazione sebbene sia assente nel processo originario. L'ordine della componente MA aggregata dipende da g, s, e, k , e dei limiti superiori sono riportati nella tabella 1 per i casi di point-time e average sampling.³

Tabella 1 – Ordine in Z di $c(Z)$ e $h(Z)$

Point-in-time Sampling	Average Sampling
$c(Z) = g, h(Z) = g - 1 - q$ per $qk < g - s \leq (q+1)k$ $q = 0, 1, \dots, g - 1$	$c(Z) = g, h(Z) = g - q$ per $qk < g - s + 1 \leq (q+1)k$ $q = 0, 1, \dots, g$
$c(Z) = g, h(Z) = g$ per $g = s$	$c(Z) = g, h(Z) = g$ per $g = s$
$c(Z) = g, h(Z) = g + q$ per $qk \leq s - g < (q+1)k$ $q = 0, 1, \dots$	$c(Z) = g, h(Z) = g + 1 + q$ per $qk \leq s - 1 - g < (q+1)k$ $q = 0, 1, \dots$

In generale il processo aggregato diventa quindi molto più complicato di quello originario. Quando però la frequenza di aggregazione, k , diverge Tiao (1972) ha mostrato che un ARIMA (p,d,q) soggetto a average sampling diventa un IMA (d,d).

La tabella 2 riporta i risultati di Tiao per la funzione di autocorrelazione delle differenze d-esime del processo aggregato, le quali sono completamente indipendenti dai parametri disaggregati. Nel caso di point-in-time sampling il processo aggregato diventa invece un IMA (d,d-1).

Tabella 2 – Autocorrelazioni limite

d/lag	1	2	3	4
0	0	0	0	0
1	0.2500	0	0	0
2	0.39394	0.01515	0	0
3	0.49296	0.04967	0.00041	0
4	0.56491	0.09353	0.00321	0.00001

³ Nella tabella 1 bisogna identificare la riga appropriata in base alla relazione tra g ed s , scegliere q come il più piccolo valore che soddisfa le ineguaglianze, e quindi determinare il limite superiore per l'ordine della componente MA. Ad esempio, se $g = 2, s = 9, k = 5$, la terza riga è quella appropriata, $5q \leq 7 < (q+1)5$ è soddisfatta per $q = 1$, e il limite superiore è $g + q = 3$.

I risultati nella Proposizione 1 così come quelli di Tiao sono a livello di rappresentazione e, sebbene ci si possa aspettare che siano validi asintoticamente anche a livello di stima, molto poco si sa su cosa succeda quando un campione finito è usato per stimare i parametri del processo aggregato. Un'osservazione interessante a questo riguardo è che gli standard errors (s.e.) degli stimatori degli elementi della funzione di autocorrelazione di un processo stazionario sono inversamente proporzionali alla radice quadrata della dimensione campionaria, T , ovvero *s.e.* $\propto T^{-1/2}$. In seguito all'aggregazione, la dimensione campionaria diventa T/k , per cui gli standard errors aggregati (s.e.a.) sono molto più ampi di quelli originari, *s.e.a.* $\propto (T/k)^{-1/2}$. Di conseguenza, qualora non si verifichi un aumento più che proporzionale nei coefficienti della funzione di autocorrelazione, l'ipotesi che questi ultimi siano tutti uguali a zero a partire da un certo ritardo sarà accettata più facilmente con dati aggregati. Ad esempio, il fatto che modelli IMA (d,d) o IMA (d,d-1) siano spesso una buona rappresentazione per serie annuali potrebbe essere la conseguenza non solo di una frequenza di aggregazione elevata ma anche del limitato numero di osservazioni disponibili. Rossana e Seater (1995) mostrano che questa spiegazione è plausibile per molte serie storiche statunitensi.

3. Il caso multivariato

Assumiamo ora che x sia un processo multivariato n dimensionale, che soddisfi il sistema

$$G(L)x_t = S(L)\varepsilon_t, \tag{3.1}$$

dove $G(L) = I - G_1L - G_2L^2 - \dots - G_gL^g, S(L) = I - S_1L - S_2L^2 - \dots - S_sL^s$, le G ed S sono matrici $n \times n$ di coefficienti, e ε_t è un termine di errore white noise multivariato, $\varepsilon_t \sim WN(0, \Sigma)$. Il problema è sempre determinare il meccanismo generatore del processo aggregato multivariato $x = \{x_T\}_{T=0}^\infty = \{\omega(L)x_{ik}\}_{i=0}^\infty$.⁴

Una prima possibilità, quando $\det G(z) \neq 0$ per $z \leq 1$, consiste nel premoltiplicare entrambi i lati di [3.1] per la matrice aggiunta di $G(L)$, $G^a(L)$, in modo da ottenere la così detta rappresentazione in forma finale (Zellner e Palm [1974]):

$$\det G(L) x_t = G^a(L)S(L)\varepsilon_t, \tag{3.2}$$

A questo punto è possibile usare il metodo descritto nella dimostrazione della Proposizione 1 per determinare la componente AR aggregata, e una sua semplice estensione per la componente MA, come vedremo meglio in seguito.

Una seconda possibilità è suggerita da Lutkepohl (1987, cap. 6.5). Lutkepohl definisce, nella nostra notazione, $y_T = (x_{kT}, x_{kT+1}, \dots, x_{kT+k-1})'$ così che x_T è una combinazione lineare di elementi di y_T . Egli ottiene poi la rappresentazione in forma finale per y_T e la usa per derivare quella di x_T .

Entrambe queste procedure hanno due inconvenienti. Il primo è che l'ordine delle componenti AR ed MA nella rappresentazione in forma finale è molto più elevato di quello nella rappresentazione normale, in genere gn e $s + g(n-1)$ contro g ed s . Di conseguen-

⁴ Quando x è un processo multivariato può accadere che diversi schemi di aggregazione siano applicati ai suoi elementi, ad esempio quando questi rappresentano variabili stock e flusso. Questa situazione può comunque essere trattata in modo simile al caso che consideriamo, si veda Marcellino (1996a).

za, anche la rappresentazione aggregata non è parsimoniosa. Inoltre, secondo inconveniente, quest'ultima è ancora in forma finale mentre sarebbe interessante determinare la controparte aggregata della rappresentazione in [3.1]. Illustreremo quindi ora un metodo, proposto in Marcellino (1996a), per derivare il meccanismo generatore aggregato senza ricorrere a rappresentazioni in forma finale.

A questo scopo, definiamo i vettori di matrici

$$B^v = (B_1, B_2, \dots, B_{gk-g}) \text{ e } G^v = (G_1, G_2, \dots, G_g, 0, \dots, 0),$$

$1 \times gk-g$ nxn $1 \times gk$ nxn

la matrice di matrici

$$G_{gk-g \times gk}^m = \begin{pmatrix} -I & G_1 & G_2 & \dots & G_g & 0 & 0 & \dots & 0 & 0 \\ 0 & -I & G_1 & \dots & G_{g-1} & G_g & 0 & \dots & 0 & 0 \\ 0 & 0 & -I & \dots & G_{g-2} & G_{g-1} & G_g & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & G_g & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & G_{g-1} & G_g \end{pmatrix},$$

e indichiamo con $G_{\cdot k}^v$ e $G_{\cdot k}^m$ rispettivamente il vettore e la matrice di matrici ottenuti cancellando le k -esime colonne (di matrici) di G^v e G^m . Abbiamo quindi

Proposizione 2. Se x è generato da [3.1] e $\det G_{\cdot k}^m \neq 0$, $x = \{x_T\}_{T=1}^{\infty}$ soddisfa

$$C(Z)x_T = H(Z)e_T, \quad e_T \sim WN(0, \Sigma_e).$$

$C(Z)$, $H(Z)$ e Σ_e sono determinati nella dimostrazione.

Dimostrazione. Premoltiplicando entrambi i lati di [3.1] per una matrice $B(L)$ di ordine $gk - g$ in L , è semplice mostrare che la matrice di coefficienti di L^i nel prodotto $B(L)G(L)$ è data dalla i -esima colonna del vettore (di matrici) $B^v G^m - G^v$. Come nel caso univariato, vorremmo determinare $B(L)$ in modo tale che tutti i coefficienti che non sono riferiti ad un multiplo di L^k in $B(L)G(L)$ siano uguali a zero. Questi coefficienti possono essere raggruppati nel vettore di matrici $B^v G_{\cdot k}^v - G_{\cdot k}^m$ e pertanto i coefficienti di $B(L)$ devono soddisfare il sistema

$$B^v G_{\cdot k}^m - G_{\cdot k}^v = 0. \quad [3.3]$$

Se $\det G_{\cdot k}^m \neq 0$, tale sistema ammette l'unica soluzione $B^v = G_{\cdot k}^v (G_{\cdot k}^m)^{-1}$. I coefficienti della componente AR aggregata, $C(Z)$, sono pertanto le k -esime colonne del vettore di matrici $G_{\cdot k}^v (G_{\cdot k}^m)^{-1} G^m - G^v$.

Se poi definiamo $N(L) = B(L)S(L)\omega(L)$ e indichiamo con N_i la matrice di coefficienti di L^i in $N(L)$, allora i coefficienti di $H(Z) = (I - H_1 Z - \dots - H_h Z^h)$ e Σ_e sono le soluzioni del sistema non lineare

$$\begin{cases} \sum_{i=0}^h H_i \Sigma_e H_i' = \sum_{i=0}^{gk-g+s+k-1} N_i \Sigma N_i', \\ -H_j \Sigma_e + \sum_{i=1}^{h-j} H_{i+j} \Sigma_e H_i' = -N_{jk} \Sigma + \sum_{i=1}^{gk-g+s+jk+k-1} N_{i+jk} \Sigma N_i', \quad j = 1, \dots, h. \end{cases} \quad [3.4]$$

Come nel caso univariato, una soluzione può essere determinata sfruttando il filtro di Kalman.

È quindi possibile caratterizzare esattamente il processo ARMA aggregato anche nel caso multivariato.⁵ Si noti che i limiti superiori per l'ordine delle componenti AR ed MA aggregate coincidono con quelli nella tabella 1 e sono pertanto molto più bassi di quelli ottenibili con i due procedimenti alternativi precedentemente illustrati. Ovviamente, il metodo nella Proposizione 2 può essere usato anche per i processi univariati. Una sua semplice modifica permette di derivare il meccanismo aggregato per processi ARMAX, ovvero quando l'insieme dei regressori include delle variabili esogene e loro ritardi, si veda Marcellino (1996a).

Analizziamo ora il caso nel quale il processo aggregato x è ottenuto a partire da un processo n -dimensionale in tempo continuo, X , tramite *exact discretization o continuous averaging*, le controparti del point-in-time e average sampling, rispettivamente. Un riferimento generale per questa situazione è Bergstrom (1990), mentre il risultato che descriveremo è dimostrato in Comte (1994), al quale rimandiamo per ulteriori dettagli e riferimenti bibliografici. Seguendo Comte, assumiamo che X sia un $VAR(g)$, per cui:

$$dD^{g-1}X_t = (\Phi_0 X_t + \Phi_1 DX_t + \dots + \Phi_{g-1} D^{g-1}X_t)dt + \omega dW, \tag{3.5}$$

dove Φ_i e ω sono matrici $n \times n$, $i = 1, \dots, n$, DX è la derivata "mean square" di X e W è un processo Browniano n -dimensionale.

In seguito ad exact discretization x diventa $VARMA(g, g-1)$:

$$x_{T+g} = F_{k1}x_{T+g-1} + F_{k2}x_{T+g-2} + \dots + F_{kg}x_T + \varepsilon_T, \tag{3.6}$$

dove le matrici

$$\bar{\Phi} = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \dots & & & & \\ 0 & 0 & 0 & \dots & I \\ \Phi_0 & \Phi_1 & \Phi_2 & \dots & \Phi_{g-1} \end{pmatrix}; \quad \bar{F}_k = \begin{pmatrix} F_{k1} & F_{k2} & \dots & F_{kg-1} & F_{kg} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \dots & & & & \\ 0 & 0 & \dots & I & 0 \end{pmatrix}$$

sono legate dalla relazione

$$\bar{F}_k = \bar{T}_k e^{\bar{\Phi}k} \bar{T}_k^{-1}, \quad \bar{T}_k = ([e^{(g-i)\bar{\Phi}k}]_{1j})_{1 \leq i, j \leq g}. \tag{3.7}$$

Nel caso invece di continuous averaging, quando

⁵La condizione $\det G \neq 0$ è solo apparentemente restrittiva, in quanto essa è sufficiente ma non necessaria per l'esistenza della martice $B(L)$. Marcellino (1996a) discute i cambiamenti da apportare alla procedura quando questa condizione non è soddisfatta.

$$x_t = \int_0^1 X(T+1-s)ds,$$

x diventa un VARMA (g, g) ma la sua componente AR è uguale a quella in [3.6] (ad esempio, Christiano e Eichenbaum [1987], Marcat [1991]), come nel caso discreto. Scarsa attenzione è stata invece prestata finora alla determinazione della componente MA⁶.

Dopo aver discusso la relazione tra il meccanismo generatore aggregato e quello disaggregato, possiamo passare a considerare brevemente gli effetti dell'aggregazione temporale su particolari caratteristiche delle variabili, interessanti dal punto di vista statistico e/o economico (Tabella 3). Un trattamento più rigoroso, degli esempi ed ulteriori riferimenti bibliografici sono presentati in Marcellino (1996a).

Tabella 3 – Aggregazione e particolari proprietà

Proprietà	Autore/i	Invarianza
Radici unitarie	Pierse e Snell (1995)	Si
Cointegrazione	Granger (1990)	Si
Radici unitarie stagionali	Granger e Siklos (1995)	No
Esogeneità	Campos <i>et al.</i> (1990)	No
Causalità	Sims (1971), Wei (1982)	No
Cicli comuni	Marcellino (1996a)	No
IRF	Granger and Swanson (1992)	No
Decomposizioni Trend-ciclo	Lippi e Reichlin (1991)	No
Misure di persistenza	Rossana e Seater (1995)	No

Un primo risultato interessante è che l'integrazione e la cointegrazione sono invarianti all'aggregazione temporale, escludendo la possibilità di radici unitarie negative o complesse come soluzioni di $\det G(z) = 0$ (Granger e Siklos [1995], Marcellino [1996a, 1996b]). Queste sono essenzialmente proprietà di lungo periodo, per cui tale caratteristica sembra naturale. L'esclusione di radici unitarie non positive, legate alle nozioni di integrazione e cointegrazione stagionale, è necessaria per evitare fenomeni di aliasing che possono determinare un incremento spurio nel numero di relazioni di cointegrazione di lungo periodo.

L'esogeneità debole (Engle *et al.* [1983]) non è invece invariante. Affinché questa proprietà valga è infatti necessario che tutta l'informazione su un insieme di parametri di interesse sia contenuta in un modello condizionato, così che il modello marginale per le variabili condizionanti (ovvero esogene) non è rilevante. In seguito all'aggregazione, i parametri del modello condizionato e marginale diventano una funzione complicata di quelli del modello congiunto originario, ed in generale anche il modello marginale contiene ora informazioni rilevanti sui parametri di interesse. Questa conseguenza dell'aggregazione determina anche la non invarianza di proprietà quali la stabilità dei parametri di un modello condizionato in presenza di cambiamenti in quello marginale, che è importante ad esempio per verificare la rilevanza pratica della Critica di Lucas, la non casualità secondo Granger (come notato da Sims [1971]), e quindi l'esogeneità forte e super che sono combinazioni di esogeneità debole e rispettivamente non casualità secondo Granger e invarianza dei parametri del modello condizionale.

⁶ Un tipo di aggregazione temporale diversa che sta ricevendo notevole attenzione recentemente è la così detta *time deformation*. L'idea è che il tempo rilevante per studiare fenomeni economici (economic time), s , non coincida con quello di generazione di tali fenomeni (*calendar time*), t , ma sia una funzione di quest'ultimo, $s=g(t)$, dove $t \in R$ e g deve soddisfare particolari requisiti (Stock [1988]). Applicazioni allo studio del ciclo economico e del comportamento di variabili finanziarie sono proposte, ad esempio, da Stock (1987), Ghysels e Jasiak (1994), Guillame *et al.* (1995).

La rappresentazione di Wold del processo non è neanche, ovviamente, invariante all'aggregazione temporale, implicando che tutte le proprietà che si basano su particolari valori e restrizioni sui suoi coefficienti non sono invarianti. Esempi rappresentativi sono la funzione di risposta agli impulsi, la decomposizione della varianza, l'identificazione della componente di trend e di quella ciclica di una variabile, e le misure di persistenza degli shock.

La non invarianza è poi probabilmente esacerbata in campioni finiti dal dover stimare i parametri del processo e quindi condurre test per le proprietà di interesse. Ad esempio, Pierce e Snell (1995) hanno mostrato una diminuzione nella potenza dei test per radice unitaria quando condotti con dati aggregati rispetto a dati disaggregati e a parità di lunghezza (finita) dell'intervallo campionario. In pratica, sebbene una radice non unitaria passi da γ a γ^k , cosa che dovrebbe aumentare la potenza del test, la diminuzione del numero di osservazioni disponibili dovuta all'aggregazione temporale prevale e determina la riduzione nella potenza.

È allora impossibile verificare delle caratteristiche di interesse o delle ipotesi su parametri disaggregati con dati aggregati temporalmente? La risposta è no, se il campione a disposizione è sufficientemente lungo e se sia il meccanismo generatore disaggregato che lo schema di aggregazione sono noti. In questo caso è infatti possibile, come abbiamo visto, determinare esattamente il meccanismo generatore aggregato e quindi anche le restrizioni su questo implicate dalle ipotesi di partenza a livello disaggregato. Se tali restrizioni sono accettate in base ai risultati di opportuni test condotti con i dati disponibili, anche le ipotesi di partenza sono corroborate.

Quando invece ipotesi disaggregate sono testate direttamente con dati aggregati, la loro non accettazione può essere dovuta non alla loro non validità ma agli effetti dell'aggregazione temporale. Alcuni autori hanno mostrato come questa possa essere una spiegazione plausibile per contrastanti risultati teorici ed empirici in campo economico, o comunque illustrato le modificazioni nelle implicazioni di una certa teoria economica al cambiare dalla frequenza temporale. La tabella 4 elenca alcuni di questi lavori.

Tabella 4 – Studi sugli effetti dell'aggregazione temporale

Autore/i	Argomento
Rowley e Trivedi (1975)	Investimento
Rossana e Seater (1992)	Salari reali
Bergstrom e Edin (1992)	Disoccupazione
Ermini (1988,1989)	Consumo e reddito
Cartwright e Lee (1987)	Market model
Rudebusch (1995)	Tassi di interesse
Goodhart <i>et al.</i> (1991)	Tassi di cambio
Cunningham e Hardouvelis (1992)	Effetti di shock monetari
Gamber e Joutz (1993)	Effetti di shock reali

Per concludere, l'aggregazione temporale è anche rilevante quando lo scopo dell'analisi è prevedere il comportamento futuro delle variabili. Ad esempio, previsioni trimestrali possono essere ottenute o direttamente con un modello per dati trimestrali o aggregando quelle di un modello mensile. Lutkepohl (1986a, 1986b, 1987 cap 7, 8) compara le due possibilità nell'ipotesi di un processo ARMA multivariato disaggregato e mostra che aggregare le previsioni disaggregate è in generale preferibile in quanto porta ad un minore errore quadratico medio di previsione. Ad un simile risultato arrivano anche Gonzalez (1992), Hotta e Cardoso Neto (1993), mentre Rathjens e Robins (1993) suggeriscono di combinare informazioni aggregate e disaggregate.

4. Disaggregazione

Dopo aver mostrato come sia possibile determinare il meccanismo generatore aggregato a partire da quello disaggregato per processi ARMA, consideriamo ora il problema inverso. In questo caso è rilevante non solo la caratterizzazione del meccanismo generatore disaggregato, ma anche la costruzione di stimatori per i dati disaggregati non osservabili. Discuteremo quindi entrambe le questioni, iniziando con la prima. Un'analisi più approfondita è contenuta in Marcellino (1998), al quale si rimanda per ulteriori dettagli, esempi e riferimenti bibliografici.

Assumendo che il processo aggregato x sia l'ARMA (g, h) univariato in [2.2], ovvero,

$$c(Z)x_T = h(Z)e_T, \quad e_T \sim WN(0, \sigma_e),$$

il problema è determinare le componenti AR ed MA e la varianza del termine di errore in [2.1], cioè $g(L)$, $s(L)$ e σ_ε in

$$g(L)x_t = s(L)\varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon).$$

Proposizione 3. Se x è generato da [2.2], $g(L)$ ed $s(L)$ in [2.1] non sono univocamente identificati. In particolare, $(2^k - 1)g$ differenti specificazioni per la componente AR disaggregata sono compatibili con $c(Z)$, mentre infinite specificazioni per quella MA sono compatibili con $h(Z)$. L'esatta identificazione è possibile se tutte le radici di $g(l) = 0$ sono distinte e positive, o distinte e possibilmente negative se k è dispari, e $s \leq g$, o $s \leq g - 1$ nel caso di point-in-time sampling.

Dimostrazione. Nella Proposizione 1 abbiamo mostrato che $c(Z) = c(L^k)$ è ottenuto premoltiplicando entrambi i membri di [2.1] per $b(L)$. Per individuare le potenziali componenti AR disaggregate dobbiamo quindi considerare tutte le possibili scomposizioni di $c(L^k)$ in $\tilde{b}(L)$ e $\tilde{g}(L)$, con $\tilde{b}(L)\tilde{g}(L) = c(L^k)$. $c(L^k)$ può essere scritto come

$$\prod_{i=1}^g \prod_{j=1}^k \left(1 - \frac{1}{c_{ij}} L\right), \quad [4.1]$$

ed almeno un c_{ij} per ogni i deve essere tale che $\tilde{g}(c_{ij}) = 0$ (si ricordi che l'ordine di $c(Z)$ non può essere superiore a quello di $\tilde{g}(L)$). Gli altri c_{ij} possono invece essere soluzioni o di $\tilde{b}(c_{ij}) = 0$ o ancora di $\tilde{g}(c_{ij}) = 0$, nel qual caso si ha una diminuzione nell'ordine della componente AR in seguito all'aggregazione. Quindi, per ogni i ci sono $2^k - 1$ possibili distribuzioni dei c_{ij} come radici di $\tilde{b}(L)$ e $\tilde{g}(L)$. Di conseguenza, le componenti AR disaggregate potenziali sono $(2^k - 1)g$ e possono essere scritte come

$$\tilde{g}(L) = \prod_m \left(1 - \frac{1}{c_m} L\right), \quad [4.2]$$

dove i c_m sono i c_{ij} che sono considerati radici di $\tilde{g}(l) = 0$. L'ordine delle possibili componenti AR varia quindi da g a gk .

Veniamo ora alla componente MA. Dalla Proposizione 1, i coefficienti MA aggregati, gli h_i , e σ_ε sono le soluzioni del sistema

$$\left\{ \begin{array}{l} \sum_{i=0}^h h_i^2 \sigma_\epsilon = \sum_{i=0}^{gk-g+s+k-1} n_i^2 \sigma_\epsilon, \\ -h_j \sigma_\epsilon + \sum_{i=1}^{h-j} h_{i+j} \sigma_\epsilon h_i = -n_{jk} \sigma_\epsilon + \sum_{i=1}^{gk-g+s-jk+k-1} n_{i+jk} \sigma_\epsilon n_i, \quad j = 1, \dots, h, \end{array} \right.$$

dove ricordiamo che $n(L) = b(L)\alpha(L)s(L)$. I coefficienti MA disaggregati e la varianza del termine di errore disaggregato sono ancora le soluzioni di questo sistema, ma questa volta i termini noti sono i membri di sinistra. I coefficienti di $\alpha(L)$ e quelli di $b(L)$ sono anche noti, sebbene questi ultimi dipendano dalla scomposizione di $c(Z)$ in $\tilde{b}(L)$ e $\tilde{g}(L)$. Abbiamo quindi un sistema di $h + 1$ equazioni così che, in generale, infinite soluzioni sono possibili quando l'ordine della componente MA disaggregata, s , è supposto maggiore di h .

Infine, se $g(l) = 0$ ha radici distinte e positive, o distinte e possibilmente negative se k è dispari, allora esse coincidono con quelle di $h(z) = 0$ elevate alla potenza di $1/k$, per cui la componente AR aggregata può essere univocamente determinata. In questo caso la componente MA aggregata sarà di ordine g (o $g - 1$ nel caso di point-in-time sampling), da cui gli al massimo g (o $g - 1$) coefficienti della componente MA disaggregata invertibile e la varianza del termine di errore devono soddisfare un sistema di $g+1$ (o g) equazioni, e questo è possibile solo se coincidono, rispettivamente, con i coefficienti di $s(L)$ e con σ_ϵ .

In generale non è quindi possibile determinare un solo meccanismo generatore disaggregato a partire da uno aggregato, a meno che quest'ultimo non soddisfi particolari condizioni. Alternative lievemente meno restrittive del requisito in Proposizione 3 sono derivate in Wei e Stram (1990), Nijman e Palm (1990a).⁷

In pratica, l'identificazione di un processo disaggregato è complicata ulteriormente perchè sia l'ordine che i coefficienti del processo aggregato sono stimati con un campione finito di osservazioni e potrebbero quindi essere alquanto differenti da quelli veri. Inoltre, in base al menzionato risultato di Tiao (1972), se la frequenza di aggregazione, k , è molto elevata il processo aggregato diventa completamente non informativo su quello disaggregato, a parte l'ordine di integrazione.

Passiamo ora ad analizzare il problema della stima dei dati disaggregati. Tale questione ha ricevuto una considerevole attenzione nella letteratura e, sebbene gli stimatori dei dati disaggregati siano sempre delle medie parziali ponderate di quelli aggregati, si possono distinguere almeno quattro approcci parzialmente diversi per la determinazione dei pesi. Nel primo metodo gli stimatori vengono determinati in modo da minimizzare una funzione di perdita, rispettando il vincolo di compatibilità dei dati aggregati, ovvero, le stime di dati mensili quando aggregate devono essere uguali ai dati trimestrali utilizzati per la loro costruzione (ad esempio, Boot *et al.* [1967], Cohen *et al.* [1971], Stram e Wei [1986]). Nel secondo metodo si risolve un problema simile ma si ipotizza l'esistenza di una serie di valori preliminari che non soddisfa però il vincolo di compatibilità con i dati aggregati e va pertanto modificata (ad esempio, Denton [1971], Chow e Lin [1971], Fernandez [1981] e Litterman [1983]). Il terzo metodo è una specializzazione dei primi due per il caso di un processo ARMA disaggregato (ad esempio, rispettivamente, Wei e Stram [1990] e Guerrero [1990]). Il quarto ipotizza anche un processo ARMA disaggregato ma considera i dati disaggregati come *missing observations*, e ne determina gli stimatori come tali (ad esempio, Harvey e Pierse [1984], Kohn e Ansley [1986], Nijman e Palm [1986] e Gomez e Maravall [1994]).

L'approccio che illustriamo brevemente è quello in Marcellino (1998), che permette di ottenere molti dei metodi prima citati come casi particolari e mostra che in effetti gli stima-

⁷ Per l'identificazione dei parametri di un processo in tempo continuo a partire da quelli della sua versione discretizzata si vedano Phillips (1973), Hansen e Sargent (1983).

tori ottenuti sono spesso delle semplici proiezioni lineari delle variabili disaggregate su quelle aggregate. Indichiamo sempre con x il processo disaggregato di media zero, che non deve necessariamente appartenere alla classe ARMA ma deve avere momenti secondi, con X quello aggregato, raggruppiamo i primi N e Nk elementi di x e X nei vettori \mathbf{X} e \mathbf{X} , e costruiamo la matrice

$$W_{N \times Nk} = \begin{pmatrix} w_0, w_1, \dots, w_{k-1} & 0, 0, \dots, 0 & \dots & 0, 0, \dots, 0 \\ 0, 0, \dots, 0 & w_0, w_1, \dots, w_{k-1} & \dots & 0, 0, \dots, 0 \\ \dots & \dots & \dots & \dots \\ 0, 0, \dots, 0 & 0, 0, \dots, 0 & \dots & w_0, w_1, \dots, w_{k-1} \end{pmatrix},$$

dove ricordiamo che ω_i è il coefficiente di L^i nello schema di aggregazione temporale $\omega(L)$, così che $\mathbf{X} = W\mathbf{X}$.⁸

Assumiamo anche che la funzione di perdita coincida con l'errore quadratico medio di disaggregazione (MSDE), per cui il problema che desideriamo risolvere è:

$$\min_{\hat{X}} \text{tr}(E(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})') \quad \text{s.t. } \mathbf{X} = W\mathbf{X}. \quad [4.3]$$

Proposizione 4. Lo stimatore lineare che minimizza il MSDE è

$$\hat{\mathbf{X}}^* = P^*\mathbf{X} = V_X W' V_X^{-1} \mathbf{X}. \quad [4.4]$$

dove V_x e V_X sono le matrici di varianze e covarianze di x e X . Inoltre $\hat{\mathbf{X}}^*$ è non distorto, soddisfa $\mathbf{X} = W\hat{\mathbf{X}}^*$, e $E(\mathbf{X} - \hat{\mathbf{X}}^*)(\mathbf{X} - \hat{\mathbf{X}}^*)' = V_x - V_x W' V_X^{-1} W V_x$.

Dimostrazione. Dato uno stimatore lineare generico $\hat{\mathbf{X}} = P\mathbf{X} = P W\mathbf{X}$ la funzione obiettivo si può scrivere come

$$\text{tr}(E(\mathbf{I} - P W)\mathbf{X}\mathbf{X}'(\mathbf{I} - P W)') = \text{tr}((\mathbf{I} - P W)V_X(\mathbf{I} - P W)').$$

P^* deve dunque soddisfare le condizioni del primo ordine

$$-V_X W' + P^* W V_X W' = 0,$$

e in questo caso anche quelle del secondo ordine sono soddisfatte, essendo $W V_X W' = V_x$, una matrice definita positiva. Dunque, $\hat{\mathbf{X}}^* = P^*\mathbf{X} = V_x W' V_X^{-1} \mathbf{X}$.

Inoltre,

$$\begin{aligned} E(\hat{\mathbf{X}}^*) &= P^* E(\mathbf{X}) = P^* W E(\mathbf{X}) = 0, \\ W \hat{\mathbf{X}}^* &= W V_X W' V_X^{-1} \mathbf{X} = \mathbf{X}. \end{aligned}$$

Infine,

⁸ L'analisi seguente rimane valida anche se la frequenza e/o lo schema di aggregazione cambiano nel tempo, cosa che richiede una appropriata modificazione della matrice W .

$$\begin{aligned}
 E(X - \widehat{X}^*)(X - \widehat{X}^*)' &= E(X - P^*WX)(X - P^*WX)' = \\
 E((I - P^*W)XX'(I - P^*W)') &= (I - P^*W)V_X(I - P^*W)' = \\
 [(I - P^*W)V_X]' - [(I - P^*W)V_XW'P^*]' &= [V_X - V_XW'V_X^{-1}WV_X]'.
 \end{aligned}$$

Quando x è un processo Gaussiano debolmente stazionario, \widehat{X}^* coincide con $E(X/X)$ ed è pertanto lo stimatore che minimizza il MSDE. In questo caso, \widehat{X}^* è anche la soluzione del problema

$$\min_X X'V_X^{-1}X \quad \text{s.t. } X = WX, \quad [4.5]$$

che è stata ottenuta da Stram e Wei (1986) per $\omega(L) = 1+L+\dots L^{k-1}$, e per particolari scelte di V_X da Boot *et al.* (1967) e Cohen *et al.* (1971). Se invece x è un processo ARMA, \widehat{X}^* diventa uguale allo stimatore in Guerrero (1990).

Il metodo di base può essere esteso facilmente per stimare serie a diversi livelli di disaggregazione o sfruttando serie a livelli diversi di aggregazione, per modificare valori disaggregati preliminari e per stimare missing observations e outliers, si veda Marcellino (1998) per i particolari.

Il procedimento può anche essere modificato in senso bayesiano per tener conto dell'incertezza su V_X . Questo aspetto è solitamente trascurato nella letteratura, dove si assume semplicemente che V_X sia nota, ma è invece molto importante perchè l'informazione a livello aggregato, V_X , non può essere usata per identificare univocamente V_X (Marcellino [1998], Wei e Stram [1990]), come nel caso della disaggregazione di processi ARMA. Quindi, è bene rendere esplicito che ipotesi a priori sono necessariamente introdotte.

V_X deve essere stimata con un campione finito per cui, anche assumendo che x sia stazionario in senso debole, le stime di ritardi elevati della funzione di autocorrelazione saranno alquanto imprecise. Questo problema si riduce quando il processo è un ARMA e le simulazioni di Chang (1993) mostrano che in questo caso, con l'ipotesi ulteriore di errori distribuiti normalmente, lo stimatore di Stram e Wei (1986), che coincide come ricordato con quello in [4.4], produce buoni risultati.

Per concludere, Nijman e Palm (1990b) confrontano l'efficienza previsiva di modelli ARIMA stimati con dati originali e ottenuti tramite disaggregazione, indicando una marcata preferenza per i primi solo per previsioni di breve periodo.

5. Conclusioni

La conclusione generale è che l'aggregazione temporale può determinare una notevole perdita di informazione. Questa si manifesta sia a livello di rappresentazione, il processo aggregato è in generale molto più complicato di quello disaggregato e il secondo non può essere determinato univocamente quando il primo è noto, sia a livello di stima, lo standard error dei coefficienti della funzione di correlazione aggregata è molto più ampio di quello a livello disaggregato, sia a livello di previsione, è in generale preferibile aggregare previsioni disaggregate piuttosto che usare un modello aggregato per ottenere previsioni aggregate. Di conseguenza, il suggerimento è di migliorare quanto più possibile il meccanismo di raccolta dei dati per avvicinare e possibilmente uguagliare le frequenze di raccolta e generazione.

References

- BERGSTROM, R. (1990), *Continuous Time Econometric Modelling*, Oxford, Oxford University Press.
- BERGSTROM, R. E EDIN, P.A. (1992), "Time Aggregation and the Distributional Shape of Unemployment Duration", *Journal of Applied Econometrics*, pp. 5-30.
- BOOT, J.C.G., FEIBES, W. E LISMAN, J.H.C. (1967), "Further Methods of Derivation of Quarterly Figures from Annual Data", *Applied Statistics*, 16, pp. 65-75.
- BREWER, K.R.W. (1973), "Some Consequences of Temporal Aggregation and Systematic Sampling for ARMA and ARMAX models", *Journal of Econometrics*, 1, pp. 133-154.
- CAMPOS, J., ERICSSON, N.R. E HENDRY, D.F. (1990), "An Analogue Model of Phase-averaging Procedures", *Journal of Econometrics*, 43, pp. 275-292.
- CARTWRIGHT, P.A. E LEE, C.F. (1987), "Time Aggregation and the Estimation of the Market Model: Empirical Evidence", *Journal of Business and Economic Statistics*, 5, pp. 131-143.
- CHAN, W.S. (1993), "Disaggregation of Annual Time-Series Data into Quarterly Figures: A Comparative Study", *Journal of Forecasting*, 12 pp. 677-688.
- CHOW, G.C. E LIN, A. (1971), "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time-Series by Related Series", *Review of Economics and Statistics*, 53, pp. 372-375.
- CHRISTIANO, L.J. E EICHENBAUM M. (1987), "Temporal Aggregation and Structural Inference in Macroeconomics", *Carnegie-Rochester Conference Series on Public Policy*, 26 pp. 63-130.
- COHEN, K.J., MULLER, W. E PADBERG, M.W. (1971), "Autoregressive Approaches to Disaggregation of Time Series Data", *Applied Statistics*, 20, pp. 119-129.
- COMTE, F. (1994): "Discrete and Continuous Time Cointegration", W. P. n. 9442, Paris, CREST.
- CUNNINGHAM, T.J. E HARDOUVELIS G.A. (1992), "Money and Interest Rates: the Effects of Temporal Aggregation and Data Revision", *Journal of Economics and Business*, 44 pp. 19-30.
- DENTON, F. (1971), "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization", *Journal of the American Statistical Association*, 66 pp. 99-101.
- ENGEL, E.M.R. (1984), "A Unified Approach to the Study of Sums, Products, Time Aggregation and Other Functions of ARMA Processes", *Journal of Time Series Analysis*, 5 pp. 159-171.
- ENGLER, R.F., HENDRY, D.F. E RICHARD, J.F. (1983), "Exogeneity", *Econometrica*, 51 pp. 277-304.
- ERMINI, L. (1988), "Temporal Aggregation and Hall's Model of Consumption Behaviour", *Applied Economics*, 20 pp. 1317-1320.
- ERMINI, L. (1989), "Some New Evidence on the Timing of Consumption Decisions and on Their Generating Process", *Review of Economics and Statistics*, 71, pp. 643-650.
- FERNANDEZ, R.B. (1981), "A Methodological Note on the Estimation of Time-Series", *Review of Economics and Statistics*, 63, pp. 471-476.
- GAMBER, E.N. E JOUTZ, F.L. (1993), "An Application of Estimating Structural Vector Autoregression Models with Long Run Restrictions", *Journal of Macroeconomics*, 15, pp. 723-745.
- GHYSELS, E. E JASIAK, J. (1994), "Stochastic Volatility and Time Deformation: An Application of Trading Volume and Leverage Effects", *mimeo*, University of Montreal.
- GOMEZ, V. E MARAVALL, A. (1994), "Estimation, Prediction and Interpolation for Non Stationary Time Series with the Kalman Filter", *Journal of the American Statistical Association*, 89, pp. 611-624.
- GONZALEZ, P. (1992), "Temporal Aggregation and Systematic Sampling in Structural Time-series Models", *Journal of Forecasting*, 11, pp. 271-281.
- GOODHART, C.A.E., MCMAHON, P.C. E NGAMA, Y.L. (1991), "Testing for Unit Roots with Very High Frequency Spot Exchange Rate Data", *Journal of Macroeconomics*, 13, pp. 423-438.
- GRANGER, C.W.J. (1990), "Aggregation of Time-Series Variables: A Survey", in Barker, T. and Pesaran M.H. (eds.) *Disaggregation in Econometric Modelling*, London, Routledge.
- GRANGER, C.W.J. E SWANSON, N.R. (1992), "Impulse Response Function Based on a Casual Approach to Residual Orthogonalization in Vector Autoregressions", W.P. 92-50, University of California at San Diego.
- GRANGER, C.W.J. E LEE, T.H. (1993), "The Effects of Aggregation on Nonlinearity", in R. Mariano(ed.) "Advances in Statistical Analysis and Statistics Computing", iii, Greenwich, JAI Press.
- GRANGER, C.W.J. E SIKLOS, P.R. (1995), "Systematic sampling, Temporal Aggregation, Seasonal Adjustment, and Cointegration. Theory and Evidence", *Journal of Econometrics*, 66, pp. 357-369.

- GUERRERO, V.M. (1990), "Temporal Disaggregation of Time-Series: An ARIMA-Based Approach", *International Statistical Review*, 58, pp. 29-46.
- GUILLAME, D.M., PICTET, O.V., MULLER, U.A. E DECOROGNA, M.M. (1995), "Unveiling Non Linearities Through Time Scale Transformations", *mimeo*, Zurich Olsen and Associates.
- HAMILTON, J.D. (1994), *Time Series Analysis*, Princeton, Princeton University Press.
- HANSEN, L.P. E SARGENT, T.J. (1983), "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities", *Econometrica*, 2, pp. 377-387.
- HARVEY, A.C. E PIERSE, R.G. (1984), "Estimating Missing Observations in Economic Time Series", *Journal of the American Statistical Association*, 79, pp. 125-131.
- HOTTA, L.K. E CARDOSO NETO, J. (1993), "The Effect of Aggregation on Prediction in Autoregressive Integrated Moving-average Models", *Journal of Time series Analysis*, 14, pp. 261-269.
- KOHN, R. E ANSLEY, C.F. (1986), "Estimation, Prediction and Interpolation for ARIMA models with Missing Data", *Journal of the American Statistical Association*, 81, pp. 751-761.
- LIPPI, M. AND REICHLIN, L. (1991), "Trend-Cycle Decomposition and Measures of Persistence: Does Time Aggregation Matter?" *Economic Journal*, pp. 314-323.
- LITTERMAN, R.B. (1983), "A Random Walk, Markov Model for the Distribution of Time Series", *Journal of Business and Economic Statistics*, 1, pp. 169-173.
- LUTKEPOHL, H. (1986a), "Forecasting Vector ARMA Process with Systematically Missing Observation", *Journal of Business and Economic Statistics*, 4, pp. 375-390.
- LUTKEPOHL, H. (1987), "Forecasting Temporally Aggregated Vector ARMA Process", *Journal of Forecasting*, 5, pp. 85-95.
- LUTKEPOHL, H. (1987), *Forecasting Aggregated Vector ARMA Process*, Berlin, Springer-Verlag.
- MARCELLINO, M. (1996a), "Some Temporal Aggregation Issues in Empirical Analysis", W.P. 96/39, University of California at San Diego (apparirà in *Journal of Business and Economic Statistics*).
- MARCELLINO, M. (1996b), "Aggregazione di Processi I(2)", *Atti della XXXVIII Riunione Scientifica della Società Italiana di Statistica*, Vol. 2 pp 549-556.
- MARCELLINO, M. (1998), "Temporal Disaggregation, Missing Observations, Outliers and Forecasting: A Unifying Non-Model Based Procedure", in FOMBY, T. E HILL, R. C. (eds.), *Messy data - Missing Observations, Outliers and Mixed-Frequency Data*, *Advances in Econometrics*, vol. 13.
- MARCET, A. (1991), "Temporal Aggregation of Economic Time Series", in Hansen L.P. e Sargent T.J., *Rational Expectations Econometrics*, Boulder, Westview Press.
- MOOD, A.M., GRAYBILL F.A. E BOES D.C. (1974), *Introduction to the Theory of Statistics*, New York, McGraw-Hill.
- NUMAN, T.E. E PALM, F.C. (1986), "The Construction and Use of Approximations for Missing Quarterly Observations: A Model-Based Approach" *Journal of Time Series Analysis*, 4, pp. 47-58.
- NUMAN, T.E. E PALM, F.C. (1990a), "Parameter Identification in ARMA Processes in the Presence of Regular but Incomplete Sampling", *Journal of Time Series Analysis*, 4, pp. 47-58.
- NUMAN, T.E. E PALM, F.C. (1990b), "Predictive Accuracy Gains from Disaggregate Sampling in ARIMA Models", *Journal of Business and Economic Statistics*, 8, pp. 405-415.
- PHILLIPS, P.C.B. (1973), "The Estimation of Parameters in Systems of Stochastic Differential Equations", *Journal of Econometrics*, 65, pp. 351-362.
- PIERSE, R.G. E SNELL, A.J. (1995), "Temporal Aggregation and the Power of tests for a Unit Root", *Journal of Econometrics*, 65, pp. 333-345.
- RATHJENS, P. E ROBINS, R.P. (1993), "Forecasting Quarterly Data Using Monthly Information", *Journal of Forecasting*, 12, pp. 321-330.
- ROSSANA, R.J. E SEATER, J.J. (1992), "Aggregation, Unit Roots and the Time Series Structure of Manufacturing Real Wages", *International and Economic Review*, 33, pp. 159-179.
- ROSSANA, R.J. E SEATER, J.J. (1995), "Temporal Aggregation and Economic Time Series", *Journal of Business and Economic Statistics*, 13, pp. 441-451.
- ROWLEY, J.C.R. E TRIVEDI, P.K. (1975), *Econometrics of Investment*, New York, Wiley.
- RUDEBUSCH, G.D. (1995), "Federal reserve Interest Rate Targeting, Rational Expectations, and the Term Structure", *Journal of Monetary Economics*, 35, pp. 245-274.
- SIMS, C.A. (1971), "Discrete Approximations to Continuous Time Distributed Lags in Econometrics", *Econometrica*, 39, pp. 545-563.
- STOCK, J.H. (1988), "Estimating Continuous Time Processes Subject to Time Deformation", *JASA*, 83, pp. 77-84.
- STRAM, D.O. E WEI, W.W.S. (1986), "A Methodological Note on Disaggregation of Time Series Totals", *Journal of Time Series Analysis*, 7, pp. 293-302.

TIAO G.C. (1972), "Asymptotic Behaviour of Temporal Aggregates of Time Series", *Biometrika*, 59, pp. 525-531.

WEI, W.W.S. (1981), "Effect of Systematic Sampling and Temporal Aggregation on Causality - A Cautionary Note", *Journal of the American Statistical Association*, 77, pp. 316-319.

WEI, W.W.S. E STRAM, D.O. (1990), "Disaggregation of Time Series Models", *Journal of the Royal Statistical Society, Series B*, 52, pp. 453-467.

WEI, A.A. (1984), "Systematic Sampling and Temporal Aggregation in Time Series Model", *Journal of Econometrics*, pp. 271-281.

ZELLNER, A. E PALM, F.I. (1974), "Time Series Analysis and Simultaneous Equation Models", *Journal of Econometrics*, 2, pp. 17-54.

LA STIMA INDIRETTA DI SERIE STORICHE E LE SUE IMPLICAZIONI PER LA MODELLISTICA ECONOMETRICA

Giulio Cainelli* e Claudio Lupi**

* IDSE-CNR, Milano

** ISPE, Roma

1. Introduzione

Numerosi istituti nazionali di statistica - fra cui l'Istat - utilizzano metodi indiretti di disaggregazione di serie storiche per la stima trimestrale delle principali grandezze di contabilità nazionale. Infatti, in assenza di rilevazione diretta di un dato fenomeno economico-sociale, queste tecniche consentono di ripartire una serie storica annuale in una a più elevata frequenza facendo uso di uno o più indicatori di riferimento e mantenendo il vincolo che la somma (o la media) dei valori trimestrali sia pari al dato annuo. In altre parole, con queste procedure si tenta di utilizzare la dinamica infrannuale dell'indicatore per stimare quella della serie incognita trimestrale. Tra le procedure che utilizzano informazioni ausiliarie un ruolo di grande rilievo è stato assunto, a partire dal lavoro di Chow e Lin (1971), dai cosiddetti metodi "ottimali". La caratteristica fondamentale di queste tecniche consiste nel fornire una soluzione coerente e simultanea alle due fasi che logicamente contraddistinguono ogni processo di trimestralizzazione con indicatori: la stima preliminare dei valori trimestrali della serie incognita e la ripartizione delle discrepanze tra i valori annuali noti della serie da trimestralizzare e quelli ottenuti dall'aggregazione temporale della stima preliminare. In particolare, sia $\{Y_t\}_1^T$ una serie osservata con frequenza 1 (ad esempio annuale) e sia $\{X_t\}_1^T$ una seconda serie (o insieme di serie) strettamente correlata con la prima (l'indicatore) che è osservabile anche a frequenza $m > 1$ (ad esempio trimestrale, $m=4$) sotto forma di $\{x_{\tau}\}_1^{mT}$. Chow e Lin derivano un metodo per stimare la serie trimestrale coerente con $\{Y_t\}_1^T$, $\{y_{\tau}\}_1^{mT}$, utilizzando l'informazione contenuta in $\{x_{\tau}\}_1^{mT}$. Lo stimatore di Chow e Lin (1971) è

$$\hat{y} = x\beta_{GLS} + L\hat{U} \quad [1]$$

Gran parte di questo lavoro è stata realizzata mentre Giulio Cainelli e Claudio Lupi erano rispettivamente ricercatore e primo ricercatore presso il Dipartimento di Contabilità Nazionale ed Analisi Economica dell'Istat. Una versione preliminare è stata prestata al convegno CIDE-ISTAT "Problemi di misurazione e riflessi sulla modellistica econometrica" (Roma, 13-14 gennaio 1997). Gli autori desiderano ringraziare i partecipanti al convegno e, in particolare, Tommaso Di Fonzo, Carlo Giannini, Enrico Giovannini e Ugo Trivellato per i commenti espressi nel corso della presentazione. Patrizia Ordine ha fornito numerosi suggerimenti che hanno contribuito a migliorare notevolmente l'esposizione. Nonostante il lavoro sia il risultato di un progetto comune e della stretta collaborazione tra i due autori, i paragrafi 1, 2, e 5 sono da attribuire principalmente a Giulio Cainelli, i paragrafi 3 e 4 a Claudio Lupi. Le conclusioni sono state redatte in comune. Le opinioni espresse sono degli autori non implicano alcuna responsabilità da parte delle istituzioni di appartenenza

dove \hat{y} è il vettore ($mT \times 1$) delle stime disaggregate, x è il vettore ($mT \times k$) degli indicatori a frequenza m , $\beta_{GLS} = (X' \Omega_U^{-1} X)^{-1} X' \Omega_U^{-1} Y$ con X e Y rispettivamente indicatori e variabile a frequenza 1, $\hat{U} = Y - X \beta_{GLS}$, $\Omega_U = A \Omega_u A'$, con A matrice di aggregazione tale che $Ax = X$ e Ω_u matrice (nota) di covarianze dei disturbi nella relazione che si assume intercorra tra variabile ed indicatore (i) a livello disaggregato. La matrice di lisciamiento, $L = \Omega_u A' \Omega_U^{-1}$, assicura il vincolo $A \hat{y} = Y$.

Dal punto di vista metodologico, al di là dei problemi di scelta degli indicatori,¹ questa procedura presenta diversi punti critici. In primo luogo, la relazione statica assunta fra i valori disaggregati può, in numerosi casi, rappresentare una semplificazione eccessiva. Secondariamente, anche se la matrice di covarianze tra i residui disaggregati viene ipotizzata essere nota, in realtà occorre stimarla imponendo delle restrizioni di identificazione. Nella pratica corrente, la restrizione più comunemente utilizzata è quella di considerare $u \sim \text{AR}(1)$. Ciò assicura una forma estremamente semplice della matrice di covarianze ma implicitamente impone una struttura dinamica ben precisa al modello teorico (Cfr. Hendry e Mizon, 1978; Mizon, 1995). Nell'ambito di questo filone d'indagine, l'interesse dei ricercatori si è concentrato essenzialmente sullo sviluppo delle metodologie statistiche di disaggregazione temporale delle serie.² Minore attenzione è stata invece prestata alle implicazioni che queste procedure possono avere sulle proprietà statistiche delle serie stimate e sulla specificazione econometrica delle equazioni di comportamento. È chiaro che al variare delle assunzioni di base il risultato finale tende anch'esso a cambiare. Non altrettanto ovvio è il fatto che, anche utilizzando gli stessi stimatori, le proprietà dinamiche delle serie trimestrali dipendono in generale dall'insieme delle scelte operate nel corso della trimestralizzazione. In questo lavoro ciò viene mostrato con riferimento al livello di aggregazione contemporanea a cui viene effettuata la trimestralizzazione. Si tratta evidentemente di un aspetto molto particolare, ma non per questo meno importante. Si noti che il punto di vista dell'analisi è sostanzialmente diverso da quello riguardante problemi di tipo *generated regressors* (Pagan, 1984), che non verranno trattati in questo lavoro.

L'articolo è organizzato come segue: il secondo paragrafo è dedicato ad una breve discussione di carattere informale di alcune peculiarità della specificazione econometrica nell'ambito delle equazioni di trimestralizzazione. Successivamente vengono riportati e discussi i risultati relativi al confronto di serie trimestrali dei consumi di Alimentari e Bevande stimate utilizzando livelli differenti di aggregazione contemporanea. Il confronto viene effettuato attraverso l'uso di test non parametrici, in modo da limitare al minimo eventuali arbitrarietà nell'analisi. Il quarto paragrafo analizza le implicazioni per la specificazione econometrica di equazioni dinamiche di comportamento. Nel quinto paragrafo viene proposto ed applicato un criterio di scelta formale tra differenti livelli di aggregazione. L'ultimo paragrafo prospetta alcune conclusioni.

2. Alcuni aspetti della specificazione econometrica delle equazioni di trimestralizzazione

I metodi indiretti di disaggregazione temporale di serie storiche sono basati su modelli econometrici stilizzati: la relazione lineare tra indicato e indicatore a livello annuale. La specificazione di una relazione econometrica presuppone che questa goda di "buone" pro-

¹ Per una discussione dettagliata dei criteri di scelta tra indicatori alternativi e delle problematiche relative alla specificazione delle equazioni di trimestralizzazione si rimanda a Cainelli e Lupi (1996) e Lupi e Parigi (1994).

² Fra coloro che si sono occupati di queste problematiche si vedano Barbone *et al.* (1981), Bassie (1958), Bollino (1994), Cainelli e Lupi (1996), Chow e Lin (1971, 1976), Da Empoli *et al.* (1979), Denton (1971), Di Fonzo (1987, 1990), Gennari e Giovannini (1993), Gregoir (1994), Guerrero (1990), Istat (1983, 1985, 1987, 1992), Lupi e Parigi (1994), Marcellino (1996), Stram e Wei (1986).

prietà statistiche. Tuttavia, il concetto di modello econometrico che sottende una procedura di trimestralizzazione è molto diverso rispetto alla nozione di modello econometrico come riduzione valida del processo che genera i dati (Hendry e Richard, 1982). Nei modelli econometrici "tradizionali", infatti, la scelta delle variabili che spiegano l'andamento della grandezza da modellare discende dalla teoria economica. Per contro, nelle procedure di trimestralizzazione, l'uso di modelli econometrici stilizzati è finalizzato alla produzione statistica e pertanto la scelta degli indicatori deve, per quanto possibile, ricadere su quantità in grado di misurare in modo alternativo la grandezza che si intende trimestralizzare. Ad esempio, nella trimestralizzazione dei consumi finali interni delle famiglie non si tenterà di stimare una funzione aggregata del consumo, bensì di utilizzare nel modello una misura alternativa dei consumi finali di contabilità nazionale che sia direttamente osservabile a frequenza trimestrale. In altri termini, le relazioni econometriche alla base delle procedure di disaggregazione temporale non sono costituite da equazioni di comportamento, quanto piuttosto da equazioni di misurazione. L'uso di equazioni di comportamento all'interno di una procedura di disaggregazione temporale avrebbe ovviamente conseguenze decisamente negative per l'analisi econometrica applicata.

La particolare natura dei modelli di trimestralizzazione genera due ordini di problemi. Il primo - probabilmente il più grave - consiste nel fatto che spesso i modelli econometrici stimati in questo contesto non risultano correttamente specificati. Il secondo discende dal fatto che per la necessità di ricorrere a modelli estremamente semplificati, una procedura "dal generale al particolare" (Hendry, 1987), appare in queste applicazioni di difficile realizzazione.

Nell'ambito della letteratura sulla disaggregazione temporale, un aspetto generalmente trascurato è quello della definizione del livello di aggregazione contemporanea al quale realizzare le stime disaggregate.³ Nelle applicazioni, questa scelta è solitamente lasciata a valutazioni di carattere qualitativo, senza cioè utilizzare criteri econometrici in grado di definire, dal punto di vista quantitativo, la precisione delle stime trimestrali in relazione ai differenti livelli di aggregazione degli indicatori. Di seguito vogliamo mostrare come questo non sia un problema di poco conto. Inoltre, è nostra opinione che l'applicazione di criteri econometrici adeguati possa essere di fondamentale importanza nell'individuare il grado di aggregazione che consente di ottenere stime più accurate del fenomeno in esame.

3. Effetti dell'aggregazione contemporanea nella stima trimestrale dei consumi di alimentari e bevande

Per comprendere più in dettaglio quali possano essere i problemi insiti nell'utilizzare dati stimati nelle diverse forme è utile rifarsi ad un caso concreto. Nella fattispecie abbiamo esaminato gli effetti dell'aggregazione contemporanea nel processo di trimestralizzazione dei consumi di Alimentari e Bevande ed i suoi effetti nell'analisi econometrica applicata. Le serie utilizzate nel presente esercizio sono costituite dai consumi stagionalizzati e grezzi espressi ai prezzi del 1990 per il periodo 1970.1-1994.4. Si tratta delle serie dei consumi finali delle famiglie precedenti all'ultima revisione di contabilità nazionale dell'ottobre 1996.

Per comodità di notazione, indichiamo con A e B rispettivamente la stima trimestrale di Alimentari e di Bevande ottenuta trimestralizzando direttamente la voce aggregata e con SA^i e SB^i la

³ Metodi che permettono la disaggregazione temporale di serie storiche in presenza di un vincolo di aggregazione contemporanea sono stati proposti da Rossi (1982) e Di Fonzo (1990). L'approccio di questi lavori è assai diverso rispetto a quello da noi utilizzato, escludendo *a priori* la possibilità di un guadagno di significatività della relazione stimata direttamente su dati aggregati (si veda il paragrafo 5).

stima ottenuta aggregando le singole componenti trimestralizzate separatamente. L'apice i assume valore d o g a seconda che si tratti della serie trimestrale destagionalizzata o grezza.

Un primo confronto tra le differenti versioni delle stime viene effettuato stimando i semplici modelli

$$\begin{aligned} A_t^i &= c + bSA_t^i + u_t^{Ai} \\ B_t^i &= c + bSA_t^i + u_t^{Bi} \end{aligned} \quad [2]$$

I risultati di questa stima e alcuni test di ipotesi sono riportati in tabella 1.

Tabella 1 – Risultati delle stime OLS del modello statico (1)

Modello	c	b	R ²	Ac(5) ^a	Coint ^b	Wc=0 ^c	Wb=1 ^d	Wjoint ^e
A ^d	-5.304 (45.365)	1.000 (0.002)	0.999	12.467 [0.000]	-9.767 [<0.010]	0.014 [0.907]	0.014 [0.907]	0.007 [0.993]
B ^d	46.043 (25.769)	0.977 (0.013)	0.984	19.124 [0.000]	-7.316 [<0.010]	3.192 [0.077]	3.199 [0.077]	1.600 [0.207]
A ^g	120.91 (102.90)	1.005 (0.004)	1.005	60.093 [0.000]	-5.975 [<0.010]	1.381 [0.243]	1.391 [0.241]	0.696 [0.501]
B ^g	286.30 (60.331)	0.860 (0.030)	0.896	40.835 [0.000]	-3.171 [<0.050]	22.519 [0.000]	22.578 [0.000]	11.289 [0.000]

(a) test di autocorrelazione dei residui fino al quinto ordine (Breusch e Pagan, 1980); (b) test di cointegrazione (Banerjee e Hendry, 1992); (c) test di Wald per l'ipotesi $c=0$; (d) test di Wald per l'ipotesi $b=1$; (e) test di Wald per l'ipotesi congiunta $c=0$ e $b=1$.

Nota: In parentesi tonda sono riportati rispettivamente gli errori standard delle stime e i livelli di significatività marginale dei test.

Alcuni risultati sono immediatamente evidenti. In primo luogo, in tutti i casi si registra la presenza di cointegrazione fra le variabili stimate ai diversi livelli di aggregazione contemporanea.⁴ Ciò è spiegabile in relazione al fatto che i totali annui per versioni differenti della stessa serie sono identici. In secondo luogo, si può notare che nelle equazioni riguardanti la voce Alimentari, non si può rifiutare l'ipotesi nulla di correttezza ($c=0$ e $\beta=1$), mentre per le stime relative alle Bevande i test sono separatamente significativi a circa l'8% per i dati destagionalizzati e sono chiaramente significativi, al di là di qualsiasi ragionevole ampiezza del test, per i dati grezzi. Inoltre è evidente che i modelli relativi alle serie destagionalizzate sono "migliori" di quelli relativi alle serie grezze. Infine si può osservare come i residui dei modelli statici siano chiaramente autocorrelati. In un precedente lavoro (Cainelli e Lupi, 1996) abbiamo mostrato come per ottenere residui "bianchi" fosse necessario utilizzare un modello dinamico e come tuttavia, anche in questo caso, permanessero alcuni problemi di eteroschedasticità. Ne discende che l'uso di una versione o l'altra della serie in esame non sia ininfluente dal punto di vista dell'analisi econometrica. Come già richiamato nell'introduzione, non vogliamo qui fare riferimento a problemi di tipo *generated regressors* (Pagan, 1984), quanto piuttosto a quelli che riguardano l'individuazione delle *relazioni dinamiche* corrette nelle applicazioni che fanno uso di dati stimati secondo procedure indirette.

Allo scopo di indagare più in dettaglio sulle possibili conseguenze per l'analisi econometrica derivanti dall'uso di versioni differenti (tutte *a priori* plausibili) degli stessi dati, si

⁴ Si noti che, al fine di ottenere un test di cointegrazione più efficiente, si è fatto riferimento al test di Banerjee e Hendry (1992), che a differenza del Dickey-Fuller, non impone *common factors*. Di conseguenza, il modello su cui è calcolato il test è un semplice modello dinamico in forma di ADL.

è proceduto all'analisi delle proprietà delle differenze tra le versioni delle serie in esame. Per comodità di notazione indicheremo con DA^i e con DB^i rispettivamente le serie degli scostamenti tra A^i e SA^i e tra B^i e SB^i . La prima informazione utile da rilevare riguarda la normalità di DA^i e DB^i . In tabella 2 sono riportati i risultati dei test.

Tabella 2 – Test di normalità

Variabile	Valore del test ^a	Significatività ^b
DA^d	3.865	0.145
DB^d	2.974	0.226
DA^s	6.552	0.038
DB^s	2.074	0.355

(a) test di Jarque e Bera (1980); (b) livello di significatività marginale.

Dai test effettuati emerge soltanto per DA^s evidenza di assenza di normalità nelle differenze tra le versioni stimate a diversi livelli di aggregazione. In tutti gli altri casi le differenze sono compatibili con una distribuzione normale. D'altra parte, in condizioni ideali le discrepanze dovrebbero essere non solo distribuite secondo una normale, ma anche essere delle differenze di *martingale*. Infatti, in questo caso le discrepanze non hanno strutture dinamiche significative e sono dei semplici rumori bianchi. Abbiamo pertanto ritenuto opportuno applicare dei test non parametrici di martingalità a queste serie. I test in questione sono quelli sviluppati da Durlauf (1991). Si dimostra che questi test sono robusti sotto condizioni piuttosto generali e consistenti rispetto ad un'ampia gamma di alternative.⁵ I test si basano sul confronto dello spettro della serie osservata con quello di un rumore bianco. Durlauf dimostra che sotto l'ipotesi nulla che $\{x\}$ è una sequenza di differenze di *martingale*,⁶ per $T \rightarrow \infty$

$$U_T(j) = (2T)^{1/2} \int_0^{j\pi} \left(f_x(\omega) - \frac{1}{2\pi} \right) d\omega \Rightarrow B(j), j \in [0,1] \quad [3]$$

dove T rappresenta il numero di osservazioni, $f_x(\cdot)$ è una stima (anche non necessariamente consistente) della densità spettrale standardizzata della serie oggetto di studio, $B(j)$ è il ponte browniano in $[0,1]$ e \Rightarrow indica convergenza in misura di probabilità. In pratica, poiché non è richiesta la consistenza della stima della densità spettrale, nello sviluppo dei test si può usare il periodogramma normalizzato. In virtù del *continuous mapping theorem*, dalla [3] si deriva

$$\sup_{j \in [0,1]} |U_T(j)| \Rightarrow \sup_{j \in [0,1]} |B(j)| \equiv KS \quad [4]$$

$$\int_0^1 U_T(j)^2 dj \Rightarrow \int_0^1 B(j)^2 dj \equiv CVM \quad [5]$$

⁵ Uno studio dettagliato delle proprietà di questi test nei campioni finiti è contenuto in Lupi (1995).

⁶ L'ipotesi nulla è in realtà più complessa, dovendo includere alcune condizioni di regolarità. Per una descrizione dettagliata ed una discussione si rimanda al lavoro originale di Durlauf (1991).

dove KS e CVM sono rispettivamente la statistica di Kolmogorov-Smirnov e quella di Cramér-von Mises. Inoltre, per $j < k \in [0,1]$ fissati si ha

$$U_T(j) \mid \Rightarrow B(j) \sim N(0, j - j^2) \quad [6]$$

$$U_T(j) - U_T(k) \Rightarrow B(j) - B(k) \sim N(0, (j - k) - (j - k)^2) \quad [7]$$

La [4] e la [5] possono essere utilizzate per verificare se le proprietà dinamiche (in termini di spettro) di una serie osservata sono complessivamente coerenti con quelle di un rumore bianco. Tramite la [6] e la [7] è possibile effettuare il confronto limitatamente a bande di frequenza ritenute di particolare interesse.

I risultati derivanti dall'applicazione dei test costruiti sulla [4] e sulla [5] sono riportati in tabella 3 e indicano, coerentemente con quelli derivanti dalle regressioni statiche, che l'ipotesi di martingalità può essere rifiutata con un elevato grado di fiducia.

Tabella 3 – Test di martingalità

Variabile	KS^a	Significatività ^b	CVM^c	Significatività ^b
DA^d	2.429	<0.01	1.634	<0.01
DB^d	2.813	<0.01	2.631	<0.01
DA^s	2.062	<0.01	1.472	<0.01
DB^s	2.452	<0.01	2.222	<0.01

(a) test di Durlauf (1991) nella versione Kolmogorov-Smirnov; (b) livello di significatività marginale; (c) test di Durlauf (1991) nella versione Cramér-von Mises

Un'informazione estremamente importante ai nostri scopi è quella derivante dall'applicazione dei test [6] e [7] agli spettri delle serie delle discrepanze tra le diverse versioni della stessa serie. In questo modo è possibile capire quali siano le frequenze a cui esistono differenze significative. In particolare, le serie relative alle discrepanze tra differenti versioni dei dati, sia destagionalizzati che grezzi, sono state sottoposte a test formali utilizzando le due varianti per "bande di frequenza" [6] e [7]. I risultati di questi test sono riportati nelle tabelle 4 e 5 e confermano la significatività statistica delle deviazioni dall'ipotesi di *martingale difference* delle serie in questione. Per quanto riguarda DA^d e DB^d , ad esempio, è abbastanza evidente la concentrazione di densità spettrale presente a frequenze corrispondenti a periodi di poco inferiori all'anno e mezzo e, in misura minore, anche a periodi più lunghi. Ciò significa che le discrepanze tra differenti versioni destagionalizzate della stessa serie, anche apparentemente molto simili fra loro, hanno strutture dinamiche piuttosto complesse, con "cicli" a frequenze d'interesse per l'analisi del *business cycle*. Analogamente, nel caso delle serie DA^s e DB^s si riscontra una struttura correlativa in corrispondenza delle frequenze stagionali. Da ciò deriva che, evidentemente, la stagionalità delle serie finali è differente a seconda che vengano stimate in forma aggregata piuttosto che come aggregazione delle serie elementari. Questi due risultati sono spiegabili in termini di peso dell'indicatore nel determinare la dinamica della serie finale. In altre parole, se a livello disaggregato gli indicatori sono poco significativi, il loro peso nel processo di trimestralizzazione è limitato e pertanto trasferiscono soltanto in modo molto parziale la propria dinamica alla serie finale.

Tabella 4 – Test di martingalità per bande di frequenza

p^a	$U(DA^d)^b$	$U(DB^d)^b$	$U(DA^s)^b$	$U(DB^s)^b$
32.00	-0.417 [0.085]	-0.429 [0.076]	-0.432 [0.074]	-0.437 [0.070]
16.00	-0.621 [0.060]	-0.763 [0.021]	-0.809 [0.014]	-0.860 [0.009]
10.67	-0.654 [0.094]	-0.667 [0.087]	-1.145 [0.003]	-1.213 [0.001]
8.00	-0.818 [0.059]	-0.448 [0.300]	-1.520 [0.000]	-1.540 [0.000]
6.40	-0.873 [0.060]	0.632 [0.173]	-1.868 [0.000]	-1.727 [0.000]
5.33	0.624 [0.197]	2.199 [0.000]	-1.825 [0.000]	-1.884 [0.000]
4.57	1.992 [0.000]	2.705 [0.000]	-1.852 [0.000]	-2.107 [0.000]
4.00	2.176 [0.000]	2.721 [0.000]	-0.108 [0.828]	0.167 [0.892]
3.56	1.193 [0.000]	2.534 [0.000]	1.432 [0.003]	2.392 [0.000]
3.20	1.193 [0.000]	2.217 [0.000]	1.116 [0.021]	2.171 [0.000]
2.91	1.616 [0.001]	1.840 [0.000]	0.716 [0.122]	1.789 [0.000]
2.67	1.297 [0.003]	1.492 [0.001]	0.311 [0.472]	1.402 [0.001]
2.46	0.938 [0.016]	1.161 [0.003]	-0.100 [0.797]	1.014 [0.009]
2.29	0.627 [0.058]	0.795 [0.016]	-0.500 [0.130]	0.616 [0.062]
2.13	0.279 [0.249]	0.394 [0.104]	-0.842 [0.000]	0.238 [0.394]

(a) periodo in trimestri; (b) test [6]

Nota: In parentesi quadra è riportato il livello di significatività marginale

Tabella 5 – Test di martingalità per bande di frequenza

p^a	DU(DA d^b)	DU(DB d^b)	DU(DA s^b)	DU(DB s^b)
32.00	-0.416 [0.085]	-0.428 [0.076]	-0.431 [0.074]	-0.437 [0.070]
16.00	-0.204 [0.398]	-0.334 [0.167]	-0.378 [0.118]	-0.422 [0.081]
10.67	-0.033 [0.891]	0.095 [0.691]	-0.336 [0.164]	-0.353 [0.143]
8.00	-0.164 [0.498]	0.219 [0.365]	-0.375 [0.121]	-0.326 [0.177]
6.40	-0.054 [0.822]	1.080 [0.000]	-0.347 [0.151]	-0.187 [0.438]
5.33	1.496 [0.000]	1.567 [0.000]	-0.043 [0.859]	-0.156 [0.518]
4.57	1.368 [0.000]	0.505 [0.036]	-0.027 [0.910]	-0.223 [0.355]
4.00	0.183 [0.447]	0.015 [0.947]	1.744 [0.000]	2.175 [0.000]
3.56	-0.241 [0.317]	-0.186 [0.441]	1.540 [0.000]	2.324 [0.000]
3.20	-0.008 [0.973]	-0.317 [0.190]	-0.315 [0.192]	-0.220 [0.361]
2.91	-0.310 [0.199]	-0.377 [0.118]	-0.399 [0.098]	-0.381 [0.115]
2.67	-0.319 [0.187]	-0.348 [0.150]	-0.405 [0.093]	-0.386 [0.110]
2.46	-0.359 [0.137]	-0.330 [0.171]	-0.411 [0.089]	-0.388 [0.108]
2.29	-0.310 [0.199]	-0.366 [0.130]	-0.399 [0.098]	-0.398 [0.100]
2.13	-0.347 [0.150]	-0.400 [0.097]	-0.342 [0.157]	-0.377 [0.118]

(a) periodo in trimestri; (b) test [7]

Nota: In parentesi quadra è riportato il livello di significatività marginale

4. L'effetto sulla stima delle equazioni di comportamento

L'esistenza di differenze significative tra le proprietà dinamiche delle diverse versioni della stessa serie suggerisce la presenza di errori di misurazione autocorrelati nella stima trimestrale del tipo

$$\hat{x}_t = x_t + \eta_t$$

[8]

dove \hat{x}_t è la serie stimata, pari alla serie "vera" soggetta all'errore di misurazione

$$\eta_t = A(B)\varepsilon_t, \quad [9]$$

con $\varepsilon_t \sim NIID(0, \sigma_\varepsilon^2)$ e $A(B)$ invertibile.

Si assuma ora di voler stimare un semplice modello tra le due variabili $\{y\}$ e $\{x\}$ tali che $(y,x) \sim CI(1,1)$, con

$$y_t = c + \beta x_t + \xi_t, \quad \xi_t \sim NIID(0, \sigma_\xi^2) \quad [10]$$

utilizzando le stime trimestrali destagionalizzate di $\{x_t\}$, $\{\hat{x}_t\}$.⁷ L'uso di $\{\hat{x}_t\}$ invece di $\{x_t\}$ comporta che il modello stimato è

$$y_t = c + \beta \hat{x}_t + v_t \quad [11]$$

dove $v_t = \xi_t + \beta \eta_t$ è autocorrelato. La soluzione che si adotta generalmente in pratica in presenza di autocorrelazione è quella di aumentare il modello con altre variabili e in particolare con ritardi delle variabili già presenti. La conseguenza è che si tende a stimare un modello del tipo

$$y_t = c + \beta \hat{x}_t + \gamma' z_t + v_{2t}$$

dove z_t è un vettore contenente valori ritardati di Δy_t e valori correnti e ritardati di $\Delta \hat{x}_t$. Quando il DGP è dato dalla relazione statica tra le variabili "vere" questa soluzione pone chiaramente dei problemi di efficienza delle stime. Al di là di questo, c'è evidentemente una questione di identificazione delle relazioni dinamiche, contaminate dalla presenza di autocorrelazione nel residuo dell'equazione statica. Inoltre, esiste un problema di stima e di inferenza rispetto al parametro di lungo periodo, dato che, a meno che v_{2t} e le innovazioni del processo marginale per $\{\hat{x}_t\}$ non siano congiuntamente una sequenza di differenze di *martingale*, la distribuzione di $T(\hat{\beta} - \beta)$ è distorta e non è più *mixed Gaussian* (Phillips, 1988). Infatti, in Lupi (1996) si dimostra come, nella generalità dei casi, la presenza di un errore di misurazione implica che la tecnica di aumentare il modello con elementi di dinamica non consente di risolvere questo aspetto.⁸

Alcuni autori⁹ suggeriscono, quando possibile, di utilizzare dati grezzi, per evitare complicazioni indotte dal filtraggio delle serie operate dai metodi di destagionalizzazione. Paradossalmente, i risultati emersi nel paragrafo precedente riguardanti i dati grezzi fanno temere che l'uso di dati trimestralizzati in forma grezza possa portare a più problemi di quanti non possa risolverne. Il processo che caratterizza le discrepanze riscontrate nel paragrafo precedente relativamente ai dati stagionali è a sua volta fortemente stagionale. Si assuma di nuovo che le variabili d'interesse siano cointegrate e che $\{x\}$ sia misurata con un errore che, in questo caso, è

$$S(B)\eta_t = w_t \quad [13]$$

⁷ Per semplicità, assumiamo che solo la variabile x sia soggetta ad errore di misurazione. Tuttavia, l'argomento può essere generalizzato senza modifiche importanti né nel procedimento, né nelle conclusioni.

⁸ Ciò viene dimostrato anche per errori di misurazione che sono dei semplici rumori bianchi. Questo ed altri semplici schemi di errori di misurazione sono illustrati in dettaglio in Lupi (1996), dove viene riportata anche evidenza Monte Carlo sulla rilevanza empirica degli argomenti trattati. In presenza di errori di misurazione, la soluzione prospettata è quella di utilizzare stimatori di tipo *Fully Modified Least Squares* (Phillips e Hansen, 1990).

⁹ Si veda ad esempio Wallis (1974). Questa tendenza si è andata rafforzando in anni recenti: per tutti si veda Ghysels e Perron (1993).

dove $S(B)=1+B+B^2+B^3$ e w_t è *white noise*. Il fatto che le variabili siano cointegrate implica l'esistenza di una rappresentazione ECM per cui

$$\Delta y_t = c + \beta \Delta x_t + \gamma(y_{t-1} - \alpha x_{t-1}) + \zeta_t \quad [14]$$

nella quale, sostituendo la [8], si ottiene

$$\Delta y_t = c + \beta \Delta \hat{x}_t + \gamma(y_{t-1} - \alpha \hat{x}_{t-1}) + \beta \Delta \eta_t + \alpha \gamma \eta_{t-1} + \zeta_t \quad [15]$$

Utilizzando la proprietà che $\Delta S(B) = \Delta_s$ e, moltiplicando l'espressione precedente per $S(B)$ si ottiene

$$\Delta_s y_t = c + \beta \Delta_s \hat{x}_t + \gamma(S(B)y_{t-1} - \alpha S(B)\hat{x}_{t-1}) + n_t \quad [16]$$

dove $n_t = \Delta_s \eta_t - \alpha \gamma \eta_{t-1} + S(B)\zeta_t$. Di conseguenza, analogamente al caso relativo alle variabili destagionalizzate, si ha una perdita di efficienza nelle stime unitamente all'emergere di relazioni dinamiche spurie, quando si cerchi di compensare l'autocorrelazione dei residui "aumentando" il modello. Si noti anche che le differenze stagionali non sono sufficienti in questo caso ad eliminare gli effetti della stagionalità e che anche il fattore ECM stagionale $(y_{t-s} - \alpha \hat{x}_{t-s})$ presenta un andamento stagionale a sua volta.

5. Criteri di scelta tra differenti livelli di aggregazione contemporanea

Dato il potenziale effetto sulla modellistica econometrica esercitato dal differente grado di aggregazione contemporanea a cui vengono stimate le serie trimestrali, sarebbe certamente di notevole utilità poter disporre di un criterio statistico per la scelta del livello di aggregazione "ottimale". In questo senso può essere utile sfruttare quella parte della letteratura econometrica che si è dedicata al problema dell'accuratezza della stima e al guadagno o alla perdita informativa connessi all'impiego di modelli caratterizzati da un differente grado di aggregazione contemporanea. In termini formali, il punto di partenza per l'analisi dei problemi di aggregazione di relazioni lineari è dato dalla specificazione di un modello disaggregato, composto da r equazioni lineari del tipo

$$H_d: Y_i = X_i \beta_i + U_i \quad [17]$$

con $i=1, \dots, r$, Y_i un vettore $(T \times 1)$ e X_i una matrice $(T \times k)$. Il modello aggregato, corrispondente alla [17], può essere espresso come

$$H_a: Y_a = X_a \beta_a + U_a \quad [18]$$

dove $Y_a = \sum_{i=1}^r Y_i$ e $X_a = \sum_{i=1}^r X_i$

Il problema è quello di decidere se utilizzare il modello aggregato [18] o quello disaggregato [17]. Un primo criterio econometrico del tipo *goodness of fit*, che consente di operare una scelta tra modelli con differente livello di aggregazione è stato proposto da Grunfeld e Griliches (1960). In base a questo criterio, si preferisce il modello disaggregato H_d se e solo se

$$\sum_{i=1}^r \hat{U}_i' \hat{U}_i < \hat{U}_a' \hat{U}_a \quad [19]$$

dove i residui \hat{U}_i e \hat{U}_a sono ottenuti sulla base di una stima OLS delle due versioni del modello. Un limite rilevante che contraddistingue il criterio [19] discende dal fatto che si assume implicitamente che le variabili dipendenti delle r equazioni disaggregate siano espresse in funzione del medesimo insieme di k regressori e che i termini di errore delle r equazioni disaggregate (U_i , $i=1,2, \dots, r$) siano tra loro incorrelati. Recentemente è stata proposta una generalizzazione della [19] che, oltre a consentire la presenza di covarianza contemporanea fra gli errori disaggregati, permette di considerare specificazioni più "generali" delle equazioni disaggregate, ciascuna basata su un insieme di regressori differenti (Pesaran *et al.*, 1989). Seguendo Pesaran *et al.* (1989), si considerino X_i e X_a composti rispettivamente da k_i e da k_a regressori con $k_i \neq k_a$. La generalizzazione del criterio [19] è data dal confronto fra le seguenti espressioni

$$S_a^2 = \frac{\hat{U}_a' \hat{U}_a}{T - k_a} \quad [20]$$

$$S^2 = \sum_{i,l=1}^r \hat{\sigma}_{il}$$

dove $\hat{\sigma}_{il} = \{T - k_i - k_l + \text{tr}(B_i B_l)\}^{-1} \hat{U}_i' \hat{U}_l$ con $B_i = X_i(X_i' X_i)^{-1} X_i'$.

Analogamente al criterio [19] si sceglierà il modello disaggregato solo se $S^2 < S_a^2$.

Per quanto riguarda la voce Alimentari, i risultati presentati in un nostro precedente lavoro (si veda Cainelli e Lupi, 1996) mostrano una chiara superiorità del modello aggregato, mentre il risultato opposto si registra per la voce Bevande. Da ciò deriva l'indicazione che i consumi in beni Alimentari sono più affidabili se stimati direttamente a livello aggregato, e non come aggregazione delle voci elementari. Per il produttore di dati, l'implicazione è che è preferibile stimare le voci di consumo relative agli Alimentari dapprima a livello aggregato per utilizzare successivamente la stima così ottenuta come *benck-mark* per la valutazione delle singole voci elementari. Per questo secondo stadio si possono utilizzare diverse tecniche simili a quelle correntemente usate per il bilanciamento dei conti economici nazionali annuali, oppure procedure derivate sulla falsariga di Rossi (1982) e Di Fonzo (1990). Vale probabilmente la pena notare come questo approccio sia contrario all'usuale *a priori*, diffuso fra i contabili nazionali e gli statistici economici in generale, che le stime migliori si ottengono utilizzando il massimo livello di disaggregazione possibile. In realtà, i nostri risultati (illustrati più in dettaglio in Cainelli e Lupi, 1996) indicano che la scelta del livello di aggregazione contemporanea su cui basare le stime va effettuata valutando caso per caso, non esistendo una regola valida in generale. Inoltre, per quanto esposto nei paragrafi 3 e 4, è essenziale massimizzare l'uso dell'informazione disponibile anche rispetto a tale scelta.

6. Conclusioni

I metodi indiretti di disaggregazione di serie storiche hanno conosciuto, a partire dal contributo di Chow e Lin (1971), un notevole sviluppo che ha consentito a molti istituti nazionali di statistica di conseguire notevoli miglioramenti nella stima trimestrale dei principali aggregati di contabilità nazionale.

Allo sviluppo di queste procedure tuttavia non ha fatto seguito un indirizzo di ricerca in grado di evidenziare come le proprietà statistiche delle serie trimestrali finali possano dipendere, anche adottando gli stessi stimatori, dalle particolari ipotesi che implicitamente

sottendono ogni procedura di trimestralizzazione e come questo si rifletta nell'identificazione e stima modelli econometrici dinamici. Con questo lavoro si è tentato di offrire una prima, anche se ancora parziale, riflessione su questi temi. In particolare si è evidenziato come il livello di aggregazione contemporanea al quale si realizzano le stime trimestrali produca considerevoli effetti sulle proprietà univariate delle serie finali e pertanto determini l'insorgere di strutture dinamiche spurie nelle equazioni di comportamento stimate utilizzando questi dati.

Le indicazioni che si possono trarre dai risultati proposti vanno, a nostro avviso, distinte a seconda che ci si ponga dal punto di vista del produttore di informazioni statistiche o da quello dell'utilizzatore. Infatti, ponendoci nell'ottica del produttore di dati, dall'analisi condotta in precedenza emerge l'importanza di adottare, nella fase di specificazione delle equazioni di trimestralizzazione, un criterio statistico-econometrico che consenta di identificare il grado di aggregazione contemporanea "ottimale" al quale effettuare le stime. Dal punto di vista dell'utilizzatore, viene confermata da questo lavoro non soltanto l'esigenza da parte del ricercatore applicato di approfondire le proprie conoscenze in tema di metodi e fonti statistiche impiegate nella produzione dei dati che egli utilizza, ma anche l'opportunità di ricorrere a metodi robusti di stima. In questo senso sarebbe auspicabile che anche nella ricerca econometrica teorica si riscontrasse una rinnovata attenzione per queste tipologie di problemi.

References

- BANERJEE A. e HENDRY D.F. (1992), "Testing Integration and Cointegration: An Overview", *Oxford Bulletin of Economics and Statistics*, 54, 225-255.
- BARBONE L., BODO G. e VISCO I. (1981), "Costi e profitti nell'industria in senso stretto: un'analisi su serie trimestrali", *Bollettino della Banca d'Italia*, 36, 465-510.
- BASSIE V. (1958), *Economic Forecasting*, Mc Graw-Hill, New York, Mc Graw-Hill.
- BOLLINO C.A. (1994), "L'utilizzo delle tecniche di disaggregazione con indicatori per le stime di serie economiche territoriali", *Documenti di lavoro n.14/94*, Roma, ISPE.
- BREUSCH T.S. e PAGAN A.R. (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics", *Review of Economic Studies*, 47, 239-253.
- CAINELLI G. e LUPI C. (1996), "Aggregazione contemporanea e specificazione econometrica nella stima trimestrale dei conti economici nazionali", ISTAT, mimeo (di prossima pubblicazione).
- CHOW G. e LIN A.L. (1971), "Best Unbiased Interpolation Distribution and Extrapolation of Time Series by Related Series", *Review of Economics and Statistics*, 53, 372-375.
- CHOW G. e LIN A.L. (1976), "Best Linear Unbiased Estimation of Missing Observations in an Economic Time Series", *Journal of the American Statistical Association*, 71, 719-721.
- DA EMPOLI D., SIESTO V. e ANTONELLO P. (1979), *Finanza pubblica e contabilità nazionale trimestrale*. Padova, CEDAM.
- DENTON F.T. (1971), "Adjustment of Monthly and Quarterly Series to Annual Totals: an Approach Based on Quadratic Minimization", *Journal of the American Statistical Association*, 66, 99-102.
- DI FONZO T. (1987), *La stima indiretta di serie economiche trimestrali*. Padova, CLEUP.
- DI FONZO T. (1990), "The Disaggregation of M Time Series When Contemporaneous and Temporal Aggregates are Known", *Review of Economics and Statistics*, 72, 178-182.
- DURLAUF S.N. (1991), "Spectral Based Testing of the Martingale Hypothesis", *Journal of Econometrics*, 50, 355-376.
- GENNARI P. e GIOVANNINI E. (1993), "La stima trimestrale dei conti nazionali mediante modelli a parametri variabili", *Quaderni di Ricerca Economia e Ambiente*, n.5/1993, Roma, ISTAT.
- GHYSELS E. e PERRON P. (1993), "The Effect of Seasonal Adjustment Filters on Testing for a Unit Root", *Journal of Econometrics*, 55, 57-98.
- GREGOIR S. (1994), "Note sur une désagrégation temporelle basé sur des modèles dynamiques simples", lavoro presentato a: "INSEE-EUROSTAT Workshop on Quarterly National Accounts" Parigi (di prossima pubblicazione).
- GRUNFELD Y. e GRILICHES Z. (1960), "Is Aggregation Necessarily so Bad?", *Review of Economics and Statistics*, 17, 1-13.

- GUERRERO V.M. (1990), "Temporal Disaggregation of Time Series: an ARIMA-Based Approach", *International Statistical Review*, 58, 29-36.
- HENDRY D.F. (1987), "Econometric Methodology: A Personal Perspective", in T.F. Bewley (ed.), *Advances in Econometrics Fifth World Congress*, Vol. II, Cambridge, Cambridge University Press.
- HENDRY D.F. e MIZON G.E. (1978), "Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England", *Economic Journal*, 88, 549-563.
- HENDRY D.F. e RICHARD J.-F. (1982), "On the Formulation of Empirical Models in Dynamic Econometrics", *Journal of Econometrics*, 20, 3-33.
- ISTAT (1983), "I conti economici trimestrali in Italia. Anni 1970-1984", supplemento al *Bollettino Mensile di Statistica*, Roma.
- ISTAT (1985), "I conti economici trimestrali. Anni 1970-1984", supplemento al *Bollettino Mensile di Statistica*, n.12, Roma.
- ISTAT (1987), "Miglioramenti apportati ai conti economici trimestrali. Serie con base dei prezzi 1970", *Collana d'Informazione*, n.4, Roma.
- ISTAT (1992), "I conti economici trimestrali con base 1985", *Note e Relazioni*, n.1, Roma.
- JARQUE C.M. e BERA A.K. (1980), "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals", *Economics Letters*, 6, 255-259.
- LUPI C. (1995), "A Monte Carlo Analysis of Two Spectral Tests of the Martingale Hypothesis", mimeo (di prossima pubblicazione).
- LUPI C. (1996), "Stationary Measurement Errors Among Cointegrating Variables", presentato all'Econometric Society European Meeting 1996, Bogazici University.
- LUPI C. e PARIGI G. (1994), "Temporal Disaggregation of Economic Time Series: Some Econometric Issues", lavoro presentato a: "INSEE-EUROSTAT Workshop on Quarterly National Accounts" Parigi (di prossima pubblicazione).
- MARCELLINO M. (1996), "Temporal Disaggregation, Missing Observations, Outliers, and Forecasting: A Unifying Non-Model Based Procedure", European University Institute, mimeo.
- MIZON G.E. (1995), "A Simple Message for Autocorrelation Correctors: Don't", *Journal of Econometrics*, 69, 267-288.
- PAGAN A.R. (1984), "Econometric Issues in the Analysis of Regression with Generated Regressors", *International Economic Review*, 25, 221-247.
- PESARAN M.H., PIERSE R.G. e KUMAR M.S. (1989), "Econometric Analysis of Aggregation in the Context of Linear Prediction Models", *Econometrica*, 57, 861-888.
- PHILLIPS P.C.B. (1988), "Reflections on Econometric Methodology", *Economic Record*, 64, 344-359.
- PHILLIPS P.C.B. e HANSEN B.E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, 57, 99-125.
- ROSSI N. (1982), "A Note on the Estimation of Disaggregated Time Series When the Aggregate is Known", *Review of Economics and Statistics*, 64, 695-696.
- STRAM D.O. e WEI W.W.S. (1986), "A Methodological Note on the Disaggregation of Time Series Totals", *Journal of Time Series Analysis*, 7, 293-302.
- WALLIS K. F. (1974), "Seasonal Adjustment and Relations Between Variables", *Journal of the American Statistical Association*, 69, 18-31.

THE USE OF CHAIN INDICES IN THE NETHERLANDS

Sake de Boer, Jan van Dalen and Piet Verbiest

Statistics Netherlands – Department for National Accounts

1. Introduction

An important purpose of the compilation of national accounts is measuring changes in economic variables. Changes in the production and use of goods and services are caused by a combination of two factors: a change in price and a change in quantity and quality (in national accounts often denoted as: volume change).

Part of the work on national accounts is the decomposition of value changes into volume changes and price changes. The most important purpose is the estimate of real growth rates (volume indices). The second goal is the estimate of price changes (deflators).

An important characteristic of this work when imbedded in a national accounts framework is that volume indices and deflators of various variables and at different levels of aggregation are interrelated in a systematic way. This is achieved by using supply and use tables or input-output tables as an integrating and balancing framework.

Price and volume indices of aggregates are always compiled from price and volume indices of individual goods and services. Direct observation of price and volume changes of aggregates is - by definition - impossible. Nearly all items in the national accounts are aggregates or aggregates of aggregates. Examples of the latter are total household consumption expenditure, total imports and total exports, but also total output and total intermediate consumption of industries. So the question rises how price and volume changes of individual goods can be added to price and volume changes of aggregates. Various methods are available to solve this problem. Different index formulae are available with different weighting schemes. So one has to make a choice. It is beyond the scope of this paper to discuss in depth the theoretical and practical considerations with respect to this choice. For a broader discussion of the theoretical and practical considerations with respect to the choice of index number formulae and weighting schemes the reader is referred to e.g. Allen (1975) and to Al et al. (1986).

Besides aggregates, special attention must be paid to a number of balancing items. Most important of course is GDP. GDP is the difference between two aggregates: total output value and total intermediate consumption of goods and services by all industries. Hence price and volume changes of GDP are the result of the price and volume changes of total output and of total intermediate consumption of all industries together. Since total output and total intermediate consumption are aggregates whose volume and price changes are a result of changes at a detailed level, price and volume changes of GDP are also a result of changes at a detailed level.

The views expressed in this paper are those of the authors, and do not necessarily reflect the views of Statistics Netherlands. The authors would like to thank Gert den Bakker and Steven Keuning for their useful comments on earlier versions of this paper

2. The choice of index number formulae

Most countries - including the Netherlands - use a combination of a Laspeyres volume index and a Paasche price index. For the formulae we refer to annex I. Characteristic for the Laspeyres volume index is that the volume changes of individual goods are weighted together with the value in a former year (the "base year"). Characteristic for the Paasche price index is that the price changes of individual goods are weighted together with their value in the current year. The deflated values derived with this index formula combination can easily be explained as "values in prices of the base year".

Alternative solutions for the weighting problem are e.g. the combination of a Paasche volume index and a Laspeyres price index and the combination of a Fisher volume index and a Fisher price index. All those combinations have in common that there is a decomposition in price and volume changes without residuals. The Fisher volume index is the geometric mean of the Laspeyres and Paasche volume indices; the Fisher price index is the geometric mean of the Laspeyres and Paasche price indices.

3. Fixed or changing weighting schemes

Applying the Laspeyres volume index number formula, volume changes are weighted with the values in a "base year". Next question is which year should be chosen as the base year. Generally speaking there is a choice between a fixed base year and a changing base year. The latter method is often called the "method of chain indices".

With the method of fixed weights for a series of years, the weights are derived from a single year in the past. An advantage of this method is that in longer series of values in constant prices deflated parts of aggregates exactly add up to the deflated aggregate. However, a very serious disadvantage is that volume changes of aggregates are calculated with outdated weights. This disadvantage is especially severe when relative prices change rapidly. As a result economic growth is often significantly overestimated.

Applying the chained Laspeyres volume index means that weights change every year and are derived from the previous year. Since those weights are more up-to-date, a better approximation of the "real world" volume changes is obtained than with the method of fixed weights. Time series results can be obtained by multiplying separately estimated year-to-year volume indices: hence the name chain indices. An important advantage of the chain index method is that the above mentioned overestimations of growth rates are avoided. There is also a disadvantage: in time series in constant prices the deflated parts of an aggregate no longer exactly add up to the deflated aggregate. As a result "mathematical discrepancies" will appear that cannot be removed without disturbing the underlying "actual" volume and price movements.

The contrast between fixed weight methods and changing weight methods is to a certain extent not absolute but only exists in relative terms. Even countries that apply fixed weights, periodically change the weights and choose another base year. In chaining the sub-series they also have a non-additivity problem.

Confronted with the choice between fixed weight and changing weight methods, a preference for yearly changing weights is expressed in recent international guidelines. The following quotations are from SNA and ESA.

SNA-1993 says:

(section 16.41) "If the objective is to measure the actual movements of prices and volumes from period to period indices should be compiled only between consecutive time periods.

Changes in prices and volumes between periods that are separated in time are then obtained by cumulating the short-term movements: i.e., by linking the indices between consecutive periods together to form "chain indices". Such chain indices have a number of practical as well as theoretical advantages. For example, it is possible to obtain a much better match between products in consecutive time periods than between periods that are far apart, given that products are continually disappearing from markets to be replaced by new products, or new qualities. Chain indices are also being increasingly demanded by economists and others for analytical purposes and are being increasingly used for special purpose indices, such as consumer price indices, in order to have indices whose weighing structures are as up-to-date and relevant as possible".

ESA-1995 says:

(section 10.62) "The preferred measure of year to year changes in volume is a Fisher volume index which is defined as the geometric mean of the Laspeyres and the Paasche indices. Changes in volume over longer periods being obtained by chaining, i.e. by cumulating the year to year volume movements.

(section 10.64) "Chain indices that use Laspeyres volume indices to measure changes in volume and Paasche price indices to measure year to year price movements provide acceptable alternatives to Fisher indices."

But ESA-1995 says:

(section 10.65) "Although the preferred measure of volume and price is a chain index, it must be recognized that the lack of additive consistency can be a serious disadvantage for many types of analysis." (section 10.66) "It is therefore recommended that disaggregated constant price data, i.e. direct valuation of current quantities at base-year prices, are compiled in addition to the chain indices for the main aggregates."

4. Short history of deflation practice in the Netherlands

For the afterwar period, Statistics Netherlands has published annual estimates of values in constant prices and volume indices of national accounts variables from 1948 onwards. In addition, for the period 1900 to 1920 annual volume growth rates of national product are available. In 1987, new macro economic data for the period 1921 to 1939 have been published. Recently, new volume changes and deflators for that period have been published as well.

In the afterwar period, two periods can be distinguished with respect to the estimation methods: 1948 to 1980 and 1981 up to present.

1948 to 1980

For this period data have been published for GDP, imports, exports, domestic final demand categories and value added of industry groups. The level of aggregation was high; e.g. the number of industry groups at the beginning was 5 and later on 14. The estimates were carried out in constant prices of fixed base years (successively 1953, 1958, 1963, 1970 and 1975). The deflation was carried out after the estimates in current prices had already been finished. Generally speaking current values were deflated at a rather detailed level (and with various index number formulae). However, the balancing of the estimates took place at a very high level of aggregation.

1981 to present

During the 70's there arose a need for more coherence between statistics of values, volumes and prices. The co-ordination instrument should be the input-output tables. At the outset the ideas went to a system of fixed weight price indices.

In the discussions the most important user of national accounts data, Netherlands Bureau for Economic Policy Analysis (CPB), has played a very important role. CPB has always (with one exception) constructed models based on values in prices of the previous year. During the time that Statistics Netherlands did not yet provide that kind of data, CPB itself compiled the data. Because of the data requirements for their more complicated and more disaggregated models the CPB was and is strongly in favour of a system of chained (Laspeyres) volume indices and chained (Paasche) price indices. This could be realized in a series of input-output tables in prices of the previous year. An important motivation of the CPB's opinion can be found in their bad experiences during the oil-crises in the 70's with the use of fixed weight indices and values in constant prices of a fixed base year in their model at that time.

The result of the discussions has been recorded in a common paper of Balk (Statistics Netherlands) and Eijgenraam (CPB) in 1979. They concluded that there was a need for a dynamically defined price index number; in such a theory more-period-indices are defined as a chain of single-period-indices; for single-period-indices from a theoretical point of view so-called superlative indices are preferable (Diewert, 1978); good approximations are the Tornqvist and Fisher indices; for single-period-indices also Laspeyres and Paasche indices are expected to be good approximations.

The result of the discussions during the late 70's has been that from the reporting year 1981 onwards two important methodological (and also organizational) changes have been introduced in the Dutch national accounts. The first is that for various variables in the national accounts time series of volume indices and deflators have been calculated by chaining the year-to-year indices. The basis for this was the compilation of input-output tables (from 1986 onwards replaced by supply and use tables) in constant prices of the previous year. An important organizational change was that from 1981 onwards, the annual input-output tables in current prices and in prices of the previous year, value changes, volume changes and price changes have all been compiled simultaneously (see De Boer and Broesterhuizen, 1986). As a result, yearly Laspeyres volume changes are available, weighted with the values of the previous year, in combination with Paasche deflators.

The advantage of using chain indices and the relatively minor disadvantage of the resulting non-additivity of constant price series has also been elaborated by Al et al. (1986).

At present, the annual National accounts publications of Statistics Netherlands contain the following time series: values in current prices, annual volume growth rates, volume indices, annual price changes, indices of deflators and values in prices of a base year (recently: 1990). The series of volume indices and deflators are chain indices derived from multiplying the annual changes. The values in prices of a constant base year are calculated with the help of chained volume indices. Hence in this case, sub-series are non-additive to aggregates in our publication.

5. Opinion of users of Netherlands' national accounts

The authors have asked a number of institutions in the Netherlands that use national accounts data for model construction about their opinion on the choice between fixed weight indices and chain indices. In addition their experience with the use of chain indices provided by Statistics Netherlands has been enquired.

The CPB is the most important user of national accounts data in the Netherlands. It played an important role in the discussions preceding the introduction of chain indices in the Netherlands (see section 4). The CPB had and still has a strong preference for chaining single-period- indices based on annually changing weights. In their opinion, values in constant prices for a longer period, based on fixed weights, yield meaningless data. They say that their experience with the use of chain indices is excellent. For them, the non-additivity of values in constant prices is no problem. Discrepancies from aggregations are not eliminated.

Other users that have been approached are Netherlands Economic Institute (NEI), Groningen University and The Netherlands' Central Bank.

It is the opinion of the model builders at NEI that chain indices are a good base for the construction of economic models. However, it is necessary that model builders carefully "tune" the specification of their models to the characteristics of the data. They think that the often mentioned problem of the aggregation discrepancies is not a real problem, but a matter that calls for special attention when presenting the data to the public.

Researchers at the economic faculty of Groningen University don't use constant price series at all. Their macro-economic models require figures on volume changes. So the issue of non-additivity is not relevant for them. If only constant price figures were available, these could only be used if they yielded the correct volume indices.

The department for monetary and economic policy of The Netherlands' Central Bank (DNB) says that indices with annually changing weights have the advantage of a strong connection with actual developments. Hence this department has no problems with Statistics Netherlands' practice.

The department for economic research of DNB uses annual changes as well as values in current and constant prices in their macro-economic models. If there are problems with mathematical discrepancies in model results, the solution mostly is found by an adjustment of the changes in stocks.

6. Results from changing versus fixed weighting schemes and different index number formulae

SNA-1993 states:

(section 16.43) "In general, Laspeyres indices, whether volume or price, tend to increase more (or decrease less) than Paasche indices, but if fixed base indices are replaced by chain indices, the index number spread between Laspeyres and Paasche is likely to be greatly reduced. The relationship between a fixed base index and the corresponding chain index is not always the same, however, as it must depend upon the paths followed by individual prices and quantities over time".

SNA-1993 states:

(section 16.44) "If individual prices and quantities tend to increase or decrease monotonically over time it can be shown that the chain Laspeyres will tend to increase less than the fixed weight Laspeyres while the chain Paasche will tend to increase more than the fixed Paasche. In these circumstances, therefore, chaining will reduce the index number spread, possibly almost eliminating it."

We will illustrate the differences between the results of different methods for the estimation of real annual growth rates on the basis of Dutch data. We present results for two periods. Firstly, we have made alternative estimates for the period 1986-1993, based on

detailed data in the supply and use tables in current prices and in constant prices of the previous year for that period. Secondly, we give some results from the sensitivity analysis that Den Bakker made for the interwar period (1921-1939) of the influence of the choice of the index number formulae and the choice of the weighting scheme (see Den Bakker, 1991).

We discuss the results of alternative estimates shortly in this section. For a more comprehensive discussion and more results we refer the reader to annex II (period 1986-1993) and annex III (period 1921-1939).

Period 1986-1993

In the Dutch supply and use tables, 250 industries, 850 commodities and 10 final demand categories are defined. Supply and use tables are available for the period 1986-1993. For the years 1987-1993, the value in current prices, the value in prices of the previous year, the volume index and the deflator are available for every cell of the supply and the use table. For our calculations, the price and volume indices of the cells are considered as individual indices (of course this is an assumption: almost every cell of the supply and use tables is already an aggregate). At higher levels of aggregation indices are calculated with different index number formulae and with fixed and changing weighting schemes.

Table 1 gives the results of the alternative estimates of the growth rates of GDP and the corresponding time series of volume indices. The reader will find similar results for imports and final expenditure categories in Annex II. The results in the column "Paasche (weights t, 1986)" are obtained by deflation with fixed base (1986) Laspeyres deflators and henceforth will be called "fixed weight Paasche volume indices".

Table 1 – Volume growth rates (t/t-1) and time series of Gross Domestic Product (market prices) according to different index number formulae (1986-1993)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t, 1986)
Volume growth rates (%)					
1987	1,4	1,4	1,4	1,4	1,4
1988	3,4	2,6	2,6	2,6	2,7
1989	4,8	4,7	4,6	4,4	4,6
1990	4,2	4,1	4,0	3,9	3,5
1991	2,3	2,3	2,2	2,2	2,0
1992	2,0	2,0	2,0	2,0	2,0
1993	1,3	0,8	0,8	0,8	0,7
Time series of volume indices (1986=100)					
1987	101,4	101,4	101,4	101,4	101,4
1988	104,9	104,1	104,0	104,0	104,1
1989	109,9	108,9	108,8	108,6	108,9
1990	114,5	113,4	113,1	112,8	112,7
1991	117,1	116,0	115,6	115,3	115,0
1992	119,5	118,3	117,9	117,5	117,3
1993	121,0	119,2	118,9	118,5	118,2

At first sight the differences between the results for different formulae seem to be rather small. However, one must take into consideration that they concern large macro totals. In the Netherlands every 0.1 % of GDP equals about 600 million guilders (280 million ECU). Secondly, if the differences every year show the same sign (- or +) in a time series yearly differ-

ences, however small, can cause serious long term deviations. Thirdly, only one year with a larger deviation can cause large deviations in all succeeding years of a time series.

When discussing the results we will consider the chain Fisher volume indices as the "standard" with which we compare other results. The reason is that Fisher indices can be seen as good approximations of superlative indices that can be seen as ideal indices from a theoretical point of view (see also section 4).

The sequence of the columns of the tables in this paper has been chosen such that a certain following order of results could be expected: highest growth rates are expected with the fixed weight Laspeyres formula and lowest with the fixed weight Paasche formula and the results of the chain formulae are expected to be in between.

The results in table 1 and annex II show that most substantial divergences are with Laspeyres fixed weight indices (first column). The Laspeyres, Fisher and Paasche chain type indices in most cases yield comparable results and the differences between the Paasche fixed weight and chain indices are generally speaking much smaller than between the Laspeyres fixed weight and chain indices. A remarkable and for the statistical practice very important conclusion which can be drawn is that, as ESA states, the Laspeyres chain indices provide good approximations of the "ideal" Fisher chain indices and that the Laspeyres fixed weight indices do less or not at all. This is especially important for the cases where the differences between Laspeyres and Paasche fixed weight indices and between the Laspeyres fixed weight index and the Laspeyres chain index are substantial. E.g. see: GDP (1988 and 1993).

Table 2 – Growth rates ($t/t-1$) of value added of three industry groups according to different index number formulae (1986-1993) (%)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t, 1986)
Agriculture and horticulture					
1987	-4,2	-4,2	-4,5	-4,7	-4,7
1988	5,3	5,2	5,1	5,0	4,4
1989	7,8	7,5	7,1	6,6	6,4
1990	10,0	9,5	9,2	8,9	9,5
1991	3,2	2,8	2,0	1,2	0,8
1992	4,2	4,9	4,4	4,0	4,6
1993	1,6	2,5	1,5	0,6	3,2
Petro-chemical industry					
1987	1,2	1,2	-0,8	-2,7	-2,7
1988	-0,6	5,6	6,3	7,0	8,1
1989	9,6	4,0	7,3	10,8	8,1
1990	9,0	6,7	5,8	4,9	8,0
1991	-2,8	-8,2	-8,3	-8,4	-6,7
1992	-2,8	-1,6	-0,9	-0,1	-2,1
1993	5,1	2,7	2,8	2,9	2,8
Construction of buildings and dwellings					
1987	1,4	1,4	1,4	1,4	1,4
1988	9,5	9,6	9,6	9,7	9,7
1989	2,1	2,2	2,2	2,3	2,3
1990	0,2	0,1	0,1	0,2	0,2
1991	-2,5	-2,4	-2,3	-2,3	-2,3
1992	-3,7	-3,8	-3,7	-3,6	-2,8
1993	-8,1	-7,4	-7,2	-7,0	-6,6

Generally speaking the results in table 1 and annex II also bear out the statement of SNA that Laspeyres volume indices yield higher growth rates than Paasche volume indices and that the differences between the Laspeyres and Paasche chain indices are smaller than between the corresponding fixed weight indices.

Table 2 gives the results of alternative calculations of the annual volume growth rates of value added of three industries which play quite a different role in the Dutch economy and are under the influence of quite different circumstances. Annex II presents similar results for total output and intermediate consumption of these three industries.

Agriculture and horticulture

This is an industry which includes very different activities. As a result of different volume and price developments of output and input in the underlying sub-industries (livestock raising, arable farming and horticulture) the average production structure of total agriculture can change rapidly. Hence, one can expect rather substantial deviations of volume indices calculated with different index number formulae, especially for value added, which is, as the balancing item of output and intermediate consumption, influenced by changes in both output and input.

Annex II shows that, with some exceptions for 1991 and 1993, deviations in output and intermediate consumption are rather modest. Differences in the results are more substantial with value added (table 2). For 1987- 1991 chain indices give the best approximations of the "ideal" Fisher indices. However, remarkable for 1992 and 1993 the chain Laspeyres and the chain Paasche indices act relatively badly.

Petro-chemical industry

This industry is rather sensitive to fluctuations on the international markets. Because it absorbs for the greater part semi-manufactured products from crude oil refineries, it is also relatively sensitive to substantial price fluctuations. The consequence is that the resulting volume growth rates from different index number formulae show larger deviations. It should be noted that the fixed weight Paasche indices as well as the chain Laspeyres and the chain Paasche indices sometimes also show larger deviations. However, table 2 shows that larger deviations again most frequently occur with the fixed weight Laspeyres indices. As could be expected the most substantial deviations are found with value added (a balancing item).

Construction of buildings and dwellings

This industry shows strong fluctuations in volume growth rates. However price movements and changes in production structure are rather gradual. This means that the estimates of volume growth rates are less dependent on the type of index number formula. Annex II shows that gross output and intermediate consumption yield only very small (or no) differences between the results of different formulae. Table 2 shows that the same holds in most cases for value added. However, for 1993 both fixed weight indices show substantial deviations from the three chain indices.

Period 1921-1939

The interwar figures provide a very interesting data set for testing the differences between growth rates arising from the use of different index number formulae and weighting schemes. The reason is that the interwar period is characterized by large fluctuations in economic growth rates and price movements.

Table 3 gives the average annual growth rates of GDP for three sub periods during the interwar period, estimated with different index number formulae. The results are derived from Den Bakker (1991). The reader will find more results in annex III.

Table 3 – Average annual volume growth rates of GDP at market prices according to different index number formulae (Interwar period) (%)

YEAR	Laspeyres (weights 1921)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t, 1921)
1922-1929	5,2	4,7	4,6	4,5	4,5
1930-1934	-1,7	-1,5	-1,5	-1,6	-2,1
1935-1939	4,5	3,8	3,9	4,0	4,5

An interesting point is that during the years 1921-1939 periods of economic growth and decline alternated.

The years 1921/29 and 1934/39 are periods of growth. Table 3 shows that for those years the fixed weight Laspeyres indices give higher volume growth rates than the Paasche indices and also higher growth rates than the chain Laspeyres indices. Further it is quite clear that, just like the results for 1987-1993, the chain Laspeyres indices give a much better approximation of the "ideal" chain Fisher indices than the fixed weight Laspeyres indices.

The years 1929/34 are a period of economic decline. For those years the difference between the fixed weight Laspeyres and Paasche indices is not systematic. However, in most cases fixed weight Laspeyres indices show sharper declines than the chain Laspeyres indices. Again a very important conclusion is that the chain Laspeyres gives a much better approximation of the chain Fisher than the fixed weight Laspeyres indices. Also for this period the differences between the chain Laspeyres, Fisher and Paasche indices are - relatively - small.

7. Non-additivity

When using chain indices to construct time series of values in constant prices of a certain base year, the problem arises that the deflated parts of an aggregate do not add up to the deflated aggregate itself. The magnitude of this "mathematical discrepancy" varies with the way the aggregate is split up in different sub groups. In the Dutch National Accounts the mathematical discrepancies in constant price tables are not eliminated, but these tables are left non-additive. Foot-notes explain the non-additivity to the users.

In the publication National Accounts of the Netherlands (see: CBS, 1995) constant price tables are not presented separately, but always in conjunction with other tables and in a fixed sequence: current prices, volume changes, chain volume indices, values at constant prices (by applying these volume indices). Price changes and chain indices of deflators are given as well.

To give an impression of the order of magnitude, table 5 presents mathematical discrepancies for the supply and disposition of goods and services. The differences can be quite substantial.

Table 5 – Supply and disposition of goods and services at prices of 1986

	1986	1987	1988	1989	1990	1991	1992	1993
1 Domestic product (gross, market prices)	437860	444051	455664	476987	496589	507866	518153	522104
2 Imports (cif)	205440	213989	230257	245783	256104	266656	272161	266465
<i>Mathematical discrepancy</i>	0	-2	-169	-305	-313	-448	-460	38
3 Disposable for final expenditure (gross)	643300	658038	685752	722466	752380	774074	789854	788607
4 Final consumption expenditure								
general government	67940	69729	70720	71809	72985	74103	75347	76475
households	260230	267349	269603	279005	290786	299795	307364	310453
5 Fixed capital formation (gross)								
enterprises	78160	78718	82202	86784	87573	87471	87109	84451
general government	11190	11435	12018	12075	12893	13204	14183	14023
6 Changes in inventories								
enterprises	3830	-118	-57	-1019	-1252	-939	-577	767
8 Exports (fob)	221950	230924	251652	268387	282708	295997	304507	309155
<i>Mathematical discrepancy</i>	0	3	-387	5425	6687	4443	1922	-6717
9 Total final expenditure (gross)	643300	658038	685752	722466	752380	774074	789854	788607

8. Concluding remarks

In this paper we discussed the use of chain indices in the Netherlands. Dutch practice with chain indices is from 1980 onwards. In the discussions preceding the introduction of chain indices, CPB, the most important user of national accounts data, played an important part making a plea for annually changing weights and chain indices. CPB still has a strong preference for chaining one-period growth rates based on annually changing weights. They call their experience with the use of chain indices excellent.

Three other users of national accounts data have been asked about their opinion on the choice between fixed weight indices and chain indices. All of them say that chain indices are a good base for the construction of economic models.

Further:

– it is necessary that model builders carefully tune the specification of their model to the characteristics of the data.

– non-additivity need not be a real problem, if researchers give special attention to the presentation of the data to the public.

– for policy makers it is important that chain indices have a strong correspondence to actual developments.

We also have carried out a sensitivity analysis of the choice of index number formulae and weights in the deflation of national accounts data. For the period 1986-1993 alternative estimates are presented of the volume growth rates of GDP, private consumption of households, gross fixed capital formation, imports and exports, and for output, intermediate consumption and value added of three industries. For the period 1921-1939 alternative estimates are given of the volume growth rates of GDP.

Considering the chain Fisher indices as a standard, most important conclusions that can be drawn from those results are:

– in nearly all cases both chain Laspeyres and chain Paasche volume growth rates and indices are good approximations of chain Fisher indices. This affirms the statement in ESA-1995 (section 10.64) that chain Laspeyres volume indices (in combination with chain Paasche price indices) provide acceptable alternatives to chain Fisher indices.

– the results show that most substantial deviations from the chain Fisher indices are found with the fixed weight Laspeyres formula. This is very important since fixed weight Laspeyres volume indices often are used in constant price estimation of national accounts data.

– fixed weight Paasche indices generally speaking act better than fixed weight Laspeyres indices, but in some cases there are also severe deviations from the chain indices.

References

AL, P.G., B.M. BALK, S. DE BOER AND G.P. DEN BAKKER, The use of chain indices for deflating the national accounts, *Statistical Journal of the United Nations*, volume 4, number 4, July 1987 p. 347-368.

ALLEN, R.G.D., (1975), *Index numbers in theory and practice*, Aldine Publishing Company, Chicago.

BAKKER, G.P. den (1991), The choice of index number formulae and weights in the national accounts (a sensitivity analysis based on macro-economic data for the interwar period), *National Accounts Occasional Paper nr. NA-044*, Voorburg, Statistics Netherlands.

BALK, B.M. AND C.J.J. EIJGENRAAM (1979), *Aspecten van een statische en dynamische definitie van de prijsindex*, Advisory Committee for the Price Statistics of Statistics Netherlands.

BOER, S. DE AND G.A.A.M. BROESTERHUIZEN, (1986), The simultaneous compilation of current price and deflated input-output tables, National Accounts Occasional Paper nr. NA-013, Voorburg, Statistics Netherlands.

CBS, National Accounts of the Netherlands (1995), Voorburg, Statistics Netherlands.

DIEWERT, W.E. (1978), Superlative Index Numbers and consistency in Aggregation, *Econometrica*, vol. 46, no 4.

EUROSTAT (1996), European System of Accounts (ESA 1995).

UNITED NATIONS ET AL. (1993), System of National Accounts (SNA 1993).

Annex I. Index number formulae

1) Fixed weights

Laspeyres volume index

$$\frac{\sum p_0 q_t}{\sum p_0 q_0} = \frac{\sum p_0 q_0 \frac{q_t}{q_0}}{\sum p_0 q_0}$$

Paasche volume index

$$\frac{\sum p_t q_t}{\sum p_t q_0} = \frac{\sum p_t q_t}{\sum p_t q_t \frac{q_0}{q_t}}$$

Fischer volume index

$$\sqrt{\text{Laspeyres} * \text{Paasche}}$$

2) Annual changing weights

Laspeyres volume index

$$\frac{\sum p_{t-1} q_t}{\sum p_{t-1} q_{t-1}} = \frac{\sum p_{t-1} q_{t-1} \frac{q_t}{q_{t-1}}}{\sum p_{t-1} q_{t-1}}$$

Paasche volume index

$$\frac{\sum p_t q_t}{\sum p_t q_{t-1}} = \frac{\sum p_t q_t}{\sum p_t q_t \frac{q_{t-1}}{q_t}}$$

Fischer volume index

$$\sqrt{\text{Laspeyres} * \text{Paasche}}$$

Annex II. Results 1986-1993

Table I gives the results of the alternative estimates of the growth rates of GDP, imports and final demand categories. Table II gives the corresponding time series of volume indices.

When discussing the results as in section 6 we will consider the chain Fisher volume indices as the "standard" with which we compare other results.

The results in Table I show that most substantial divergences occur with Laspeyres fixed weight indices. The Laspeyres, Fisher and Paasche chain type indices in most cases yield comparable results and the differences between the Paasche fixed weight and chain indices are generally speaking much smaller than between the Laspeyres fixed weight and chain indices. A remarkable and for the statistical practice very important conclusion which can be drawn from Table II is that, as ESA states, the Laspeyres chain indices provide good approximations of the "ideal" Fisher chain indices and that the Laspeyres fixed weight indices do less or not at all. This is especially important for the cases where the differences between Laspeyres and Paasche fixed weight indices and between the Laspeyres fixed weight index and the Laspeyres chain index are substantial. See: GDP (1988 and 1993), Consumption of households (1991), Capital formation (1992 and 1993), Exports (1988, 1989, 1990 and 1993), Imports (1990, 1991 and 1993).

Generally speaking the results in Table I also bear out the statement of SNA that Laspeyres volume indices yield higher growth rates than Paasche volume indices and that the differences between the Laspeyres and Paasche chain indices are smaller than between the corresponding fixed weight indices.

Table II presents the time series of volume indices corresponding with the annual growth rates in table I. It will not be surprising that the conclusions that can be drawn from the results in table I and II are quite similar.

Table I – Growth rates (t/t-1) according to different index number formulae (1986-1993): macro totals (%)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t, 1986)
Gross Domestic Product (market prices)					
1987	1,4	1,4	1,4	1,4	1,4
1988	3,4	2,6	2,6	2,6	2,7
1989	4,8	4,7	4,6	4,4	4,6
1990	4,2	4,1	4,0	3,9	3,5
1991	2,3	2,3	2,2	2,2	2,0
1992	2,0	2,0	2,0	2,0	2,0
1993	1,3	0,8	0,8	0,8	0,7
Consumption expenditure of households					
1987	2,7	2,7	2,7	2,7	2,7
1988	0,7	0,8	0,8	0,8	0,8
1989	3,5	3,5	3,5	3,5	3,5
1990	4,2	4,2	4,2	4,2	4,1
1991	3,3	3,1	3,1	3,1	2,9
1992	2,5	2,5	2,5	2,5	2,4
1993	1,1	1,0	1,0	0,9	0,9
Gross fixed capital formation					
1987	0,9	0,9	0,9	0,9	0,9
1988	4,6	4,5	4,5	4,5	4,5
1989	5,0	4,9	4,9	4,9	4,8
1990	1,7	1,6	1,6	1,5	1,4
1991	0,2	0,2	0,2	0,2	-0,1
1992	0,8	0,6	0,6	0,6	0,4
1993	-2,5	-2,8	-2,8	-2,9	-3,0
Exports					
1987	4,0	4,0	3,9	3,7	3,7
1988	10,3	9,0	9,0	9,0	9,1
1989	6,8	6,6	6,5	6,3	6,2
1990	5,4	5,3	5,3	5,2	5,0
1991	4,9	4,7	4,7	4,6	4,7
1992	3,1	2,9	2,8	2,8	2,9
1993	2,1	1,5	1,4	1,4	1,4
Imports					
1987	4,2	4,2	4,0	3,9	3,9
1988	7,7	7,6	7,5	7,4	7,4
1989	6,6	6,7	6,7	6,6	6,4
1990	4,4	4,2	4,2	4,1	4,0
1991	4,3	4,1	4,1	4,1	3,9
1992	2,2	2,1	2,0	2,0	2,1
1993	-1,5	-2,1	-2,2	-2,3	-2,4

Table II – Time series of volume indices according to different index number formulae (1986=100)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t,1986)
Gross Domestic Product (market prices)					
1987	101,4	101,4	101,4	101,4	101,4
1988	104,9	104,1	104,0	104,0	104,1
1989	109,9	108,9	108,8	108,6	108,9
1990	114,5	113,4	113,1	112,8	112,7
1991	117,1	116,0	115,6	115,3	115,0
1992	119,5	118,3	117,9	117,5	117,3
1993	121,0	119,2	118,9	118,5	118,2
Consumption expenditure of households					
1987	102,7	102,7	102,7	102,7	102,7
1988	103,5	103,6	103,6	103,6	103,5
1989	107,1	107,2	107,2	107,1	107,1
1990	111,6	111,7	111,7	111,6	111,5
1991	115,3	115,2	115,1	115,0	114,7
1992	118,2	118,1	118,0	117,9	117,4
1993	119,5	119,3	119,2	119,1	118,5
Gross fixed capital formation					
1987	100,9	100,9	100,9	100,9	100,9
1988	105,5	105,5	105,5	105,5	105,5
1989	110,8	110,7	110,7	110,7	110,6
1990	112,6	112,4	112,4	112,4	112,1
1991	112,8	112,7	112,6	112,6	112,0
1992	113,8	113,4	113,3	113,2	112,4
1993	111,0	110,2	110,1	109,9	109,0
Exports					
1987	104,0	104,0	103,9	103,7	103,7
1988	114,8	113,4	113,2	113,0	113,1
1989	122,6	120,9	120,5	120,1	120,1
1990	129,1	127,4	126,9	126,4	126,1
1991	135,4	133,4	132,8	132,2	132,0
1992	139,5	137,2	136,6	135,9	135,9
1993	142,5	139,3	138,5	137,8	137,7
Imports					
1987	104,2	104,2	104,0	103,9	103,9
1988	112,2	112,1	111,8	111,6	111,6
1989	119,7	119,6	119,3	118,9	118,8
1990	124,9	124,7	124,2	123,8	123,5
1991	130,3	129,8	129,3	128,9	128,3
1992	133,2	132,5	132,0	131,4	131,0
1993	131,2	129,7	129,1	128,5	127,9

Tables IIIA, IIIB and IIIC give the results of alternative calculations of the annual volume growth rates of total output, total intermediate consumption and value added of three industries which play quite a different role in the Dutch economy and are under the influence of quite different circumstances.

Agriculture and horticulture

This is an industry which includes very different activities. As a result of different volume and price developments of output and input in the underlying sub-industries (livestock raising, arable farming and horticulture) the average production structure of total agriculture can change rapidly. Hence, one can expect rather substantial deviations of volume indices calculated with different index number formulae, especially for value added, which is, as the balancing item of output and intermediate consumption, influenced by changes in both output and input. Table IIIA shows that, with some exceptions for 1991 and 1993, deviations in output and intermediate consumption are rather modest. Differences in the results are more substantial with value added. For 1987-1991 chain indices give the best approximations of the "ideal" Fisher indices. However, remarkable for 1992 and 1993 the chain Laspeyres and the chain Paasche indices act relatively badly.

Table IIIA – Growth rates (t/t-1) of gross output, intermediate consumption and value added of three industry groups according to different index number formulae (1986-1993) (%)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t, 1986)
Agriculture and horticulture					
<i>Gross output</i>					
1987	-1,4	-1,4	-1,6	-1,8	-1,8
1988	1,9	2,0	1,9	1,8	1,7
1989	3,3	3,4	3,2	3,0	2,9
1990	3,8	3,5	3,4	3,2	3,5
1991	2,2	2,0	1,6	1,3	1,1
1992	2,0	2,4	2,1	1,8	1,8
1993	0,5	0,8	0,4	-0,1	0,6
<i>Intermediate consumption</i>					
1987	0,7	0,7	0,6	0,5	0,5
1988	-0,5	-0,4	-0,5	-0,5	-0,4
1989	0,0	0,2	0,1	0,1	0,2
1990	-1,3	-1,6	-1,5	-1,5	-1,4
1991	1,2	1,2	1,3	1,3	1,4
1992	0,0	0,1	0,0	-0,1	0,1
1993	-0,5	-0,6	-0,6	-0,7	-0,3
<i>Value added (basic prices)</i>					
1987	-4,2	-4,2	-4,5	-4,7	-4,7
1988	5,3	5,2	5,1	5,0	4,4
1989	7,8	7,5	7,1	6,6	6,4
1990	10,0	9,5	9,2	8,9	9,5
1991	3,2	2,8	2,0	1,2	0,8
1992	4,2	4,9	4,4	4,0	4,6
1993	1,6	2,5	1,5	0,6	3,2

Petro-chemical industry

This industry is rather sensitive to fluctuations on the international markets. Because it absorbs for the greater part semi-manufactured products from crude oil refineries, it is also relatively sensitive to substantial price fluctuations.

The consequence is that the resulting volume growth rates from different index number formulae show larger deviations (see: table IIIB). It should be noted that the fixed weight Paasche indices as well as the chain Laspeyres and the chain Paasche indices sometimes also show larger deviations. However, table IIIB shows that larger deviations again most frequently occur with the fixed weight Laspeyres indices. As could be expected the most substantial deviations are found with value added (a balancing item).

Table IIIB – Growth rates (t/t-1) of gross output, intermediate consumption and value added of three industry groups according to different index number formulae (1986-1993) (%)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t,1986)
Petro-chemical industry					
<i>Gross output</i>					
1987	4,8	4,8	3,9	3,1	3,1
1988	8,9	9,7	9,7	9,7	9,7
1989	1,7	0,6	0,8	0,9	0,2
1990	3,1	3,3	2,9	2,6	3,2
1991	-4,1	-5,3	-5,4	-5,4	-5,3
1992	-2,2	-2,6	-2,6	-2,6	-3,3
1993	-0,8	-1,8	-1,8	-1,9	-2,2
<i>Intermediate consumption</i>					
1987	6,2	6,2	5,9	5,7	5,7
1988	12,5	11,5	11,3	11,2	11,5
1989	-0,9	-1,3	-2,3	-3,4	-4,1
1990	0,9	1,5	1,7	1,8	1,6
1991	-4,6	-4,3	-4,4	-4,5	-4,6
1992	-2,0	-2,9	-3,0	-3,2	-3,5
1993	-3,1	-2,9	-3,0	-3,1	-3,4
<i>Value added (basic prices)</i>					
1987	1,2	1,2	-0,8	-2,7	-2,7
1988	-0,6	5,6	6,3	7,0	8,1
1989	9,6	4,0	7,3	10,8	8,1
1990	9,0	6,7	5,8	4,9	8,0
1991	-2,8	-8,2	-8,3	-8,4	-6,7
1992	-2,8	-1,6	-0,9	-0,1	-2,1
1993	5,1	2,7	2,8	2,9	2,8

Construction of buildings and dwellings

This industry shows strong fluctuations in volume growth rates. However price movements and changes in production structure are rather gradual. This means that the estimates of volume growth rates are less dependent on the type of index number formula. Table IIIC shows that gross output and intermediate consumption yield only very small (or no) differ-

ences between the results of different formulae. The same holds in most cases for value added. However for 1993 both fixed weight indices show substantial deviations from the three chain indices.

Table III C – Growth rates (t/t-1) of gross output, intermediate consumption and value added of three industry groups according to different index number formulae (1986-1993) (%)

YEAR	Laspeyres (weights 1986)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t,1986)
Construction of buildings and dwellings					
<i>Gross output</i>					
1987	5,1	5,1	5,1	5,1	5,1
1988	12,2	12,2	12,3	12,3	12,3
1989	4,6	4,6	4,6	4,6	4,6
1990	0,5	0,5	0,5	0,5	0,5
1991	0,5	0,5	0,5	0,5	0,6
1992	-2,2	-2,2	-2,2	-2,2	-2,2
1993	-4,4	-4,4	-4,4	-4,4	-4,4
<i>Intermediate consumption</i>					
1987	6,7	6,7	6,7	6,7	6,7
1988	13,4	13,4	13,4	13,4	13,4
1989	5,6	5,6	5,6	5,5	5,5
1990	0,6	0,6	0,6	0,6	0,5
1991	1,7	1,7	1,7	1,7	1,8
1992	-1,6	-1,6	-1,6	-1,7	-1,8
1993	-3,0	-3,1	-3,2	-3,2	-3,3
<i>Value added (basic prices)</i>					
1987	1,4	1,4	1,4	1,4	1,4
1988	9,5	9,6	9,6	9,7	9,7
1989	2,1	2,2	2,2	2,3	2,3
1990	0,2	0,1	0,1	0,2	0,2
1991	-2,5	-2,4	-2,3	-2,3	-2,3
1992	-3,7	-3,8	-3,7	-3,6	-2,8
1993	-8,1	-7,4	-7,2	-7,0	-6,6

Annex III. Results 1921-1939

Table IVA gives the annual volume growth rates of GDP during the interwar period, estimated with different index number formulae. The results are derived from Den Bakker (1991).

Generally speaking for the interwar period the results from the alternative index number formulae show more differences than for the years 1986-1993. The reason is of course that relative price and volume changes were more substantial during the interwar period.

An interesting point is that during the years 1921-1939 periods of economic growth and decline alternated.

The years 1922 to 1929 and 1935 to 1937 are periods of steady growth. Table IVA shows that for those years (except 1929) the fixed weight Laspeyres indices give higher annual volume growth rates than the Paasche indices and also higher growth rates than the chain Laspeyres indices. Further it is quite clear that, just as for 1987-1993, the chain Laspeyres indices give a much better approximation of the "ideal" chain Fisher indices than the fixed weight Laspeyres indices. Besides it is remarkable that for 1924, 1936 and 1937 the fixed weight Paasche index as well as the fixed weight Laspeyres index is higher than all three chain indices.

The years 1930 to 1934 (except 1933) are a period of economic decline. For those years the difference between the fixed weight Laspeyres and Paasche indices is not systematic. However, in most cases fixed weight Laspeyres indices show sharper declines than the chain Laspeyres indices. Again a very important conclusion is that the chain Laspeyres gives a much better approximation of the chain Fisher than the fixed weight Laspeyres indices. Also for this period the differences between the chain Laspeyres, Fisher and Paasche indices are -relatively- small.

Table IVA – Growth rates ($t/t-1$) of GDP at market prices according to different index number formulae (Interwar period) (%)

YEAR	Laspeyres (weights 1921)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t,1921)
1922	6,5	6,5	6,1	5,6	5,6
1923	2,7	2,1	2,0	1,9	1,7
1924	8,1	7,2	7,3	7,5	7,8
1925	3,9	3,6	3,3	3,1	3,1
1926	7,5	6,8	6,7	6,6	6,0
1927	5,1	4,6	4,6	4,5	4,2
1928	5,3	4,9	4,8	4,8	4,7
1929	2,3	2,4	2,3	2,3	2,9
1930	-1,6	-1,2	-1,1	-1,1	-1,7
1931	-5,4	-4,6	-4,5	-4,4	-5,4
1932	-1,2	-0,9	-1,1	-1,3	-2,3
1933	-0,2	0,4	0,4	0,3	0,6
1934	-0,2	-1,1	-1,3	-1,4	-1,5
1935	4,4	2,7	3,0	3,2	2,9
1936	6,4	5,5	5,3	5,2	5,6
1937	7,2	5,7	5,8	5,9	7,0
1938	-4,4	-3,2	-2,9	-2,7	-2,9
1939	9,1	8,1	8,3	8,6	9,8

Table IVB – Volume indices of GDP at market prices according to different index number formulae (Interwar period), 1921=100

YEAR	Laspeyres (weights 1921)	Laspeyres (chain)	Fisher (chain)	Paasche (chain)	Paasche (weights t, 1921)
1922	106,5	106,5	106,1	105,6	105,6
1923	109,4	108,7	108,2	107,7	107,4
1924	118,3	116,5	116,2	115,8	115,8
1925	122,9	120,7	120,0	119,3	119,4
1926	132,1	128,9	128,1	127,3	126,6
1927	138,8	134,9	133,9	132,9	131,9
1928	146,1	141,4	140,4	139,3	138,1
1929	149,4	144,8	143,7	142,5	142,1
1930	147,0	143,1	142,1	141,0	139,7
1931	139,0	136,5	135,7	134,8	132,1
1932	137,3	135,3	134,2	133,1	129,1
1933	137,0	135,9	134,6	133,5	129,9
1934	136,7	134,3	133,0	131,6	128,0
1935	142,7	138,0	136,9	135,9	131,7
1936	151,9	145,6	144,2	142,9	139,1
1937	162,9	153,8	152,5	151,3	148,9
1938	155,7	148,9	148,1	147,2	144,6
1939	169,9	161,0	160,4	159,8	158,8

Table IVB shows the time series of volume indices of GDP for the years 1921 to 1939. Despite strong price and volume fluctuations during that period the chain Laspeyres and Paasche indices show much similarity with the “ideal” chain Fisher indices. The highest deviation is 1.4%. In contrast, the series of fixed weight Laspeyres indices yields very high deviations from the Fisher indices and for many years of the series. Highest deviations are for 1937 (10,4%) and for 1939 (9,5%). The series of fixed weight Paasche indices shows higher deviations than the chain indices (highest for 1935: 5,2%), but the deviations are less than those for the fixed weight Laspeyres.

MACRO MODELING WITH CHAIN-TYPE GDP

Chris Varvares, Joel Prakken and Lisa Guirl

Macroeconomic Advisers, LLC

1. Introduction

Any empirical investigation would benefit from an improvement in the quality of the underlying data. In macroeconomic modeling and forecasting, increasing the timeliness, and reducing the sample bias and variance of the data upon which these activities depend, will help to improve both our understanding of the structure of the economy and its behavior through time, and, by extension, our ability to forecast it and pursue appropriate macroeconomic policies. Unfortunately, in most cases, improvements in economic data do not come without a cost. Traditionally, such improvements required the collection of more data, of different types, from an increased number of collection points, as well as the application of more advanced processing, and so on, all of which involve a sometimes significant cost. Moreover, governmental statistical agencies are being asked to do more with less. These times of rapid technological and structural change may be raising the importance of high quality economic data for both economic policy-makers and private-sector practitioners, while coincidentally tight governmental budgets are resulting in the resources provided to the statistical agencies being frozen or reduced.

Fortunately, the old axiom “work smarter, not harder” may have some relevance. The collection and processing of economic data is likely to benefit significantly from “working smarter” until such time that additional resources become available which also allow us to capture the benefits of “working harder.” Specifically, there are two ways in which working smarter will pay big dividends with tolerable additional costs. *First*, the re-engineering of economic data, as in the switch from fixed-weight to chain-type GDP, results in a dramatic improvement in the quality of the aggregate data, with, in principle, no additional data being required. *Second*, the same technological advances which are the greatest motivating factor requiring the re-engineering of our aggregate data, will also, in time, improve the quantity, quality, and timeliness of the raw data, while lowering its collection cost. Private business is rapidly applying electronic data collection technologies to its own data. Once our statistical agencies can tap into these data networks, near real-time economic data drawn from huge samples will not be far behind. The subject of this paper, however, is to address one small, but important, part of the first way to higher quality data, namely the adaptations required in macro modeling and forecasting in switching to chain-type GDP.

Far and away, the primary reason to re-engineer aggregate economic statistics and replace fixed-weight measures with chain-type measures is to eliminate the substitution bias that can afflict fixed-weight indices. This becomes especially important when rapid and sus-

The authors are President, Chairman, and Economist, respectively, at Macroeconomic Advisers, LLC, St. Louis, Missouri, USA. They can be reached by e-mail at: mail@macroadvisers.com

tained changes in relative prices occur, as is now the case with computers and other high-technology items. Macroeconomic analysis and modeling requires the ability to accurately decompose changes in nominal aggregates into the price and real components. Fixed-weight indices, as we shall see below, are simply not well suited to this task, and are easily dominated by chain-type measures. We should offer an even stronger statement on this point, namely, fixed-weight indices can be *significantly misleading* when it comes to the measurement of real GDP and other key aggregates. As a result, despite the considerable upheaval in the user community that a switch to chain-type GDP may cause, it is well worth it.

Fixed-weight measures, as the name implies, employ the weights of a specific time period in the identities that aggregate the detailed components. Statisticians have long recognized the problem of substitution bias, and they also recognized that it becomes more severe the further the distance in time from the period supplying the fixed weights. As a result, it has been customary to periodically "re-weight" the data using a more recent time period as the source of the fixed weights. This reduces the degree of the substitution bias in the "current" data, but increases it in data further back in time. Thus economic history changes each time this periodic updating of the weighting period occurs. In contrast, chain-type measures are not dependent upon the weights of any *specific* period because they depend on the weights of *every* period and, therefore, there is no need to re-weight them periodically. Chain-type indices are, in this sense, absolute, and independent of the base year. Importantly, comparisons across business cycles, and the measurement of important trends are not subject to revision every time the statistical agency judges that it is time to re-weight the index.

Lest we leave the impression that the switch to chain-type GDP in the US was met with ready acceptance, we should point out some of the serious problems it has caused as well. Chain-type GDP, as it is implemented in the US, is an approximation of a Fisher Ideal index. The immediate and most difficult challenge to anyone accustomed to working with the fixed-weight data is, principally, the nonadditivity of the real components. For generations of macroeconomists who learned in the first chapter of their first book on macroeconomics that $GNP = C + I + G + X - M$, the news that this identity would no longer "ident" came as a rude shock. There are other requirements imposed on users of the data or problems that must be overcome as a result of the switch to chain-type GDP. Some either are a property of all chain-type measures, or arise out of the details of the particular implementation of the Fisher Ideal index in the US. Including the nonadditivity problem just mentioned, these are:

- 1.1) **Nonadditivity:** Real components of GDP no longer sum to GDP.
- 1.2) **Re-estimation:** All behavioral equations employing the data must be re-estimated.
- 1.3) **Different Weighting Schemes for Deep History and Recent History:** The data is published using two different weighting methods depending on whether the data is recent history or deep history.
- 1.4) **PQ = Y\$:** The particular weighting scheme employed in the US results in the quarterly data not being true Fisher Ideal indices. As a result, the product of price times real does not equal nominal.
- 1.5) **Differences in Level of Disaggregation:** The construction of chain-type aggregates utilizes the lowest level of detail so that a user aggregating two subaggregates, even using the correct aggregation scheme, will never get a result that exactly matches the official figures.
- 1.6) **Annual Averaging Does Not Hold:** The particular weighting scheme utilized by BEA has the property that the quarterly data do not aggregate across time, i.e., the average of any GDP component across the four quarters of a year does not necessarily equal the annual value for that component.

All of these properties pose considerable mechanical difficulties to anyone accustomed to frequently manipulating large amounts of both detailed and aggregated macroeconomic time series, and for macro modelers especially. Dealing with these difficulties is not impossible, but it is not easy either. Done right, it requires modelers to correctly complete three major tasks: 1) re-estimate all the behavioral relationships, 2) re-structure the model to be able to accommodate the needs of the various users of the model for alternative weighting schemes, while insuring that the simulation properties are independent of the aggregation scheme, 3) re-structure the aggregating relationships in the model to reflect the new method(s) of calculating all real quantities, prices, and nominals.

Each of the major tasks will be addressed below, in sections 3, 4, and 5 respectively. But first, in Section 2 we provide some background on the issue of fixed-weight versus chain-type GDP and the specifics of the implementation of Chain-Type GDP in the US. In section 6 we offer some concluding remarks.

2. Background: implementation of chain-type GDP in the US

The reason any of this is of interest is the necessity of decomposing growth in nominal magnitudes across time between that part arising from changes in prices and that part arising from changes in quantities. This is clearly not a problem at the level of an elemental component of GDP. The nominal magnitude is known, the price index is known, so the real quantity (denominated in dollars) is just the nominal quantity divided by the price index. However, this becomes an issue as soon as it is necessary to *aggregate* two or more components with different price indices.

2.1. Measuring Fixed-Weight GDP and Prices

Begin by considering a simple economy with only consumer and investment goods. In time period t , $Q_{C,t}$ is the physical quantity of consumer goods produced, $Q_{I,t}$ is the physical quantity of investment goods produced, while $p_{C,t}$ and $p_{I,t}$ are their respective prices. The level of nominal consumption is $C\$_t = p_{C,t}Q_{C,t}$, the level of nominal investment is $I\$_t = p_{I,t}Q_{I,t}$, and nominal GDP ($Y\$_t$) is computed by summing nominal consumption and investment:

$$Y\$_t = C\$_t + I\$_t = p_{C,t}Q_{C,t} + p_{I,t}Q_{I,t} \quad [1]$$

Measured in 1987 prices, the level of real consumption in period t is $C87_t = p_{C,87}Q_{C,t}$, the level of real investment is $I87_t = p_{I,87}Q_{I,t}$. We can define price indices, P , equal to 1 in the base year, such that $P_{C,t} = p_{C,t}/p_{C,87}$ and $P_{I,t} = p_{I,t}/p_{I,87}$. The real quantities can be re-written as $C87_t = p_{C,87}Q_{C,t}$, $P_{C,t}/P_{C,t} = C\$_t/P_{C,t}$, i.e. as the ratio of the nominal magnitude to the price index.

Thus expression [1] can be re-written as:

$$Y\$_t = C\$_t + I\$_t = p_{C,t}Q_{C,t} + p_{I,t}Q_{I,t} = P_{C,t}C87_t + P_{I,t}I87_t \quad [1']$$

This is the more familiar form where physical quantities are replaced with dollar-denominated real magnitudes, and prices are replaced with price indices which have a value of 1 in the base period.

Real GDP (Y87) is computed by summing real consumption and investment:

$$Y87_t = C87_t + I87_t = P_{C,87}Q_{C,t} + P_{I,87}Q_{I,t} \quad [2]$$

Finally, the implicit deflator of Gross Domestic Product (P87t) is computed as the ratio of the current value of quantities to their value in 1987 prices:

$$P87_t = (P_{C,t}Q_{C,t} + P_{I,t}Q_{I,t}) / (P_{C,87}Q_{C,t} + P_{I,87}Q_{I,t}) \quad [3]$$

$$= (C\$_t + I\$_t) / (C87_t + I87_t)$$

This is how the Bureau of Economic Analysis (BEA) in the US constructed fixed-weight real GDP and its companion price level, except that BEA keeps track not just of two quantities and prices, but over one thousand. For annual data, "t" is defined as a calendar year; for quarterly data, "t" is a calendar quarter. Yearly data can be computed either by averaging the individual prices and quantities annually before aggregation or, with the same result, averaging quarterly data for Y87 and P87 annually after aggregation. Given the definitions of real output and the aggregate price level, it is identically true that nominal GDP is the product of real fixed-weight GDP and the implicit deflator:

$$Y\$_t = P87_t Y87_t \quad [4]$$

At the time of the last benchmark revisions using fixed-weight data, these constructs were "re-based" from 1982 to 1987, so that the price indices equaled 1 in 1987, and in the computation of real GDP, quantities were valued at the prices that obtained in 1987 (e.g. $I87_t = P_{I,87}Q_{I,t}$), instead of in 1982. That is, the components of real GDP were *re-weighted* using 1987 prices. With all price weights equal to 1, expression [2] makes clear that additivity holds throughout.

In the fixed-weight approach to aggregating components of real GDP, the additivity property makes it quite natural to think in terms of the *levels* of real GDP and its components. Strictly speaking, however, Y87, C87, and I87 are indices; and, as the Chief Statistician of the BEA is fond of saying, "there is nothing real about real GDP." Nevertheless, this hasn't stopped users of fixed-weight GDP from thinking it is *real*, and thinking of GDP as fundamentally a "level" concept.

To see how the growth rates of real fixed-weight GDP are affected by the periodic re-basing, begin from expression [2] and calculate the ratio of this period's real GDP to last period's real GDP:

$$Y87_t/Y87_{t-1} = (P_{C,87}Q_{C,t} + P_{I,87}Q_{I,t}) / (P_{C,87}Q_{C,t-1} + P_{I,87}Q_{I,t-1}) \quad [5]$$

With some manipulation, [5] can be re-written as:

$$Y87_t/Y87_{t-1} = \beta_{C,t-1} (Q_{C,t} / Q_{C,t-1}) + \beta_{I,t-1} (Q_{I,t} / Q_{I,t-1}), \quad [6]$$

where the weights are defined as the lagged share of real consumption in real GDP, $\beta_{C,t-1} = C87_{t-1}/Y87_{t-1}$, and the lagged share of real investment in real GDP, $\beta_{I,t-1} = I87_{t-1}/Y87_{t-1}$. Since $C\$ = P_{C,87}$, the real share can be written as $\beta_{C,t-1} = (C\$_{t-1}/Y\$_{t-1}) / (P87_{t-1}/P_{C,t-1})$.

There are three things to note about expression [6]. First, the period-to-period change in the physical quantities is independent of the base year. Second, in the weights, $\beta_{C,t-1}$ and $\beta_{I,t-1}$, the nominal component share in GDP is independent of the base year. Third, the ratios of $P87_{t-1}/P_{C,t-1}$ and $P87_{t-1}/P_{I,t-1}$, *do depend on the base year*. For example, in the base period,

the ratio is one, but in any other period the ratio could differ from one as a result of relative price changes. Simply moving the base period, changes $P_{t-1}/P_{C,t-1}$ at every point in time. As a result, the weights, $\beta_{C,t-1}$ and $\beta_{I,t-1}$, also depend on the base period. This is why the periodic re-basing of fixed-weight GDP results in sometimes dramatic revisions in measured growth rates of GDP.

Note also that the weight of the i th component depends inversely on the relative price of the i th component. As the price of the i th component falls relative to the overall price, P_{87} , the weight of the i th component increases in the calculation of the growth of real GDP. Thus, in situations where supply-induced declines in relative prices, such as with computer and other high-tech products today, raise the relative growth of the associated real components, the weight of the fastest growing components increase with time. This is the crux of the substitution bias problem. As we shall see below, chain-type GDP avoids this difficulty because the growth of real GDP turns out to be the weighted average of the growth of the physical quantities of its components, where the weights are shares of the i th *nominal* component to *nominal* GDP, which is, of course, invariant to the base period.

2.2. Chain-type GDP: the basic approach

The fundamental difference between fixed-weight GDP and chain-type GDP is that fixed-weight GDP values product at the prices of some fixed period, while chain-type GDP values product at "current" prices. Obviously, one can't directly calculate the level of real GDP while valuing product at current prices, since this is just nominal GDP, which we already observe directly. One can, however, calculate the *change* in real GDP valued at current prices. Thus, chain-type GDP is fundamentally a concept about the *change* in real GDP.

In the discussion that follows, $G_t = X_t/X_{t-1}$ is referred to as the growth in X . Typically, for macroeconomic time series, G will have a value in the neighborhood of 1, something slightly greater than 1 if you're lucky, and something less than 1 if you're not. This leads directly to the concept of "chaining." If you can't compute the level of real GDP directly, but you can compute the growth in GDP, then the current level of the index can always be computed by the construction of a *chain* of growth terms from an initial index value.

$$Y_t = Y_0 G_1 G_2 G_3 \dots G_t \quad [7]$$

If, in the base year, Y_0 is set equal to 100, then the units of Y_t are that of a "quantity index." If, instead, Y_t is set equal to the value of nominal GDP in the base year, then Y_t is denominated in dollars. This latter approach is taken in the US, where the base year is 1992. Note that chaining can work in both directions in time.

The computation of G can be accomplished using any of several weighting schemes. Consider first a one-period Laspeyres *chain-type* index comparing the value of this period's quantities to the value of last period's quantities valued at last period's prices:

$$G_{L,t} = (P_{C,t-1}Q_{C,t} + P_{I,t-1}Q_{I,t}) / (P_{C,t-1}Q_{C,t-1} + P_{I,t-1}Q_{I,t-1}) \quad [8]$$

or alternatively, in terms of price indices and dollar-denominated real quantities,

$$G_{L,t} = (P_{C,t-1}C_t + P_{I,t-1}I_t) / (P_{C,t-1}C_{t-1} + P_{I,t-1}I_{t-1}) \quad [8']$$

For purposes of making the exposition general, we have dropped the "87" designation from C87 and I87, as can be seen in expression [8']. We prefer expression [8'] because it employs the price indices and dollar-denominated real quantities we are accustomed to working with. Note that fixed-weight GDP is a Laspeyres index, but with price weights from a fixed period.

After a little manipulation, [8'] can be re-written as a weighted average of the growth of C and I:

$$G_{L,t} = \beta_{t-1}(C_t / C_{t-1}) + (1 - \beta_{t-1})(I_t / I_{t-1}), \quad [9]$$

where the weight, β_{t-1} , is the lagged share of nominal consumption in nominal GDP, or $\beta_{t-1} = C_{t-1}/Y_{t-1}$. The key to getting weights which are nominal shares rather than real shares is that the prices in the numerator of expression [8] are from the same period as the prices and quantities in the denominator of [8], i.e. they are "current" prices. If instead, the prices in [8] were from some fixed (base) year, as in [5], the weights would turn out to be real shares.

Next, consider a one-period "Paasche index" comparing the value of this period's quantities to the value of last period's quantities valued at this period's prices:

$$G_{P,t} = (P_{C,t}Q_{C,t} + P_{I,t}Q_{I,t}) / (P_{C,t}Q_{C,t-1} + P_{I,t}Q_{I,t-1}), \quad [10]$$

or alternatively,

$$G_{P,t} = (P_{C,t}C_t + P_{I,t}I_t) / (P_{C,t}C_{t-1} + P_{I,t}I_{t-1}). \quad [10']$$

With a little manipulation, this too can be re-written as a weighted average of the growth in C and I:

$$G_{P,t} = [\beta_t(C_t / C_{t-1})^{-1} + (1 - \beta_t)(I_t / I_{t-1})^{-1}]^{-1} \quad [11]$$

Here, the weight β_t is the *current* share of nominal consumption in nominal GDP. Note again, that since the changes in the real quantities are independent of the base year, and since the nominal component shares are independent of the base year, the growth term itself is independent of the base year.

It is a property of the Laspeyres index that it overstates the value of quantity changes while a Paasche index understates them. A better measure is a Fisher Ideal index which is the geometric average of the Laspeyres and Paasche indices:

$$G_{F,t} = [G_{L,t} G_{P,t}]^{.5} \quad [12]$$

A corresponding measure of the Fisher Ideal Index for the aggregate price level (P_t) can be constructed in an analogous manner by reversing the roles of the prices and quantities in expressions [8'] and [10'] and applying expression [12] to the resulting Laspeyres and Paasche indices. As we have just seen, expressions [8'] and [9] are interchangeable ways to calculate the Laspeyres index and expressions [10'] and [11] are interchangeable ways to calculate the Paasche index, and we can thus describe the weighting process either in terms of price weights or shares of nominal GDP. However, while it is sometimes more intuitive to describe the difference between the Laspeyres and Paasche indices in terms of the nominal-GDP share weights, as we shall see below, it is computationally cleaner and

offers more flexibility to use the price-weight formulations. Once the Fisher Ideal growth term [12] is calculated, expression [7] can be used to calculate the corresponding level of real chain-type GDP or chain-type GDP index.

A nice feature of the Fisher Ideal measures is that if real GDP is denominated in dollars, then nominal GDP is still the product of real output and the price level:

$$Y\$_t = P_t Y_t. \quad [13]$$

Comparing expression [13] to expression [4] clarifies that, at least in concept, the chain-type methodology for aggregation has nothing to do with the measurement of nominal GDP, only its decomposition into real output and price. The formulae also make clear, however, that as a general rule chain-type GDP is not the arithmetic sum of its components which is problem 1.1 listed in section 1.

2.3. Implementation details

What we have just outlined is the *conceptual approach* used by BEA to construct its chain-type measures of real GDP and the aggregate price level, but the actual implementation of this approach is complicated enough to warrant some additional discussion. We begin by considering the annual data which, with the exception discussed in the following paragraph, are constructed as described above with the time period “t” defined as a calendar year; note that, in this process, the underlying data on prices and quantities are averaged to an annual basis before aggregating them.

The aforementioned exception occurs in the ultimate year of the historical time series because BEA has decided not to include a completed calendar year into the annual weighting scheme until after the first annual revision for the year in question. This usually occurs in July of the following year. Therefore, in the latest available full year of data, GDP is not always calculated by growing the previous year’s level with the Fisher index for the latest year; doing so implicitly would utilize the Paasche index based on the latest year’s preliminary estimates of prices, or, alternatively, preliminary estimates of shares in nominal GDP. Instead, it is calculated using only the Laspeyres index for the latest year, which, the reader will recall, is based on the previous year’s prices. Hence, at the end of the otherwise “Fisher Ideal” time-series for GDP is appended a “Laspeyres tail.” With the annual benchmark revisions, usually in July, data for the latest full year will be revised and re-weighted to become a true Fisher Ideal index. This is the annual version of problem 1.3, the use of two different weighting schemes.

Things get even more complicated when we consider the quarterly data because they are not true Fisher Ideal indices, even before the start of the Laspeyres tail. They would be true Fisher Ideal indices if they were constructed as described above with “t” defined as a calendar quarter. But in fact, to reduce volatility in the quarterly series, BEA decided to weight quarterly changes in quantities (and prices) by prices (and quantities) that are *annual averages from two adjoining years, rather than the current and lagged quarterly weights*. BEA felt that the quarter-to-quarter volatility in the quantities and prices that serve as the weights in the aggregation formulae would impart *unacceptable* volatility to the aggregated components. It is not clear what exactly was meant by *unacceptable*, or what, if any, experimentation or other evidence led them to this conclusion.

Figure 1 - Weighting quarterly quantity changes: fisher ideal history

Annual Prices for Laspeyres Index				Annual Prices for Paasche Index			
1994				1995			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Quarterly Fisher Index							

To understand how this weighting scheme works, consider the years 1994 and 1995. The quarterly Laspeyres growth terms used to chain to the level of real GDP from 1994:Q3 through 1995:Q2 all are calculated using price weights computed as annual averages over calendar year 1994, while the corresponding Paasche indices use 1995's prices. This is shown in Figure 1. The use of 1995's prices is then extended into the Laspeyres tail of the quarterly data, which currently begins in the third quarter of 1995. This is the quarterly version of problem 1.3 (See Figure 2). Note that in the Laspeyres tail, recent quarterly data are once again constructed as fixed-weight indices, since the annual prices from the so-called "anchor year" (1995 in this discussion) are used as the weights in the aggregation of the real components.

Figure 2 - Weighting quarterly quantity changes: the Laspeyres tail

Annual Shares for Laspeyres Tail							
1995				1996			
Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Quarterly Laspeyres Index							

One question that immediately springs to mind is this: how could the annual average of quarterly data constructed this way possibly equal the annual data constructed as described above? The answer is, it shouldn't. However, BEA makes an *ex post* adjustment to the quarterly data so that the average of the four quarterly values within a calendar year does in fact just equal the reported annual figure. Yet, more than one year after the conversion to chain-type GDP as the featured measure, BEA has still not formally and publicly documented how this "interpolation," as they call it, is accomplished. This is problem 1.6. There is not much more to say about this problem. Historically, the published quarterly data do average to their annual counterparts. In simulation, we use annual averaging of the quarterly data, component by component. Until we can learn more about BEA's interpolation procedures, this is about all we can do.

As should be clear from the above discussion, the *quarterly* data are not true Fisher Ideal indices. As a result, they do not have the property that the nominal magnitude is the product of

the real magnitude and the chain-type price index. This is problem 1.4, and it applies only to the quarterly aggregates. However, one can still compute an implicit deflator for any aggregate as the nominal magnitude divided by the real. These implicit deflators are identical to the chain-type indices for annual data, and close, but not exactly equal to the chain-type indices for quarterly data. Solutions to this problem will be discussed further in section 5.

Another implementation detail that is a problem from a modeling perspective, is that every published aggregate time series is constructed from the lowest level of detail available. In the US national accounts there are roughly 1300 “elemental” components of GDP. Since even the largest structural macro models of the US attempt to model far far fewer of the elemental real and price components, modelers are stuck with constructing aggregates from aggregates. There is no guarantee the resulting aggregation will exactly match the official published figures. This is problem 1.5. A solution to this problem in simulation is discussed in section 5.

The use of different weighting schemes for deep history and recent history, and especially the use of a fixed-weight Laspeyres index in the Laspeyres tail poses an interesting problem. Since the Laspeyres tail is extended forward in time with the first release of each additional quarter of data, there arises the possibility of substitution bias creeping back into the recent historical and forecasted data. This may not be a serious problem for a forecast that extends just a few quarters into the future, but certainly is for a forecast with a time horizon measured in years. Therefore, we have structured the aggregating identities in the macroeconomic Adviser’s model such that users have the option of generating quarterly forecasts of aggregates that are constructed as a fixed-weight Laspeyres index, or as a true quarterly Fisher index, or as an annual Fisher index, that is, extending the use of annual weights. These details are covered in section 5.

In the following three sections we address the three major tasks that modelers must complete in order to properly convert a fixed-weight model to a chain-type model. This discussion draws on our own experience in re-structuring the Washington University Macro Model (WUMM) to accommodate chain-type GDP. This model was built by the principals of Macroeconomic Advisers, is maintained by MA, and is made available to its clients under license. WUMM is one of the class of “large” quarterly structural macro models that traces its origins back to the MIT-Penn-SSRC modeling effort. We offer below our particular set of solutions to the problems indicated above. There certainly may be other reasonable solutions.

3. Model re-estimation with chain-type GDP

Least squares estimators are both biased and inconsistent if the independent variables in a regression are mismeasured. Hence, if chain-type data are “true”, then estimation with them should produce parameter estimates that more accurately reflect underlying utility and production functions than do the biased and inconsistent estimates derived from fixed-weight constructs. This doesn’t mean, however, that regressions will “fit” better with chain-type data. That depends upon the variances and covariances of both the dependent and independent variables in one’s regressions, old and new. Generalizations are not really possible beforehand. We do suspect, however, that the use of annual weights in the computation of the quarterly data and the *ex post* adjustment made to the quarterly figures in order to achieve consistency with the annual numbers have implications for serial correlation in the residuals of our equations. Since these “interpolations” remain undocumented, we remain unsure of their effects.

When running regressions with chain-type data, one should think carefully about the issue of scaling, although this has turned out to be less of a concern than it was prior to the actual switch to chain-type GDP. In the months leading up to the switchover to chain-type GDP, BEA had considered featuring the pure index form of GDP and its components. That is, both GDP and the price index for GDP, as well as every other component and price index would have a value of 1 (or 100) in the base year. BEA eventually elected to feature the 1992-dollar denominated indices. In the days before chain-type GDP, we were regressing consumption on household income and wealth with everything measured in 1987 prices. The estimated coefficients had familiar and straightforward interpretations as marginal propensities to consume. We take a certain comfort from these numbers because we can place them in the context of a long literature on the consumption function and so have some immediate sense of what constitutes an acceptable empirical and theoretical result.

Table 1 – Estimates of service consumption

Dep Var: Service Consumption (Per Capita)
Sample: 1966:Q4 - 1991:Q4

	(1) 1987 Prices	(2) Chain Index	(3) Chain Dollars
Rebate income	0.08141 (1.43877)	0.00470 (1.33123)	0.07689 (1.33123)
Labor income	0.26483 (7.31120)	0.01603 (7.41878)	0.26240 (7.41878)
Medical transfers	0.99825 (3.84184)	0.05653 (3.65765)	0.92088 (3.65765)
Other transfers	0.46669 (3.58657)	0.02747 (3.69483)	0.45020 (3.69483)
Equity Values	0.02361 (2.72113)	0.00173 (3.51620)	0.02833 (3.51620)
Other Net Worth	0.05001 (7.11594)	0.00311 (7.77426)	0.05089 (7.77426)
Adj R-Squared	0.99944	0.99945	0.99945
Durbin-Watson	1.78539	1.84071	1.84071
Rho	0.84515	0.81802	0.818
% R.M.S.E.	0.68510	0.63960	0.63960

If the series on consumption and income we use in our regressions were measured as index numbers rather than in dollars, the estimated coefficients would not be subject to familiar interpretations until we scaled the results back into an intelligible form. True, not all regressions are subject to this nuisance. For example, many of the relationships in our model are log-linear with coefficients interpreted as elasticities that are independent of how the data are scaled. Still others are in terms of nominal quantities and so are unaffected by the new methodology. But in cases where ease of interpretation is an issue, we decided long ago to use the real quantity indices denominated in 1992 dollars before performing any empirical work. As it turned out, BEA is featuring the 1992-dollar denominated data, and so it is readily available.

Initially we worried that re-estimation of WUMM using the new data could so alter some of its key relationships that the fundamental properties — perhaps even the stability — of the model might be affected. In fact, we've found that most equations look pretty much the same whether estimated on the old or new data. As a case in point, consider one of the largest components of GDP in WUMM: consumer spending on services. It is modeled as a function of tax rebates and surcharges, disposable labor income, medical transfer income (mostly Medicare and Medicaid benefits), other kinds of transfer income (mostly Social Security benefits), the equity portion of household net worth, and other net worth. Our old regression is summarized (we've spared the gory details of the distributed lags) in Column 1 of Table 1.

Everything is measured in 1987 prices and per capita. T-statistics are shown in parentheses, and some regression summary statistics are also shown at the bottom of the table. The coefficients are subject to the usual interpretation; i.e., the MPC out of labor income is 26 cents on the dollar, the MPC out of equity values is 2.4 cents, and so forth.

Next, we re-ran the regression using the quantity index for spending on consumer services (per capita) as the dependent variable and computing real incomes and wealth through division by the alternative price index for total PCE rather than its implicit deflator. These results are shown in column 2 but, in this form, it is hard to make sense of the coefficients — the scaling problem. Finally, in column 3 is shown the same regression as in column 2 but with the quantity index for PCE-services scaled to 1987 dollars. Now columns 1 and 3 are more comparable and, as you can see, the qualitative characteristics of the regression haven't changed too much.

Which equation is "better?" With the chain-type data, t-scores generally are somewhat higher. In the case of the estimated MPCs out of net worth, the differences are large enough to get our attention, especially since the link between equities and the consumption of services — a very important relationship in the model — is about 20% larger when estimated with the chain-type data. The adjusted R^2 seems to give the slightest of edges to the "old" equation, but that comparison is misleading because the regressions have different dependent variables to which are applied slightly different corrections for first-order serial correlation. To circumvent that obstacle, we computed the percent root-mean-square error of each equation without the advantage of the rho-correction, which is how equations are coded in our model. For the old equation, the RMSE is 0.6851%; for the new, 0.6396% — a 6.6% improvement. On balance, we'd say, "usher out the old and bring in the new!"

In other equations where relative prices play a more direct role the parameter estimates can and do change more, but the bottom line is this: you can proceed pretty much as usual with your regression work. We recommend that if given the choice of using pure index numbers or dollar-denominated values, use the dollar-denominated values. It does no harm, and can assist in the interpretation of some parameters. The fit of some familiar specifications will probably get better, some might get worse, and some coefficients will change by amounts that could be worth worrying about; but don't expect to unearth some startling new macro relationship that has gone unnoticed over the years.

4. Simulation should be independent of the aggregation scheme

If it is true that the economy evolves through time independent of our measurement of it, then it should also be true that model simulations should be independent of the aggregation scheme used. This becomes an issue because of the use of multiple aggregation

schemes. If there were only one weighting option, e.g., all historical and forecast data were constructed as true quarterly Fisher Ideal indices, then the aggregating identities in the model would be coded for this method of aggregation and we would be done. However, in the US, the construction of recent historical data as fixed-weight Laspeyres indices, in the Laspeyres tail, requires that we offer an option of simulating the model so that it produces aggregate measures constructed this same way. Users whose forecast horizon is one, two, or three quarters, and there are many, want to forecast the data that BEA will be releasing. On the other hand, long-term projections that produce fixed-weight Laspeyres aggregates are subject to substitution bias creeping back into the simulation. Therefore, a true chain-type aggregation method must also be available for these longer projections.

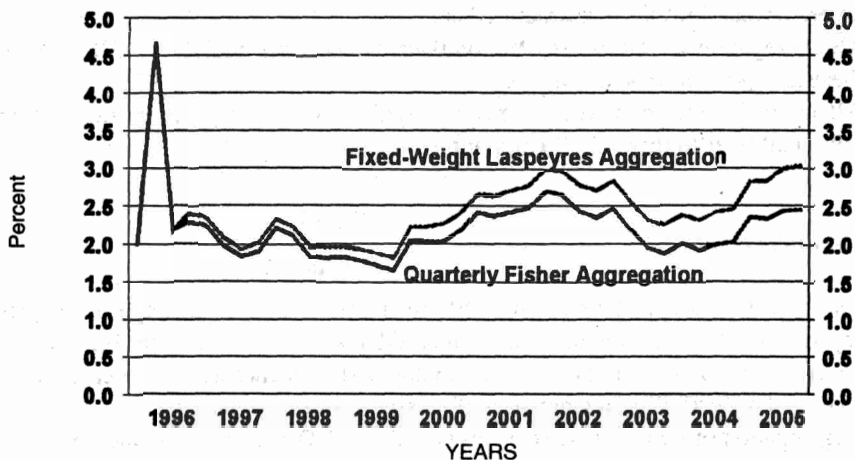
We have chosen to offer three types of aggregation schemes in simulation: annual Fisher, Laspeyres Tail, and quarterly Fisher. Each of these can be selected interactively by a user simulating the model. Table 2 summarizes the alternative weighting schemes used historically and available in simulation. Section V covers the details of how the three alternative aggregation schemes are handled in simulation.

Lest you think we have gone to a lot of trouble for no good reason, we provide in figure 3 a comparison of the simulated growth of GDP constructed as a fixed-weight Laspeyres index that is the extension of the Laspeyres tail, and the growth of GDP constructed as a quarterly Fisher Ideal index. One of the primary reasons for the difference in the two reported growth paths is the assumed 12% annual rate decline in the price of the computer component of Producers' Durable Equipment (PDE) relative to the price of other PDE. Even with this quite modest relative price decline (the actual has been running at 15% - 20%) for computer equipment, the difference in GDP growth quickly reaches a couple of tenths of a percentage point and exceeds 1/2 percentage point by the end of the ten-year simulation horizon.

Table 2 – Aggregation schemes in chain-type GDP

When used:	Type of Aggregation	Weights
Historical Data		
“Quarterly Deep History” Quarterly data where Paasche weights are available	Fisher	Quarters 3,4,1,2 - Annual Weights Laspeyres from year including qtrs 3 & 4 Paasche from year including qtrs 1 & 2
“Quarterly Laspeyres Tail” Quarterly data where Paasche weights are not available	Laspeyres	Laspeyres Annual Weight from the anchor year - currently 1995
“Annual Deep History”	Fisher	Laspeyres Annual Weight from Prior Year Paasche Annual Weight from Current Year
“Annual Recent History”	Laspeyres	Laspeyres Annual Weight from the anchor year - currently 1995
Forecast Data		
Quarterly Short-term Forecast	Laspeyres	Laspeyres Annual Weight from the anchor year - currently 1995
Quarterly Long-term Forecast	Fisher	Laspeyres Quarterly Weight from Prior Qtr Paasche Quarterly Weight from Current Qtr
Quarterly Long-term Forecast	Fisher	Laspeyres Annual Weight from “Prior” Year Paasche Annual Weight from “Current” Year

Figure 3 - Substitution bias in simulation – Real GDP growth



How then can one construct a model with underlying simulation properties independent of the aggregation scheme used in simulation? The answer is to remove from the simultaneous block of the model, all aggregates that depend on the aggregation scheme. This can be accomplished by either replacing these aggregates with their elemental components, or by replacing these aggregates with additional variables constructed such that they are always computed the same way in simulation regardless of the aggregation scheme in use. For example, in WUMM, the real value of nonfarm business output less housing, a variable we call QB, is the key “scale” variable in a number of important behavioral equations. Of course, as BEA reports this series, it is an annual Fisher Ideal index in deep history, and a fixed-weight Laspeyres index in recent history. Will it be yet a third type of index in simulation? To cut through this problem we construct a variable we call QBADJ. In WUMM, QBADJ is always a quarterly Fisher index built from the elemental components in the model. This series is then used in the estimation and simulation of several key equations. There are eight such “auxiliary” aggregates in WUMM.

This approach reduces the problem of multiple aggregation schemes to basically a table-writing problem. All official aggregates in the model are calculated post simultaneously, according to the aggregation scheme selected by the user. Since none of the “official” aggregates appear in the simultaneous block of the model, switching aggregation schemes has no impact on any of the simulation values of any variables, other than the official aggregates themselves, or other series which are recursive.

5. Restructuring the identities in the model

Another major task facing a macro modeler in the switch to chain-type GDP is the restructuring of a large number of identities in the model. This task is necessitated by problem 1.1, the nonadditivity of real GDP, or, more generally, the fact that the method of aggregating real quantities and prices has changed. In addition, the fact that for quarterly data real times price no longer yields exactly the nominal magnitude, problem 1.4, requires restructuring each of these identities. The fact that BEA uses different weighting schemes for deep history and recent history, with newly-released aggregate data constructed as fixed-weight Laspeyres indices, problem 1.3, requires that the new identities be specified to

allow for multiple weighting schemes. Otherwise, users performing long-term projections would again face the problem of substitution bias creeping back into their forecasts.

The two types of identities to be restructured are:

1) All aggregation identities for real components and price indices of GDP. We shall refer to these as *aggregation identities*.

2) All identities that construct nominal magnitudes from real quantities and price indices. We shall refer to these as *PXQ identities*.

5.1. Aggregation identities in model simulation

As we indicated earlier, thinking of GDP growth as a weighted average of the growth of its components, where the weights are nominal GDP shares, is quite intuitive. However, the price-weight formulation turns out to be computationally cleaner and more easily accommodates the kind of flexibility in simulation discussed above.

The aggregation of two or more real components of GDP can be illustrated by continuing with our two good economy. Below we repeat expressions [8'], [10'], and [12]. These constitute the basic set of aggregation calculations that must be performed for each real aggregate and each price aggregate.

$$G_{L,t} = (P_{C,t-1}C_t + P_{I,t-1}I_t) / (P_{C,t-1}C_{t-1} + P_{I,t-1}I_{t-1}) \quad [8']$$

$$G_{P,t} = (P_{C,t}C_t + P_{I,t}I_t) / (P_{C,t}C_{t-1} + P_{I,t}I_{t-1}) \quad [10']$$

$$G_{F,t} = [G_{L,t} G_{P,t}]^5 \quad [12]$$

To illustrate how these expressions are coded to allow for the possibility of multiple weighting schemes, we write [8'] and [9'] below, substituting price weight variables for the price indices.

$$G_{L,t} = (W_{C,L}C_t + W_{I,L}I_t) / (W_{C,L}C_{t-1} + W_{I,L}I_{t-1}) \quad [8'']$$

$$G_{P,t} = (W_{C,P}C_t + W_{I,P}I_t) / (W_{C,P}C_{t-1} + W_{I,P}I_{t-1}) \quad [10'']$$

There are two price weight variables for each real GDP component, a Laspeyres price weight and a Paasch price weight. Once the aggregating identities are coded this way, switching aggregation schemes is accomplished by re-computing the price weight terms in accordance with the aggregation scheme.

For the Laspeyres tail, the values of $W_{C,L}$ and $W_{C,P}$ are the annual average price of C in the anchor year, currently 1995. The values of $W_{I,L}$ and $W_{I,P}$ are the annual average price of I in the anchor year. You will notice that this generates values of $G_{L,t}$ and $G_{P,t}$ that are identical. And indeed, this is the desired result, since we want expression [12] to yield the fixed-weight Laspeyres index. Because the aggregates are fixed-weight indices in the Laspeyres tail, the price weights need only be computed once, and then applied iteratively over the forecast horizon.

For the quarterly Fisher index, $W_{C,L}$ and $W_{C,P}$ are the lagged and current price index for C, respectively, and $W_{I,L}$ and $W_{I,P}$ are the lagged and current price index for I, respectively. Feeding the resulting values of $G_{L,t}$ and $G_{P,t}$ into expression [12] yields a true quarterly Fisher index. In this case, the price weights must be re-computed each quarter in the simulation horizon.

For the annual Fisher index, $W_{C,L}$ and $W_{I,L}$ are the annual average price indices for C and I, respectively, in the year containing quarters 3 and 4, and $W_{C,P}$ and $W_{I,P}$ are the annual average price indices for C and I, respectively, in the year containing

quarters 1 and 2. Thus, as the Fisher index is computed forward through time, the price weights are re-computed every four quarters.

Two points are in order about the discussion above. First, we are treating C and I as if they are elemental components of GDP, when in fact they are themselves aggregations of elemental components. We sometimes refer to this process as calculating "a Fisher of Fishers." In WUMM, there are roughly 40 components of GDP at the lowest level of aggregation, none of which are truly elemental in that BEA produces more disaggregated data. Second, we have carried out the entire exposition in terms of aggregating real quantities. There is a parallel and entirely analogous set of expressions for aggregating prices, where the roles of prices and quantities are reversed.

Once the growth terms are calculated, the level of the index must be computed. A condensed version of expression [7] is shown below.

$$Y_t = Y_{t-1}G_t \quad [7']$$

This expression shows how the current level of GDP is constructed by chaining to the previous level with the growth index. However, as noted in section 1, the construction of the aggregate data from the lowest level of elemental components available will mean that construction of the aggregate with anything but the lowest level of detail will not exactly replicate the official figures. Therefore, we calculate historically a residual term R_t :

$$R_t = Y_t / Y_{t-1} G_t \quad [14]$$

These residuals are calculated historically for every quantity and price aggregate in the model, maintained in the model database, and can be displayed with the model simulation software. Because the term G depends on the aggregation method employed, so too do the residual terms. Switching aggregation methods requires recalculating historically the residual terms. For simulation purposes we need to incorporate the residual, and [7'] becomes:

$$Y_t = Y_{t-1}G_tR_t \quad [15]$$

In simulation these residuals can be altered by the user, but our experience suggests that they are best set equal to one in the forecast horizon. Values other than one imply a boost or a drag to the growth of Y relative to the underlying components. Figure 4 shows the GDP aggregation residual.

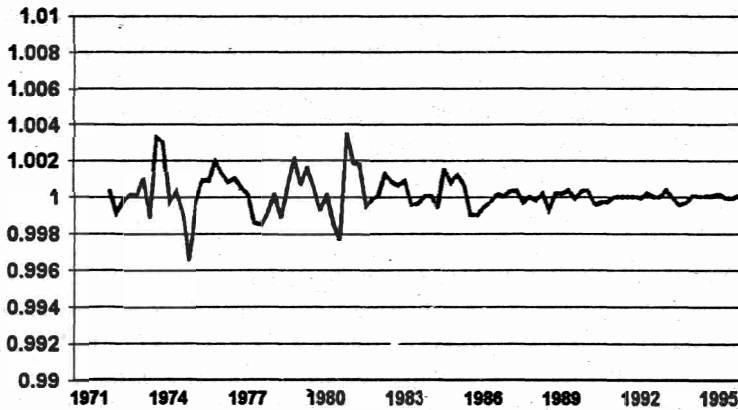
It should be made clear that this residual does not make the components of GDP additive. They will still fail to add-up to GDP. Presenting the components in a table with GDP, usually involves calculating the additive residual and including it as a separate line in the table.

There has been some discussion in modeling circles in the US about the properties of these aggregation residuals, and whether anything useful could be found from an analysis of them historically. Our own view is that such an analysis would yield little of value. Because of the existence of the Laspeyres tail, any statistical analysis of the properties of these residuals in deep history, does not tell us anything about the properties of the residuals in the Laspeyres tail, or the behavior of the residuals in the forecast period. And, the five or six quarters in the Laspeyres tail are an insufficient number of observations to perform any formal tests.

If you are beginning to suspect that several of the modeling difficulties discussed here can be traced back to the decision in the US to employ two different weighting schemes over history, you are correct. A single aggregation method, such as quarterly Fisher, would

have obviated the need for multiple aggregation methods, would have allowed us to learn something useful from the aggregation residuals, and as we will see below, would have allowed us to drop one whole set of residuals altogether.

Figure 4 - The GDP aggregation residual



Note: Calculated using Quarterly Fisher Index

5.2. PxQ identities in simulation

Conceptually, for true Fisher Ideal indices of real output and prices, it is the case that expression [13], repeated here, holds identically.

$$Y\$_t = P_t Y_t . \quad [13]$$

In the US implementation of chain-type GDP [13] does hold for annual data, but does not hold for quarterly data. Therefore, both historically and in the forecast period, we calculate a "PxQ residual" of the form:

$$R_t = Y\$_t / P_t Y_t \quad [16]$$

A residual of this type is computed for every time period and for every PxQ identity. These residuals are stored in the model database, can be viewed with the model simulation software, and can be manipulated by the user in the forecast period. These residuals are independent of the aggregation scheme. Incorporating the residual in the PxQ identity [13] yields:

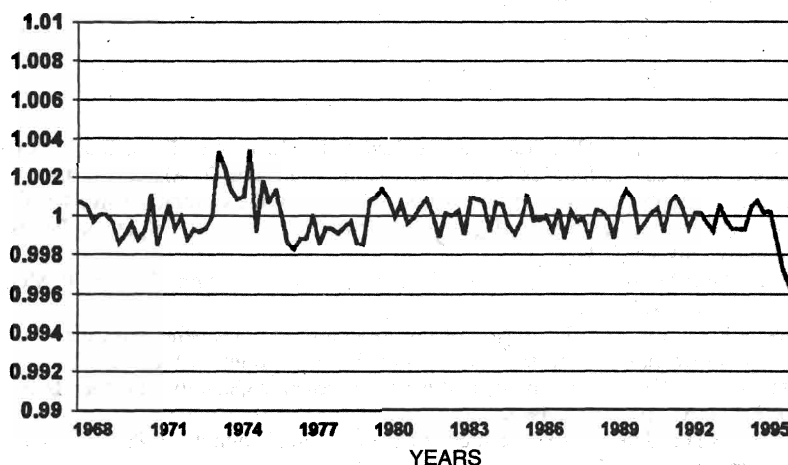
$$Y\$_t = P_t Y_t R_t . \quad [17]$$

There may be some exploitable information in these residuals, especially if one is operating in the Laspeyres tail. As Figure 5 below indicates, there is sometimes a clearly discernible trend in the residual within the range of the Laspeyres tail.

This residual appears to be equal to 1 with approximately random errors, up until the start of the Laspeyres tail. From that point the errors are decidedly one-sided. We are at a loss to explain why this happens, and have not received an answer from BEA as to why this is occurring.

The large deviation from 1 at the end of the historical data makes it imperative that users have a way of extending this residual into the forecast period. Otherwise they would impart a large jump in the first forecast quarter to any nominal magnitude so constructed.

Figure 5 - The PxQ residual for nonpetro imports



6. Summary and concluding remarks

The switch to chain-type GDP eliminates the substitution bias problem that plagues the fixed-weight data we have been accustomed to working with. This alone makes it worth the facing the difficulties occasioned by the switch. Nevertheless, modelers must be prepared to take the necessary steps to re-structure their models to accommodate the new regime. The major tasks are:

1) Re-estimate the behavioral relationships. Generally, where aggregates are employed on the left or right hand side of equations, eliminating the bias inherent in the fixed-weight data will improve the regressions. However, one must pay special attention to issues of scaling, especially where coefficients have familiar interpretations. For this reason, it makes sense to use dollar-denominated quantity indices where possible.

2) If more than one aggregation scheme is to be employed, and it may well be a good idea, modelers must be prepared to re-structure their models to remove any aggregate series from the simultaneous block of the model and reduce aggregation to a table-writing problem by making it totally post simultaneous.

3) Be prepared to add possibly thousands of lines of additional code to accommodate the new aggregation method. Even then, it will probably be necessary to incorporate aggregation and PxQ residuals to insure that the model can replicate the official figures.

The year and one half we spent re-structuring, and re-estimating the model was difficult to say the least. It was made more difficult by the lack of clear documentation of the important implementation details discussed in this paper, and others that we have not included. The implementation details are important, and should be subject to the review of the user community, well in advance of the actual switch to chain-type GDP. This will give users adequate time to make the needed changes.

Finally, even today, more than one year beyond the switch to chain-type GDP in the US, there is a gross lack of appreciation for the problems related to working with chain-type GDP. To the extent that groups such as ours have handled the ugly details and incorpo-

rated into our models the needed modifications, perhaps we have made it a kind of black box. Indeed, this is the case in our system. Users can be totally ignorant of the details of chain-type GDP and still generate results that observe the rules of its construction.

It behooves the statistical agencies anticipating a switch to chain-type GDP to start well in advance educating the user community about why chain-type GDP is a superior product, and how to use it intelligently.

References

BRAITHWAIT, STEVEN D., "The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost of Living Indexes", *American Economic Review* 70, March 1980, 64-77.

LANDEFELD, J. STEVEN, AND ROBERT P. PARKER, "Preview of the Comprehensive Revision of the National Income and Product Accounts: BEA's New Featured Measures of Output and Prices", *Survey of Current Business*, July 1995, 31-37.

MACROECONOMIC ADVISERS, LLC, "Chain-Type Aggregation Weighting Options In Model Simulation", Technical Note, March 1996.

PRAKKEN, JOEL L. AND LISA T. GUIRL, "Macro Modeling & Forecasting with Chain-Type Measures of GDP", *Macroeconomic Advisers, LLC, Special Analysis*, August 29, 1995.

TRIPLETT, JACK E., "Economic Theory and BEA's Alternative Quantity and Price Indexes", *Survey of Current Business*, April 1992, 49-52.

YOUNG, ALLAN H., "Alternative Measures of Change in Real Output and Prices", *Survey of Current Business*, April 1992, 32-43.

I CONTI NAZIONALI CALCOLATI MEDIANTE INDICI A CATENA: ALCUNI PRIMI RISULTATI PER IL CASO ITALIANO

Sandra Maresca

Istat, Dipartimento di Contabilità Nazionale e Analisi Economica

1. Introduzione

Il nuovo sistema europeo dei conti economici nazionali (d'ora in poi SEC95), in linea con quanto affermato nell'ultima versione del "System of National Accounts" (SNA93), raccomanda l'impiego di indici a catena per misurare le variazioni di volume e di prezzo. In realtà nonostante la decisa presa di posizione del nuovo SNA a favore degli indici a catena, tra gli esperti della materia è ancora aperto il dibattito sulla superiorità del nuovo approccio, ma soprattutto sulle conseguenze che questo comporta nell'ambito di uno schema contabile come quello dei conti economici nazionali.¹ La raccomandazione di applicare il metodo del concatenamento, d'altra parte, è in linea con quanto già da tempo evidenziato dalla teoria riguardante i numeri indice.² In sede internazionale la formulazione consigliata è l'indice di Fisher, ma vengono accettati quali sostitutivi l'indice di Laspeyres per le variazioni di volume e di Paasche per quelle di prezzo.

Questo lavoro si propone di individuare, nell'ambito della contabilità nazionale, le possibilità che si aprono con l'adozione di un nuovo approccio alle misure in volume, diverso dall'attuale fondato sulle tradizionali valutazioni a base fissa. Nel paragrafo che segue sono passati rapidamente in rassegna gli aspetti metodologici del metodo a ponderazione variabile, avendo un'attenzione particolare agli elementi che qualificano tale approccio e che lo contraddistinguono dal suo diretto sostituto. Il terzo paragrafo è dedicato all'esame delle attitudini della metodologia del concatenamento nell'ambito degli schemi della contabilità nazionale con riferimento alle nuove indicazioni stabilite in sede internazionale sulle valutazioni in termini di volumi. Nel quarto paragrafo vengono riportati i risultati di alcune prime sperimentazioni riguardanti il calcolo degli indici a ponderazione variabile applicati agli aggregati dei conti economici italiani. Nella parte finale vengono tratte alcune prime conclusioni.

¹ A tale proposito, in sede comunitaria è stata istituita una Task Force che si occupa delle problematiche relative alle valutazioni a prezzi costanti nella quale la scrivente è stata chiamata a rappresentare l'Istituto Nazionale di Statistica.

² La bibliografia in merito è ampia; in particolare si rinvia al lavoro di Allen (1975) e alla rassegna curata da Biggeri (1986).

L'autore desidera ringraziare M. Grassini e I. Visco per gli utili suggerimenti ed E. Giovannini, L. Picozzi, S. Pisani, e A. Viviani per le osservazioni alle versioni preliminari di questo lavoro. Le considerazioni espresse in questo lavoro non impegnano l'Istituto Nazionale di Statistica, ma scaturiscono comunque dal lavoro svolto nell'ambito del progetto di ricerca sui temi della deflazione avviato dal Dipartimento di Contabilità Nazionale ed Analisi Economica

2. L'indice a catena: aspetti metodologici

2.1. Il concatenamento degli indici a base mobile

L'indice a catena nasce con l'esigenza di introdurre un indicatore delle variazioni di prezzo o di volume che non tenga solo conto dei valori assunti dalle variabili considerate in due tempi precisi (l'istante corrente e l'istante base), ma che incorpori l'andamento complessivo presentato dal fenomeno nell'intervallo temporale esaminato. In altri termini, l'indice dovrebbe dipendere sia dalle intensità manifestate nei punti estremi, sia da quelle assunte durante l'intervallo stesso.³

In forma generale, l'indice a catena $I_{t/0}^c$ si ottiene dal prodotto dei successivi indici $I_{1/0}$, $I_{2/1}$, ..., $I_{t/t-1}$ riferiti ai sub-intervalli $[0,1]$, $[1,2]$, ..., $[t-1,t]$:

$$I_{t/0}^c = \prod_{j=1}^t I_{j/j-1} \quad [1]$$

qualunque sia la formula adottata per il calcolo degli indici $I_{j/j-1}$ e la variabile in osservazione.

Gli indici a base mobile della [1] possono essere calcolati utilizzando le formulazioni comunemente riportate dalla teoria dei numeri indice, vale a dire le formule di Laspeyres, di Paasche, di Fisher, di Tornqvist, di Sato-Vartia, ecc. ciascuna rispondente ad una precisa ipotesi economica sulla relazione esistente tra prezzi e quantità.⁴

Gli indici a base fissa (o binari), che come è noto vengono costruiti ipotizzando che la specificazione dei beni rimanga invariata nell'intervallo di tempo considerato, diventano non significativi e addirittura irrealistici quando i due estremi del confronto si allontanano. Alla perdita di rappresentatività del paniere, in genere, si ovvia attraverso il cambiamento di base dell'indice e attraverso il rinnovamento del paniere dei prodotti di riferimento. L'indice a catena non solo risolve automaticamente il problema del "disfacimento" del paniere e, conseguentemente, del cambiamento di base mediante il cambiamento di pesi, ma presenta il vantaggio di incorporare tutte le variazioni economiche verificatesi nel periodo in esame.⁵

Se durante l'intervallo considerato si sono prodotte variazioni strutturali, sia sul piano delle quantità, sia sul piano dei prezzi, questo significherà generalmente che:

$$\frac{I(0,t+1)}{I(0,t)} \neq I(t,t+1)$$

Nel caso dell'applicazione diretta della formula dell'indice, la variazione per il periodo $(t,t+1)$ viene calcolata con l'ausilio dei due indici $I(0,t)$ e $I(0,t+1)$. In presenza di variazioni strutturali, la configurazione del paniere all'epoca 0 condiziona il calcolo, il che suggerirebbe di calcolare la variazione in questione utilizzando l'indice $I(t,t+1)$, in cui la composizione dell'insieme si riferisce all'intervallo $(t,t+1)$. Procedendo in questo modo di anno in anno, si ottengono serie di variazioni annuali trasformabili in indici a catena attraverso moltiplicazione. Affinché tali modificazioni rientrino esattamente nella costruzione dell'indice,

³ Nel 1887 fu A. Marshall che per primo suggerì l'introduzione nell'indice dei prezzi di pesi che si rinnovano ogni anno. Per Marshall il problema era limitato all'inserimento di nuovi beni in un indice dei prezzi e tale inserimento poteva venire effettuato cambiando i pesi ogni anno e concatenando i successivi indici annuali attraverso una semplice moltiplicazione.

⁴ In Office Statistique des Communautés Européennes (1987) sono presentate le formule utili agli schemi dei conti economici.

⁵ Le modificazioni a cui ci si riferisce riguardano le preferenze dei consumatori, i cambiamenti di qualità dei beni prodotti, le variazioni nella struttura dei consumi risultanti dall'introduzione di nuovi beni e dalla scomparsa di altri.

sono però necessarie informazioni complete su tutti i prezzi e le quantità che compongono la spesa degli operatori, per tutti i periodi del concatenamento.

L'elemento più qualificante degli indici a catena è che, a differenza che in quelli binari, i cambiamenti di prezzo o di volume di ciascun intervallo temporale sono messi in relazione con le spese che gli operatori effettuano nello stesso periodo. Le continue modificazioni dei modelli di spesa impoveriscono, infatti, l'attendibilità degli indici diretti poiché l'ipotesi sulla costanza dei consumi diventa inaccettabile su un lungo periodo di tempo; viceversa, come afferma Forsyth (1978), l'indice in grado di soddisfare tale ipotesi è quello a catena, i cui concatenamenti sono tali da rendere plausibile l'invarianza delle quantità acquistate.⁶

2.2. Vantaggi di una rappresentazione con indici a catena

Come già accennato in precedenza, uno dei maggiori problemi della costruzione dei numeri indice consiste nel fatto che i prodotti sono spesso soggetti a sparizione dal mercato per essere sostituiti da altri nuovi.⁷ Infatti, quando il set di beni esistenti sul mercato subisce variazioni continue o quando gli estremi del confronto sono piuttosto lontani, un indice diretto può fornire solo una comparazione tra una piccola quota del valore complessivo in entrambi i periodi. In queste circostanze la copertura degli indici è così povera da rendere insignificante qualsiasi confronto e la differenza tra l'indice di Laspeyres e di Paasche tende ad amplificarsi.

L'approccio dell'indice a catena fornisce una valida soluzione a questo problema: in ciascun anello della catena le informazioni che vengono sfruttate sono notevoli, dal momento che la sovrapposizione del set di beni nei due estremi del sub-intervallo è quasi totale. Concatenando i successivi anelli si ottiene un indice a catena che incorpora in ogni anello le modificazioni del set dei prodotti e che annovera una quantità di dati maggiore rispetto a ciò che si verificherebbe in un confronto diretto.

Sempre con riferimento alle modalità di costruzione dei numeri indice, si osserva che l'indice binario di Laspeyres risente, in modo consistente, dell'influenza dei pesi del tempo base; viceversa, l'indice di Paasche di quelli correnti. Con tali presupposti, divenendo distanti i termini del confronto e realizzandosi i normali effetti di sostituzione tra beni, negli indici sopra menzionati l'influenza delle ponderazioni di riferimento tenderà ad accentuarsi. Conseguentemente, l'indice di Laspeyres mostrerà una tendenza a muoversi verso l'alto (*upward-biased*) e l'indice di Paasche a muoversi verso il basso (*downward-biased*).

Tale tendenziosità dell'uno rispetto all'altro, che, come si è detto, limita l'applicabilità degli indici binari, viene ad essere ridotta nel procedimento di concatenamento degli indici a catena, dove gli indici a base fissa vengono calcolati per brevi intervalli di tempo e dove, quindi, ciascuna ponderazione ha un'influenza moderata. A tale proposito occorre sottolineare che l'impiego del principio del concatenamento è più importante dei risvolti dovuti alla formula da usare: infatti, purché i concatenamenti vengano fatti ad intervalli ragionevolmente brevi, la scelta dell'ipotesi sulla relazione tra prezzi e quantità sottostante la formula di collegamento ha scarsi effetti pratici.

D'altra parte, questo elemento (che favorisce l'uso degli indici a catena) diventa, in alcune situazioni, un ostacolo teorico per la sua costruzione. A tale proposito può essere utile ricordare i due casi teorici più rilevanti studiati da Szulc (1983). Nel primo si ipotizza che sia il modello dei prezzi relativi, che quello delle relative quantità tendano a muoversi gradualmente mantenendo la stessa direzione nell'intervallo in esame. In questo caso, venendo l'in-

⁶ Alla stessa conclusione si arriva in Triplett (1992) a proposito del "cost-of-living index".

⁷ D'altra parte, come illustra Hill (1988), confrontando in modo diretto due situazioni in cui i beni prodotti differiscono solo per la comparsa di alcuni beni addizionali non vengono incorporati tutti i benefici derivanti dal progresso tecnologico.

dice concatenato di Laspeyres a porsi più in basso del suo diretto sostituto e l'indice concatenato di Paasche più in alto del suo sostituto, la divergenza tra i due tipi di indici presenta una diminuzione. Al limite, quando tutti i prezzi e le quantità variano continuamente ed esponenzialmente ed il concatenamento viene fatto su intervalli infinitamente piccoli, gli indici a catena di Laspeyres e di Paasche coincidono (Hill 1988).

La discrepanza tenderà ad aumentare nel caso contrario, cioè quando prezzi e quantità relative mostrano un andamento oscillatorio legato all'esistenza di una correlazione negativa tra variazioni di prezzo e quantità. Se l'alto prezzo e la quantità ridotta in un concatenamento fossero seguiti, nel concatenamento successivo, da un ritorno a livelli normali di quantità e prezzi, la formula di Laspeyres sovrastimerebbe il contributo della crescita dei prezzi nel primo concatenamento, ma sottostimerebbe la caduta dei prezzi nel secondo. Persistendo tale comportamento di prezzi e quantità, risulta che l'indice a catena di Laspeyres diverge verso l'alto dal corrispondente binario; al contrario, per l'indice a catena di Paasche si avrebbe, invece, una divergenza verso il basso.

In circostanze di prezzi e quantità oscillanti (il cosiddetto fenomeno *bouncing*), il concatenamento produce amplificazioni delle variazioni intervenute nel periodo,⁸ così che gli effetti degli slittamenti degli indici, nonostante la scarsa influenza delle ponderazioni in ciascun anello, possono diventare così pronunciati da compromettere l'affidabilità del procedimento. Questo vincolo teorico viene superato facilmente nella pratica in un ottica di concatenamento annuale, visto che raramente ci si trova nella condizione di dover costruire un indice annuale di volume o di prezzo per categorie di beni le cui quantità e prezzi mostrano variazioni cicliche simultanee. Inoltre, la teoria dei numeri indice ci fornisce la soluzione a questo inconveniente sostituendo semplicemente, come viene riportato in Moulton (1993), l'indice dei prezzi al consumo di Laspeyres con l'indice ottenuto attraverso la formula della media geometrica in cui la struttura dei pesi (che viene ipotizzata costante) corrisponde al valore della spesa invece che alla quantità.

3. L'indice a catena per le misurazioni delle variazioni di volume e di prezzo nella contabilità nazionale

3.1. Le valutazioni a prezzi costanti: le indicazioni dei nuovi schemi di contabilità nazionale

Secondo gli schemi internazionali dei conti nazionali (SNA93 e SEC95) lo scopo delle valutazioni a prezzi costanti è quello di effettuare una scomposizione delle variazioni in valore in variazioni di prezzo e di volume nell'ambito di uno schema che evidenzia le relazioni contabili tra gli aggregati. A tale proposito, l'indicazione che emerge dai nuovi manuali è che la migliore rappresentazione della realtà si potrà ottenere solamente calcolando le variazioni in volume e le variazioni dei prezzi realizzatesi tra due periodi consecutivi. Inoltre, riunendo le variazioni attraverso un indice a catena, si otterrebbe, per approssimazione, l'informazione necessaria per comparare due osservazioni distanti nel tempo.

Come è noto, il maggiore limite degli indici a base fissa consiste nell'invarianza della struttura dei pesi durante tutto il periodo in esame. Quando, come spesso avviene con il trascorrere del tempo, si realizzano modificazioni sostanziali che investono i sistemi di ponderazione, si incorre in una probabile sovrastima degli aggregati espressi a prezzi costanti di un anno base piuttosto "remoto". Questa tendenza riflette il fatto che i prodotti per cui si registra una crescita consistente, generalmente, sono quelli per i quali i prezzi relativi mostrano una dinamica più

⁸ In molti hanno analizzato il fenomeno; tra gli altri si ricordano Forsyth (1978), Szulc (1983), Forsyth e Fowler (1986) e Hill (1988).

contenuta di quella media. Di conseguenza, quando si utilizzano prezzi più aggiornati, i prodotti con forte crescita sono associati a pesi ridotti e la variazione che ne deriva risulta più bassa (in caso di correlazione negativa tra prezzi e quantità relativi). Tale fenomeno (che in letteratura⁹ viene definito come *substitution bias*) produce, quando si attua un cambio di base, una dinamica più lenta dell'aggregato a prezzi costanti rispetto a quella evidenziata nella vecchia base.

A tale proposito è utile sottolineare che l'indice a catena ha il vantaggio considerevole di non essere soggetto a ribasamenti quando si procede ad un cambio di base poiché la metodologia del concatenamento provvede ad un aggiornamento annuale della base e, quindi, come si afferma nell'SNA93 "...chaining avoids introducing apparent changes in growth or inflation as a result of changing the base year".

Relativamente alle formulazioni da utilizzare, entrambi i manuali concordano nel suggerire l'indice di Fisher sia per la misura delle variazioni di volume, sia per quelle dei deflatori; una valida alternativa è comunque la combinazione dell'indice di Laspeyres per i volumi e di Paasche per i prezzi.¹⁰

Dal momento che i conti nazionali costituiscono degli schemi additivi a due dimensioni, nei quali l'addizione degli elementi conduce a sotto-totali e a un totale generale comune, l'indice da utilizzare dovrebbe garantire la coerenza tra i modi di scomposizione dell'evoluzione del valore degli elementi che sfociano in un certo totale e il trattamento di questo totale. In pratica, tale condizione richiede che le componenti dei totali possano essere addizionate quando queste vengono rivalutate ai prezzi di un anno base, così come accade con valutazioni a prezzi correnti.¹¹

In un contesto contabile quale quello dei conti economici nazionali le possibilità pratiche della metodologia del concatenamento vengono fortemente ridotte dal fatto che l'indice a catena non gode della proprietà dell'additività.¹² Infatti, con la metodologia degli indici a catena non è possibile redigere conti in valore a prezzi di un anno base in cui i totali si equilibrino poiché i conti che risultano dalla moltiplicazione dell'insieme dell'epoca di base per un indice di volume a catena presentano divergenze a livello di aggregazione.

L'SNA93 e il SEC95, per ovviare a questo inconveniente, propongono di elaborare indici a catena per seguire l'evoluzione del volume e del prezzo di ciascun macroaggregato e, parallelamente, di continuare a calcolare dati disaggregati ai prezzi di un anno base.¹³ Peraltro, qualora venissero presentate tavole disaggregate di indici a catena, l'utilizzatore, estrapolando le serie elementari riferite ad un qualunque anno base, troverebbe che la loro somma non coincide con il totale estrapolato, a sua volta, con l'indice a catena generale.

3.2. Possibilità pratiche di applicazione degli indici a catena

Nonostante si sia visto che l'approccio degli indici a catena non consenta la compilazione di conti a prezzi di un anno base, l'informazione ottenibile risponde a quella desiderata dalla maggioranza degli utilizzatori dei conti nazionali. Va osservato, infatti, che con una rappresentazione a catena si possono presentare per ogni anno conti economici a prezzi del-

⁹ La situazione che corrisponde ad un cambio di base è stata evidenziata in Young (1992) e Landefeld e Parker (1995).

¹⁰ Il ricorso alle formulazioni menzionate è giustificato dalla necessità di soddisfare la condizione della reversibilità dei fattori. Tale proprietà garantisce la scomposizione totale dell'evoluzione del valore in una componente di volume ed in una di prezzo.

¹¹ Le problematiche inerenti la deflazione degli aggregati della contabilità nazionale sono state affrontate in Siesto (1977), Quirino (1978), ONU (1979) EUROSTAT (1981).

¹² In Martini (1992) la scelta tra le funzioni possibili per la costruzione dei numeri indice viene fatta tenendo conto, oltre che delle proprietà assiomatiche (omogeneità, commensurabilità, monotonicità e identità), anche di altre proprietà quali l'associatività, la reversibilità della base e dei fattori le quali risultano importanti nei confronti che si propongono gli indici nella contabilità nazionale.

¹³ Nell'SNA93 è comunque esplicito che bisogna rassegnarsi alla perdita dell'additività (par. 16.77).

l'anno precedente in cui vengono rispettati i vincoli degli schemi contabili; inoltre, moltiplicando gli indici a base mobile si ottiene un indice a catena per periodi consistentemente lunghi. Con questa metodologia, quindi, possono essere soddisfatte le esigenze sia di coloro che sono interessati agli effetti delle variazioni che si producono da un anno all'altro, sia di chi ha bisogno di serie storiche sufficientemente lunghe. Se, dunque, si applica la formula di Laspeyres per le fluttuazioni in volume e quella di Paasche per quelle dei prezzi, si ottengono variazioni annuali e conti a prezzi dell'anno precedente (in cui viene mantenuta la proprietà dell'additività); inoltre, concatenando gli indici a base mobile, è possibile fissare l'anno base dell'indice in funzione degli usi che se ne vuole fare.

Si è affermato in precedenza che la costruzione degli indici concatenati in sostituzione degli indici ordinari supera simultaneamente i problemi di cambiamento di base e di raccordo delle serie. Tuttavia, contro l'utilizzazione di questa metodologia, vengono spesso invocati inconvenienti di ordine pratico; per questo motivo in questa sede si cercherà di verificare se tali problemi pratici costituiscono effettivamente un ostacolo alla loro applicazione. In particolare, dal momento che gli inconvenienti generalmente menzionati si riferiscono alla quantità di informazioni indispensabili, ci si propone di effettuare una comparazione tra la metodologia del concatenamento e quella classica.

Dal punto di vista teorico, il periodo scelto per formare le maglie della catena dovrà essere più corto possibile, in modo da minimizzare gli effetti delle variazioni che si producono nella struttura dei prezzi e delle quantità; nella pratica, i periodi coincideranno il più delle volte con l'anno solare, visto che i conti nazionali coprono generalmente un anno. Ai fini della deflazione degli aggregati di contabilità, il grado di dettaglio dei calcoli gioca un ruolo importante ed il livello di disaggregazione ottimale al quale operare può essere definito come il livello più fine per cui sono disponibili sia valori a prezzi correnti,¹⁴ sia indicatori di prezzo.

Partiamo dall'ipotesi che i conti nazionali siano elaborati annualmente, cosicché per ogni anno studiato si dispone di serie complete di dati a prezzi correnti, e che sia possibile modificare i coefficienti di ponderazione a livello di base. Ipotizziamo anche di poter disporre di indicatori di prezzo allo stesso livello di disaggregazione. In pratica, questi indicatori possono essere di diversa natura: possono essere calcolati attraverso valori medi unitari o con l'osservazione diretta di prezzi. Sulla base delle ipotesi fatte in OSCE (1987), si dimostra agevolmente che, per ottenere la serie dell'indice a catena, non si ha bisogno di più informazioni che per la serie del suo diretto sostituto.

Esaminando i costi e gli sforzi supplementari che richiede l'uso degli indici a catena in rapporto agli altri, si constata che una parte considerevole del lavoro di uno statistico è costituito dalla raccolta, il trattamento, l'analisi delle informazioni che si usano, ma soprattutto dalla verifica della coerenza degli indicatori di prezzo. Ora, dal momento che attraverso le serie ordinarie si compilano per ogni anno studiato conti a prezzi dell'anno base e con gli indici a catena conti a prezzi dell'anno precedente, emerge che nell'uno come nell'altro caso vengono redatte e, soprattutto, verificate lo stesso numero di tavole in prezzi di un altro anno.

Si è parlato in precedenza della disponibilità di indicatori di prezzo e, congiuntamente, di serie complete di dati a prezzi correnti con lo stesso livello di disaggregazione. Questa necessità comporta sicuramente un aggravio in termini di costruzione di indici di prezzo ufficiali e, quindi, di dettaglio per le stime correnti dei conti economici nazionali da utilizzare nelle operazioni di ponderazione per gli indici ufficiali.

L'utilizzazione della metodologia del concatenamento comporta inevitabili costi nella gestione delle procedure informatiche. Sostituendo al sistema con ponderazione fissa quello

¹⁴ Da paese a paese e a seconda che si tratti di stime provvisorie o definitive, può corrispondere un grado di disaggregazione differente.

a ponderazione variabile, sono necessari controlli continui e pesanti di coerenza e completezza di informazioni (si pensi, a tale proposito, alla maggiore possibilità di trovare valori nulli in corrispondenza di beni scomparsi dal mercato o risultanti da cambiamenti nelle nomenclature).

4. L'applicazione della metodologia del concatenamento agli aggregati della contabilità nazionale per l'Italia

Al fine di valutare le implicazioni derivanti dalle raccomandazioni sulle valutazioni a prezzi costanti stabilite in sede internazionale sono stati rielaborati gli aggregati macroeconomici che compongono il conto delle Risorse e degli Impieghi alla luce della metodologia del concatenamento.

Si fa presente che l'elaborazione svolta ha una valenza limitata dal fatto che il livello di disaggregazione adottato per ciascun aggregato è quello per cui si realizza la coerenza del sistema. Relativamente all'offerta si è lavorato sugli aggregati del valore aggiunto a 44 branche. Dal lato della domanda si è usata l'usuale disaggregazione per 40 funzioni dei consumi finali delle famiglie, quella a due voci dei consumi collettivi e quella a 21 voci per branca produttrice degli investimenti. La disaggregazione delle importazioni e delle esportazioni si è avvalsa della distinzione tra beni e servizi.

Il vincolo costituito dall'aver scelto di partire da dati per i quali è verificata la coerenza tra risorse e impieghi comporta che come serie elementari sono stati considerati in realtà degli aggregati espressi ai prezzi costanti di un anno base (il 1990); ciò riduce, nell'esercizio, gli effetti del concatenamento. Infatti, quando si calcola l'indice a catena partendo da dati in volume ricavati da valori a prezzi costanti di un anno base, le variazioni in volume al di sotto del livello di aggregazione scelto dipenderanno dai prezzi relativi dell'anno base (Dalgaard, 1994).

Per ciascun macroaggregato presente nel conto delle Risorse e degli Impieghi sono stati calcolati gli indici a catena con le formulazioni suggerite dal nuovo SEC, quali quelle di Laspeyres, di Fisher e di Paasche e, dove possibile, di Sato-Vartia. Nel caso dell'indice Sato-Vartia, la presenza di espressioni logaritmiche ne limita l'uso a quegli aggregati i cui valori correnti siano numeri positivi. Nonostante non sia annoverato tra gli indici suggeriti dal nuovo SEC (e nuovo SNA), a questa misura viene riconosciuta la capacità di soddisfare le principali proprietà elencate dalla teoria dei numeri indice. Pur tenendo conto di una formulazione molto articolata, l'indice di Sato-Vartia non si discosta mai dall'indice di Fisher.

Si è detto nelle pagine precedenti che nella metodologia del concatenamento la scelta del tipo di formula assume un ruolo marginale dovuto all'influenza moderata che gioca ciascuna ponderazione. L'evidenza empirica avalla la teoria dal momento che, dal confronto delle variazioni percentuali sull'anno precedente calcolate per tutte le tipologie degli indici a catena, non emergono sostanziali differenze. L'indice di Fisher media i pur minimi effetti dovuti alle sfasature temporali dei pesi utilizzati negli indici di Laspeyres e Paasche (non riportati nei grafici per evitare inutili appesantimenti nella lettura) e presenta, quindi, variazioni che si pongono tra i limiti superiori ed inferiori dati da questi ultimi. Alcune considerazioni di carattere generale possono essere tratte dall'osservazione dei grafici alle pagine seguenti, in cui i tassi medi annui degli indici a catena di Laspeyres e Fisher vengono confrontati con il tradizionale indice a base fissa di Laspeyres per tutti gli aggregati del conto delle Risorse e degli Impieghi.

In primo luogo, le differenze tra i tassi di variazione aumentano quanto più ci si allontana dal 1990, anno di riferimento della contabilità nazionale attuale. In particolare, la dinamica degli indici a base mobile è più accentuata prima per risultare più contenuta di quella del-

l'indice diretto dopo l'anno base. A questa uniformità di comportamento fanno eccezione gli indici a catena costruiti per le importazioni di beni e servizi (Fig. 2). Nel primo sottoperiodo, infatti, a fronte di un tasso medio annuo dell'indice a base fissa del 2,5%, l'indice a catena di Laspeyres presenta una crescita del 2,3% (e del 2,34% quello di Fisher). Nel caso specifico, il contenimento delle importazioni dovuto alla recessione del 1975 ha dato luogo, nel caso dell'indice a base mobile, ad una riduzione delle variazioni in volume più pesante rispetto al corrispondente indice a base fissa poiché la struttura dei prezzi relativi riferita al 1974 ha aggravato maggiormente la dinamica negativa registrata nel 1975.

In secondo luogo, poiché la divergenza tra i tassi di variazione dipende dalla diversa struttura di ponderazione adottata tra indici, essa dipenderà dalla maggiore o minore stabilità dei prezzi delle voci considerate nelle disaggregazioni: quanto più le articolazioni sono fini, tanto più si possono realizzare cambiamenti da un periodo all'altro e tanto più l'indice a catena si discosta dall'indice tradizionale. Nel caso dei consumi collettivi, la disaggregazione a due voci (consumi collettivi delle Amministrazioni Pubbliche e quelli delle Istituzioni Sociali Private) comporta una sostanziale uguaglianza nella dinamica degli indici (Fig.4).

Figura 1 – Tassi medi annui di variazione del prodotto interno lordo

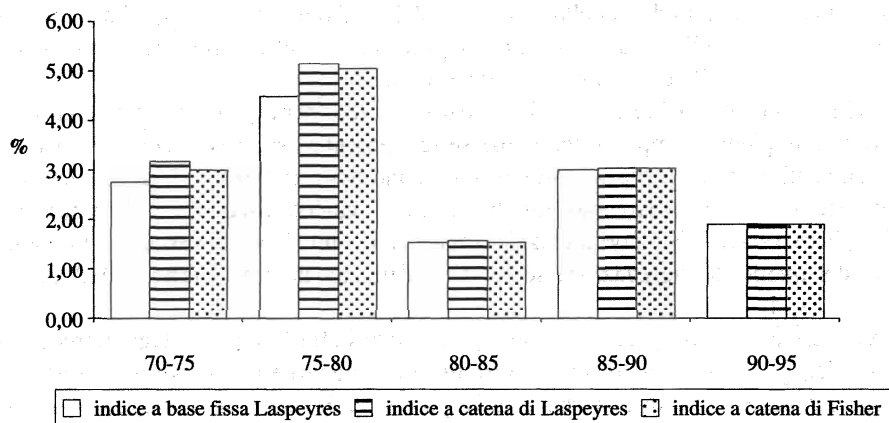


Figura 2 – Tassi medi annui di variazione delle importazioni

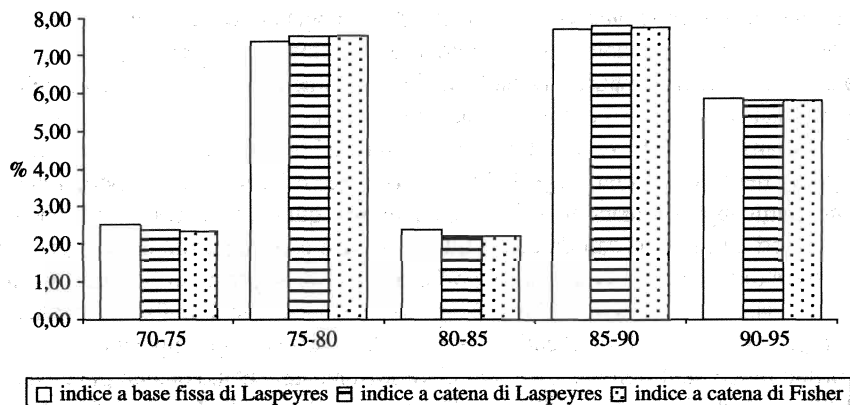


Figura 3 – Tassi medi annui di variazione dei consumi finali interni delle famiglie

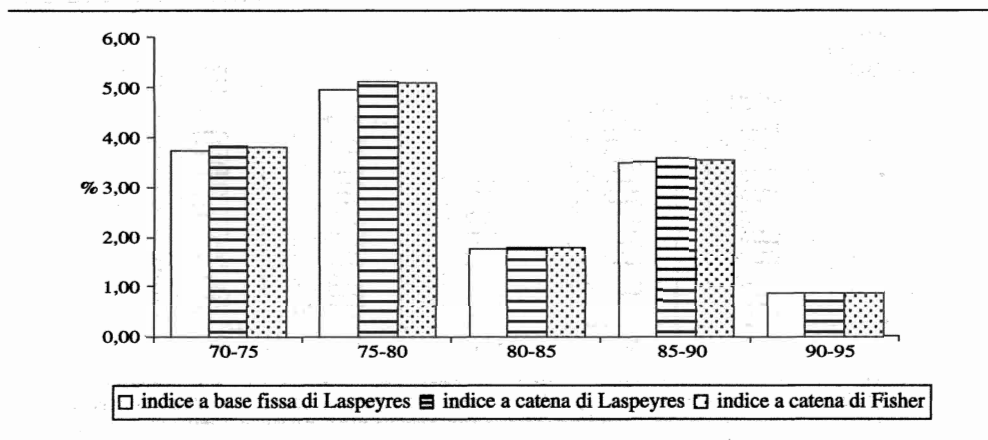


Figura 4 – Tassi medi annui di variazione dei consumi collettivi

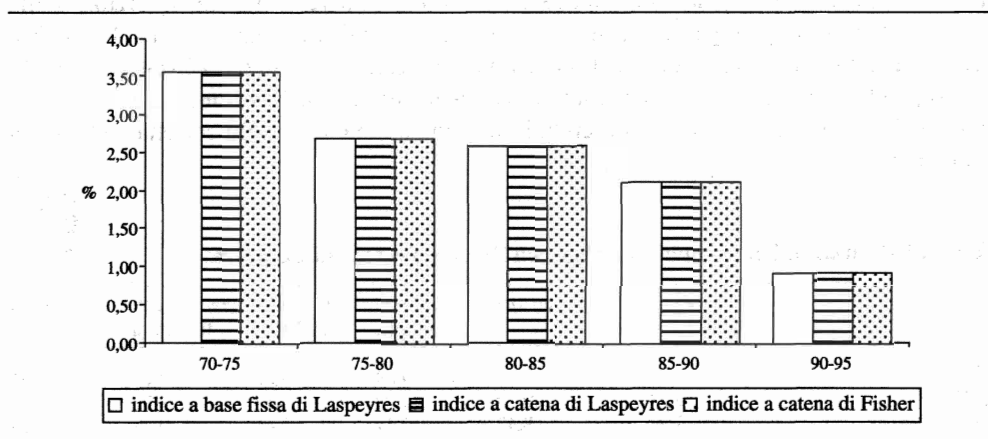


Figura 5 – Tassi medi annui di variazione degli investimenti fissi lordi

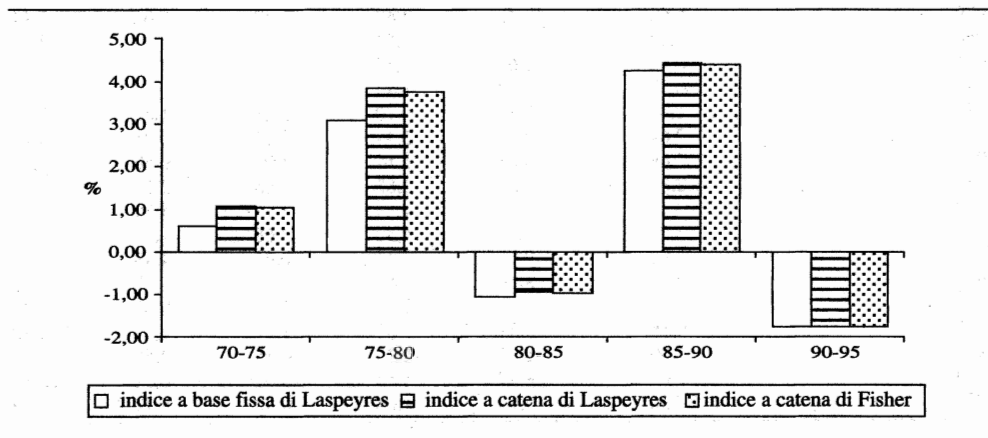
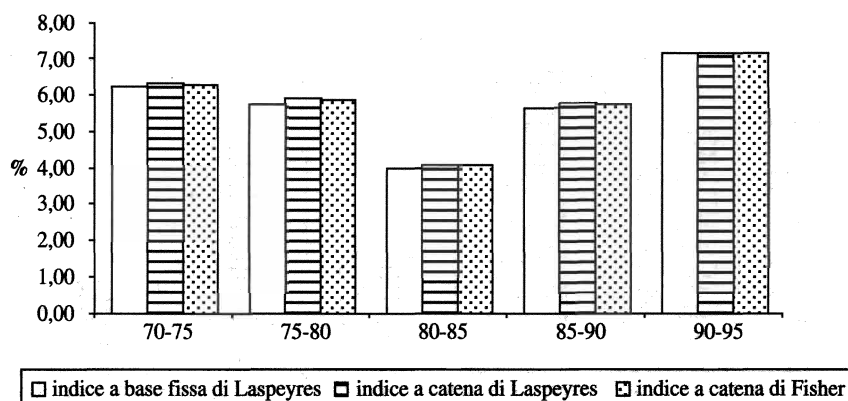


Figura 6 – Tassi medi annui di variazione delle esportazioni



Avendo verificato che in termini di variazioni percentuali gli indici presentano una forte somiglianza, si è optato per l'indice a catena di Laspeyres per la presentazione della dinamica dei macroaggregati del conto delle Risorse e degli Impieghi (Tabella 1). D'altra parte, si è detto che la superiorità del procedimento a catena si realizza a livello di singolo anello poiché ogni fenomeno è collegato ai cambiamenti che lo investono ed è per questo motivo che a tali indici viene riconosciuta la capacità di una lettura più corretta della realtà economica.

Tabella 1 – Variazioni percentuali calcolate con l'indice a catena di Laspeyres

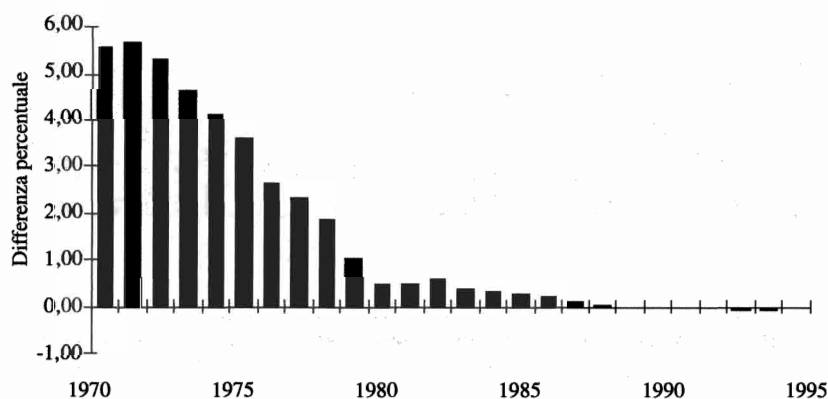
ANNI	Prodotto interno lordo	Importazioni	Consumi famiglie	Consumi collettivi	Investimenti fissi lordi	Esportazioni
1971	1,77	2,66	3,54	4,99	0,61	7,01
1972	3,31	10,49	3,99	4,83	1,63	7,65
1973	7,33	8,43	7,11	2,66	9,48	6,10
1974	5,23	5,83	3,94	2,73	2,35	9,70
1975	-1,58	-13,71	0,72	2,61	-7,96	1,33
1976	7,56	13,95	5,20	2,63	0,80	13,09
1977	3,20	1,11	4,14	2,53	2,51	8,67
1978	4,23	5,82	3,39	3,31	0,53	10,40
1979	6,58	12,45	7,40	2,87	7,06	7,44
1980	4,11	4,90	5,64	2,11	8,53	-8,65
1981	0,45	-1,64	1,32	3,25	-3,12	7,47
1982	0,35	-0,24	1,43	1,33	-4,69	-2,34
1983	1,43	-3,15	0,70	3,11	-1,06	3,47
1984	2,64	12,15	2,39	2,17	3,85	8,79
1985	2,89	4,63	3,14	3,09	0,50	3,39
1986	2,88	4,39	3,63	2,46	2,15	2,93
1987	3,19	11,47	4,12	3,17	4,79	4,54
1988	3,98	5,50	4,32	2,68	7,17	5,83
1989	2,91	9,09	3,41	0,94	4,43	8,93
1990	2,17	8,60	2,42	1,26	3,67	6,71
1991	1,14	2,72	2,68	1,68	0,84	-0,76
1992	0,58	5,57	1,01	1,15	-1,75	5,90
1993	-1,09	-8,05	-2,33	0,45	-12,56	9,08
1994	2,09	8,37	1,52	-0,01	-0,01	10,34
1995	2,96	9,63	1,66	-0,52	5,72	11,69

Oltre che con la notevole somiglianza dell'indice a catena con quello di Fisher, la decisione di ricorrere alla formulazione di Laspeyres si spiega soprattutto in virtù della proprietà dell'additività che il primo mantiene nei confronti tra anni contigui. Lo svantaggio teorico della perdita di coerenza degli schemi contabili viene ridotto con l'uso della formula di Laspeyres, che consente la costruzione di tavole riferite all'anno corrente espresse ai prezzi dell'anno precedente che soddisfino la condizione dell'additività.

Dal concatenamento delle variazioni dell'indice a catena, possono essere calcolate agevolmente le serie storiche in termini di volumi degli aggregati per il periodo 1970-1995 estrapolando il valore di un anno scelto come base: questo procedimento, che in realtà condurrebbe alla presentazione di serie di "pseudo-valori" a prezzi costanti, non rientra tra le indicazioni dei nuovi schemi internazionali di contabilità nazionale. Per questo motivo in queste pagine non si è ritenuto opportuno riportare le serie storiche ottenute dall'estrapolazione se non quella che si riferisce al prodotto interno lordo.

Seguendo tale tecnica, il PIL in base 1990 è stato calcolato estrapolando il valore aggiunto dell'anno base con la serie dell'indice a catena di Laspeyres (ovviamente per ottenere il Pil sono stati detratti i servizi bancari imputati e aggiunte le imposte sulle importazioni e sul valore aggiunto). L'analisi dei risultati (Fig. 7) mette in evidenza che le discrepanze tra la serie del PIL ufficiale e quella determinata con il concatenamento vanno aumentando, in valore assoluto, quanto più ci si allontana dall'anno base 1990.

Figura 7 – Differenze percentuali tra PIL a prezzi 1990 e PIL calcolato con l'indice a catena di Laspeyres



Dalla Fig. 7, in cui vengono mostrate le differenze percentuali tra PIL pubblicato e PIL calcolato con l'indice a catena, si constata che quest'ultima si mantiene al di sotto della prima fino al 1988 per poi risultare superiore dal 1992 in poi (la coincidenza dei valori nel 1990 e 1991 risulta automaticamente dalla metodologia). Il fenomeno appena evidenziato deriva dal fatto che il numero indice calcolato con l'aggiornamento annuale dei pesi si pone ad un livello inferiore rispetto a quello binario prima dall'anno base e superiore dopo.

Sempre dalla Fig. 7 si osserva che le discrepanze più sostenute, dell'ordine del 5%, si hanno nel primo quinquennio degli anni '70; avvicinandoci al 1990 le differenze tendono a chiudersi poiché i sistemi di ponderazione tendono a somigliarsi. A tale proposito merita un'attenzione particolare l'analisi delle variazioni annuali del prodotto interno lordo. Dalla

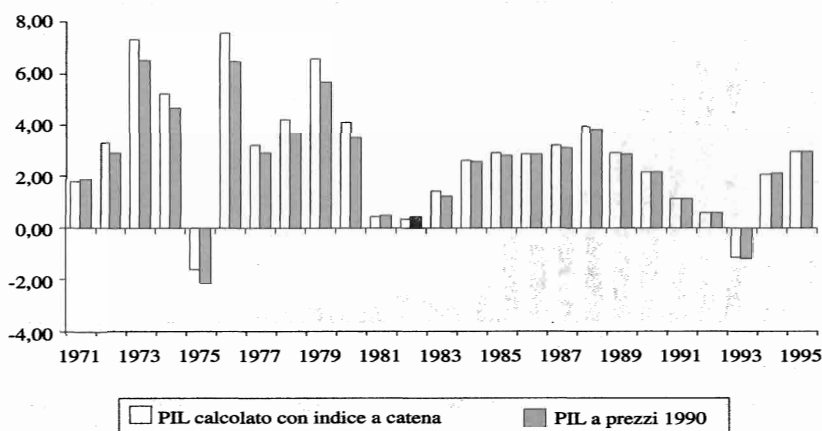
Fig. 8 è possibile riconoscere una certa regolarità: dopo il 1990 la crescita del PIL appare sovrastimata se misurata attraverso il procedimento tradizionale (tenendo fissi i prezzi dell'anno base), mentre negli anni precedenti l'anno di riferimento il meccanismo del *substitution bias* determina un risultato opposto (gli stessi risultati sono stati ottenuti da Landefeld e Parker per il PIL degli Stati Uniti).

Tale risultato appare scontato se solo si pensa agli enormi cambiamenti di prezzo e quantità dei prodotti energetici agli inizi degli anni '70 che, nell'ottica del concatenamento, producono dinamiche dell'output più vivaci rispetto a quelle derivanti da una ponderazione stemperata come quella del 1990.

Un ultimo aspetto che si è voluto verificare riguarda la perdita della proprietà dell'additività della metodologia degli indici a catena. A tale scopo ricordando che, nell'ambito della contabilità nazionale italiana viene raggiunta la coerenza degli schemi contabili per gli aggregati dell'offerta a livello delle 44 branche, il valore aggiunto di ciascuna branca viene determinato dal saldo tra produzione e costi intermedi.

Con tale presupposto, in Maresca (1995) si era proceduto ad un secondo metodo di calcolo del valore aggiunto (e quindi del PIL) utilizzando i numeri indici a ponderazione variabile. In questo caso l'aggregato saldo era stato ottenuto dalla differenza tra produzione e costi intermedi rielaborati estrapolando i dati del 1985 (all'epoca anno base della contabilità nazionale) attraverso le serie di indici a catena di volume.

Figura 8 – Variazioni annuali percentuali del PIL



Ripetendo tale simulazione con i nuovi dati di contabilità nazionale si ottiene un nuovo valore aggiunto in base 1990 che può essere confrontato con quello precedente realizzato dal concatenamento diretto dell'aggregato saldo (Tabella 2).

Le discrepanze che emergono tra le due serie del PIL quantificano gli effetti dovuti alla mancanza di additività della metodologia seguita, vale a dire alla impossibilità di aggiungere (in questo caso di detrarre) gli aggregati ottenuti con indici a catena. Tra l'altro con il secondo procedimento, ricalcolando il valore aggiunto dalla differenza tra produzione e costi intermedi significa reintrodurre il sistema di ponderazione presente nel 1990; vale a dire che la dinamica del saldo torna a dipendere fortemente dalla struttura degli aggregati nell'anno che si sceglie come base della catena.

Tabella 2 – Prodotto interno lordo calcolato con l'indice a catena di Laspeyres a prezzi 1990
(mld di lire)

ANNI	Pil ottenuto da val.agg. (a)	Pil ottenuto da prod.-costi (b)	differenze assolute c=a-b
1970	697147	751771	-54624
1971	709477	762146	-52669
1972	732970	781756	-48786
1973	786648	842849	-56201
1974	827825	879074	-51249
1975	814653	826573	-11920
1976	876287	884865	-8578
1977	904370	909989	-5619
1978	942678	945404	-2726
1979	1004572	1004170	402
1980	1045947	1045341	606
1981	1050583	1050566	17
1982	1054351	1054623	-272
1983	1069396	1068196	1200
1984	1097512	1098498	-986
1985	1129218	1129491	-273
1986	1161732	1160588	1144
1987	1198846	1197940	906
1988	1246554	1246295	259
1989	1282929	1283169	-240
1990	1310659	1310659	0
1991	1325582	1325582	0
1992	1333277	1333221	56
1993	1318749	1320114	-1365
1994	1346350	1347602	-1252
1995	1386199	1387408	-1209

Quest'ultima elaborazione ha il merito di mettere in risalto le conseguenze in termini di additività, ma essendo il Pil un aggregato che può essere ottenuto seguendo diversi approcci (dal lato dell'offerta, della domanda e della distribuzione) i risultati dipenderanno, di volta in volta, dalla metodologia di stima scelta.

5. Conclusioni

Per le sue caratteristiche, il metodo del concatenamento degli indici mostra una relativa superiorità rispetto al più tradizionale metodo degli indici a base fissa in quanto riesce a cogliere meglio la rapida evoluzione della realtà economica che ha caratterizzato gli ultimi decenni. Per questo motivo, tale metodo ha guadagnato un numero crescente di sostenitori nel corso degli anni.

In questo lavoro sono stati presentati i primi risultati ottenuti dall'applicazione della metodologia del concatenamento ai dati di contabilità nazionale italiana dato il vincolo della disponibilità dei dati attuali. In particolare sono stati calcolati gli indici a catena per gli aggregati del conto delle Risorse e degli Impieghi partendo dalla disaggregazione massima consentita e utilizzando le principali formulazioni suggerite dalla teoria dei numeri indice. Con riferimento alla disponibilità dei dati è indubbio che la realizzazione delle tavole *supply and use* annuali previste dai nuovi sistemi dei conti economici e le innovazioni in termini di costruzione degli indici dei prezzi al consumo¹⁵ giocheranno un ruolo importante per gli sviluppi futuri della tecnica del concatenamento nella contabilità italiana.

¹⁵ In EUROSTAT (1995) si ritrovano gli attuali orientamenti europei sugli indici dei prezzi al consumo.

Una prima considerazione di carattere generale che si può trarre dal confronto dell'evoluzione degli indici a catena con l'indice a base fissa di Laspeyres, utilizzato nelle attuali misurazioni in volume della contabilità nazionale, è che l'approccio a ponderazione variabile tende ad eliminare il fenomeno del *substitution bias* caratteristico del procedimento tradizionale.

Dalle elaborazioni svolte si è verificato empiricamente che l'adozione di formule alternative (di Laspeyres, di Paasche, di Fisher, di Sato-Vartia) nella tecnica del concatenamento ha una modesta rilevanza sui risultati ottenuti in termini di variazioni degli indici. Infatti, la superiorità dell'indice a catena rispetto all'indice a base fissa (sancita oramai dalle organizzazioni internazionali) si manifesta, tra l'altro, per il fatto che la scelta della formula ha scarsi effetti pratici.

In particolare, gli esercizi condotti nel presente lavoro evidenziano che l'indice a catena di Laspeyres rappresenta una buona approssimazione dell'indice di Fisher consigliato dai nuovi sistemi contabili ma ha il vantaggio non trascurabile di garantire il bilanciamento dei conti economici se espressi ai prezzi dell'anno precedente. Pertanto, se si opta per la formulazione di Laspeyres, con la metodologia del concatenamento è possibile presentare per ogni anno conti economici a prezzi dell'anno precedente in cui vengono rispettati i vincoli degli schemi contabili; inoltre, con il *linking* dei singoli anelli annuali si ottiene un indice a catena esteso a più intervalli temporali. Con gli indici a catena, quindi, viene soddisfatto compiutamente il bisogno di gran parte degli utilizzatori di conti economici interessati ai cambiamenti "reali" che si verificano da un anno all'altro e viene mantenuta la possibilità di costruire serie storiche di numeri indice.

Il principale problema che si riscontra con l'adozione della metodologia del concatenamento è dato dal mancato rispetto della condizione dell'additività, e ciò, in un ambito contabile, implica la perdita di coerenza dei conti economici che ne derivano. Per questo, i nuovi sistemi di contabilità nazionale, oltre a raccomandare il calcolo dei numeri indici a catena per la misurazione delle variazioni dei volumi e dei prezzi per ciascun macroaggregato,¹⁶ consigliano di continuare ad elaborare dati disaggregati ai prezzi di un anno base con la tecnica tradizionale. Tra l'altro, l'esperienza dei Paesi che utilizzano già fin da ora l'approccio del concatenamento (USA, Olanda, Norvegia) sembrerebbe suggerire questa soluzione; infatti, nessuno di questi Paesi ritornerebbe ad un sistema a base fissa come unico modo di valutare le grandezze in termini reali.¹⁷

Anche per la contabilità del nostro Paese, la proposta migliore sembrerebbe quella di costruire parallelamente i due sistemi ad uso interno e sperimentale per i prossimi anni; al momento dell'applicazione del SEC95, la fornitura di stime basate sul concatenamento dovrebbe essere corredata di un'adeguata informazione sull'entità e il segno delle discrepanze con il sistema a base fissa, e dovrebbero essere messe in evidenza le variazioni nella composizione degli aggregati, per mostrare agli utilizzatori l'inadeguatezza dell'uso esclusivo e acritico degli indici a base fissa.

In queste circostanze però va osservato che l'utilizzatore dei dati di contabilità nazionale si trova davanti a due variazioni, supponiamo del PIL, diverse perchè ottenute con due metodologie differenti: il calcolo diretto ed il concatenamento.

Anche se per alcune categorie di utenti l'esistenza di una doppia versione potrebbe generare incertezze sulla scelta delle informazioni da usare, occorre ribadire che solo con il calcolo degli indici a ponderazione variabile è possibile avere la migliore rappresentazione della crescita reale degli aggregati economici.

¹⁶ Uno sforzo maggiore andrebbe fatto anche in sede internazionale al fine di stabilire in modo univoco a quali macroaggregati ci si debba riferire.

¹⁷ A tale proposito è interessante evidenziare il caso dell'Istituto statistico della Norvegia che adottando il sistema a base fissa nei conti economici annuali e quello ai prezzi dell'anno precedente per le stime trimestrali si è trovato nel 1989 a dover optare per uno dei due sistemi (quello a ponderazione variabile) poiché le divergenze in termini di variazioni erano divenute insostenibili e incomprensibili per gli utilizzatori (Flottum, 1996).

A questo stato dei lavori è comunque indispensabile riflettere sulla portata del cambiamento che il passaggio al nuovo approccio potrebbe provocare nel modo di presentare i dati e le conseguenze che ne deriverebbero per chi utilizza le stime in termini di volume nel quadro di uno schema contabile chiuso.

References

- ALLEN R.G.D. (1975), "Index numbers in theory and practice", N.Y., Mc Millan.
- BIGGERI L. (1986), "Recenti sviluppi nel campo dei numeri indici", in *I numeri indici nella teoria e nella pratica*, a cura di L. Biggeri, ISCO.
- DALGAARD E. (1994), "Implementing the revised SNA" recommendations on price and volume measures, documento presentato alla 23^{ma} conferenza della I.A.R.I.W., Canada.
- EUROSTAT (1981), "Sistema europeo di conti economici integrati (SEC)", Lussemburgo.
- EUROSTAT (1996), "European System of Accounts ESA 1995" Luxembourg.
- EUROSTAT (1995), "Improving the quality of price indices" in Atti dell' International Seminar, Florence, dec. 18-20, 1995.
- FLOTTUM E. J. (1996), "The 1995 General Revision of the Norwegian National Accounts", dattiloscritto.
- FORSYTH F.G. (1978), "The practical construction of a chain price index number", J.R.Statist.Soc.A.
- FORSYTH F.G., FOWLER R.F. (1986), "La teoria e la pratica dei numeri indici a catena dei prezzi", in *I numeri indici nella teoria e nella pratica*, ISCO.
- HILL P. (1988), "Recent developments in index number theory and practice", OECD Economic Studies.
- LANDEFELD J.S., PARKER R.P. (1995), "Preview of the Comprehensive Revision of the National Income and Product Accounts: BEA's New Featured Measures of Output and Prices", *Survey of Current Business*.
- MARESCA S. (1995), "L'indice a catena per le misurazioni di prezzo e di volume", par.8.2 del cap."Il nuovo sistema di Contabilità Nazionale. Problematiche applicative in Italia", in *Atti della Seconda Conferenza Nazionale di Statistica*, Tomo 1, 1995, Sistan-Istat.
- MARTINI M. (1992), "I numeri indice in un approccio assiomatico", Milano, Giuffrè.
- MARSHALL A. (1887), "Remedies for fluctuations of general prices", *The Contemporary Review*.
- MOULTON B. R. (1993), "Basic components of the CPI: estimation of price changes", *Monthly Labor Review*, volume 116, number 12.
- OFFICE STATISTIQUE DES COMMUNAUTES EUROPEENNES (1987), "L'utilisation des indices-chaine dans les comptes nationaux".
- ONU, "Manual on National Accounts at constant prices" (1979), *Statistical papers*, SERIES M, n 64, N.Y.
- QUIRINO P. (1978), Le valutazioni a prezzi costanti nel quadro della contabilità nazionale, *Rivista di Politica Economica*, n 8-9.
- SIESTO V. (1977), *La contabilità nazionale italiana*, Bologna, Il Mulino.
- SZULC. B. (1983), "Linking price index number", Diewert W.E. Montmarquette C. (eds.), *Price level measurement*, Minister of Supply, Canada, Ottawa.
- TRIPLETT J.E.(1992), "Economic theory and BEA's alternative quantity and price indexes", *Survey of Current Business*.
- YOUNG A.H. (1992), "Alternative measures of change in real output and prices", *Survey of Current Business*.
- UNITED NATIONS, EUROSTAT, IMF, OECD, WORLD BANK (1993), "System of National Accounts 1993", Brussels, Luxembourg, Paris, Washington, D.C.

LABOR MARKET STRUCTURE AND WELFARE: A COMPARISON OF ITALY AND THE U.S.

Christopher J. Flinn

Department of Economics - New York University

1. Introduction

Institutional environments are often compared in terms of the distributions of individual and family earnings and income which are observed within them. For example, there exists a large literature which contains descriptions of income and/or earnings inequality across countries, some recent examples of which include Levy and Murname (1992), Gottschalk (1993), Gottschalk and Smeeding (1995), Atkinson *et al.* (1995). Typically, normalized measures of dispersion such as the coefficient of variation or Gini index are computed using cross-sectional earnings or income distributions for a number of countries and/or computed at various points in time within countries.¹ In this research we make the point that the ranking of institutional environments using inequality measures computed from cross-sectional earnings distributions may be very different from the ranking obtained when lifetime measures of inequality are used.

The observation that cross-sectional distributions of compensation or welfare measures may not perfectly mirror their life-cycle analogs is certainly not new. One of the clearest statements of the issue investigated in this paper is provided by Friedman (1962, p. 171):

A major problem in interpreting evidence on the distribution of income is the need to distinguish two basically different kinds of inequality; temporary, short-run differences in income, and differences in long-run income status. Consider two societies that have the same distribution of annual income. In one there is great mobility and change so that the position of particular families in the income hierarchy varies widely from year to year. In the other, there is great rigidity so that each family stays in the same position year after year. Clearly, in any meaningful sense, the second would be the more unequal society. The one kind of inequality is a sign of dynamic change, social mobility, equality of opportunity; the other, of a status society.

¹ Good examples of the later type of study, which are also relevant to the comparisons made in this paper are contained in Brandolini and Sestito (1994, 1996). These authors compare year-to-year variations in earnings inequality in Italy over the last two decades using a variety of inequality measures.

This research was conducted while the author was a fellow at ICER [Torino, Italy] and was partially supported by the C.V. Starr Center for Applied Economics at New York University. I am especially grateful to Ugo Trivellato and Enrico Rettore for making the Italian data used in this research available to me. Helpful comments were received from seminar participants at Bocconi, Padova, Penn, the IRP at the University of Wisconsin, DELTA, the Equilibrium Wage Distributions Conference (Paris), the Savings Workshop (Tilburg), the European University Institute, and NYU. Giuseppe Bertola, Tito Boeri, Douglas Gale, and Ramon Marimon provided useful suggestions which are partially reflected in this draft. I am responsible for all errors and omissions

There are at least three approaches to the analysis of long-run welfare inequality which could be taken. The first approach is a strictly empirical one in which data on compensation over the lifetimes of groups of individuals is first time-aggregated, with the distribution of the averaged compensation measures becoming the focus of the analysis [see, for example, Aaberge *et al.* (1996), Björklund (1993) and Blomquist (1981)]. This approach has as its main advantage the fact that no modelling framework, with its attendant assumptions, need be adopted to compare inequality in different institutional environments. This is a very great advantage, especially when one attempts to compare compensation distributions across institutional environments in which individual behavior takes place in qualitatively different contexts. The main disadvantages of such an approach are that the requisite data for performing it are typically not available, and the absence of an explicit modelling framework makes it difficult to determine exactly which environmental characteristics are the key ones in accounting for differences in the longrun distributions of welfare outcomes.

A second approach is to change the focus of the analysis so as to explicitly analyze income or earnings dynamics outside of a specific modelling framework. The objective of this type of study is, treating compensation over the life-cycle as an exogenous stochastic process F , to find summary measures characterizing the degree of mobility, $\sigma(F)$. Several mobility measures have been proposed by sociologists and economists, and there exist a number of empirical studies in which the mobility process is the focus of interest.² The advantage of this approach is that it is explicitly dynamic and can provide useful summary information regarding how much mobility exists in an environment – the summary measures produced are explicitly designed for comparative analysis in that they are unitless. The principle disadvantage is in not being able to explicitly value [in welfare terms] the mobility process associated with a given environment. This is only to say that static distributional and mobility analyses are only pieces of the puzzle – neither is sufficient in and of itself to determine the distribution of life-cycle welfare. Another disadvantage, shared with the first approach, is the failure to link the degree of mobility present in an institutional environment to particular features of the environment.

The third approach, which is the one taken in this paper, involves the computation and comparison of life-cycle welfare distributions across environments. Any comparison of the distribution of welfare outcomes of necessity must begin with an explicit definition of welfare determination. Here we adopt one of the standard frameworks utilized in analyzing life-cycle compensation and employment dynamics, that of off- and on-the-job search.³ There are two principle advantages of this approach. First, we are able directly to compute a distribution of welfare outcomes under the maintained assumptions of the model. Second, we are able to link characteristics of the environment, in this case the labor market, to the properties of this distribution in a more or less immediate way. The disadvantage of this type of analysis is that it must be conducted under a number of relatively artificial and, more importantly, untestable assumptions. Therefore, the results of such an analysis must be interpreted with a large degree of caution. A more minor disadvantage is the problem of comparability of distributional characteristics across environments - i.e., the units problem.

² In terms of the derivation of mobility indices, some examples of influential work on the subject are Prais (1955), Shorrocks (1978a, 1978b), and Geweke *et al.* (1986). For a recent example of empirical work focused on earnings mobility see Aktinson *et al.* (1992).

³ There are a number of other models one could consider using as a framework for this analysis, such as job matching [see Jovanovic (1979), Miller (1984), and Flinn (1986), human capital investment, or efficiency wages (Flinn (1997))]. In order to estimate a matching model one should have access to more than one wage observation per job. Human capital investment models typically lack implications for the turnover process, which is a focus of our analysis. Efficiency wage models can only properly be estimated when cast in an equilibrium framework, and it is likely that the nature of the equilibrium may differ qualitatively across environments. We therefore settled on the search framework as the simplest behavioral model which captures labor market dynamics relatively well and which could be estimated with the data available to us.

In this paper we deal with this problem by comparing only a unitless characteristic of each distribution, the coefficient of variation (CV).

We have chosen Italy and the U.S. to make our substantive points because the labor market institutions in these two countries are well-known to be quite dissimilar. Our analysis allows us to identify the effects of characteristics of the labor market environment, such as the rates of receiving offers while unemployed or employed, the rate of dismissal, and the wage offer distribution, on the distribution of life-cycle welfare.

There exist a number of studies which compare the labor markets of Italy and the U.S., both in terms of institutional differences and empirically [see, for example, Grubb *et al.* (1983), Del Boca (1988), Bertola and Ichino (1995), Bertola and Rogerson (1997), Blanchflower *et al.* (1993), Demekas (1994)]. In virtually all of the empirical studies, differences in the industrial relations systems and/or output markets are invoked to explain patterns of intertemporal relationships between comparable (typically aggregated) labor market measures in the two countries. A popular focus of this type of analysis is on the effect of the severe restrictions on layoffs in Italy on the demand for labor over the business cycle [variability in employment over the cycle is substantially less in Italy than in the U.S.].

While the Italian labor market has generally become less regulated over the past few decades, it remains to this day substantially more regulated than the U.S. market. In this paper we make comparisons of labor market dynamics across the two countries by exploiting retrospective information collected from a sample of individuals from the Lombardia region of Italy [which includes Milan] and data on the labor market experiences of white males of the same ages from the representative portion of the National Longitudinal Survey of Youth [NLSY]. To our knowledge, this is the first study to compare the two labor markets using a dynamic model in conjunction with individual-level event history data.

While we will not attempt to give a detailed account of the Italian industrial relations system, we will provide a brief overview of some of its more important features. Consider first the hiring process. Until recently, Italy was the only European country to attempt to regulate precisely whom an employer could hire through the distribution of lists with candidates ranked in terms of criteria which mainly related to their "need" for employment [i.e., whether they have dependents, whether they have a disability, etc.]. While this system was abolished in 1991, it was still partially in effect during the time period covered by our sample.

It remains notoriously difficult to fire or lay off individuals or groups of individuals within the Italian system. In order to permanently dismiss someone a quasi-legal process must be followed, with evidence of serious misbehavior on the part of employees available for substantiation; if the employer is judged to have brought a false case against an employee, punishments are severe. It is even difficult for an employer to lay off individuals or groups of individuals because of redundancy; in such a case employers are required to consult with union representatives and/or the government in an attempt to lessen the negative impact of separations on employees. Once again, individuals or groups of employees can contest layoffs, and the firm is punished if it is found to have misrepresented its case for layoffs.

Italian firms also face severe restrictions with respect to the way in which they can utilize their work force. Despite the relaxation of some regulations pertaining to the work intensity margin in the 1980s, it is difficult for Italian firms to hire workers under fixed-term contracts and to restrict hours [the share of part-time to total employment in Italy is the lowest in Europe].

The wage-setting process, while having been liberalized in the last decade, still remains highly centralized in comparison with the U.S. Baseline agreements are negotiated at national and regional levels between employer organizations and the major trade

unions. During the 1980s, firm-level bargaining with individual unions became increasingly important, and as a result wage differentials between various classes of workers have been widening. The change which occurred during the 1980s came about due to shifts in the bargaining power of employers and unions rather than due to changes in regulatory laws *per se*. The wage determination system was formally changed in new laws introduced in 1992-93.

Because of the variation in labor market regulations across states in the U.S., it is not possible to give an analogous overview of the U.S. regulatory environment, but such an exercise is really not required for our purposes. It is clear that the Italian system is much more regulated than what is found in any locality in the U.S. on every dimension.

In the context of a model of labor market dynamics estimated using individual-level data from the two countries, what do the differences in the labor market structures lead us to expect? Perhaps most obvious is the implications that dismissal rates will be higher in the U.S. While in the U.S. employees, especially union members, do have recourse to procedures to contest firings of layoffs, these procedures seem almost inconsequential compared to those in place in Italy.

Due to the extreme difficulty of dismissing or modifying the employment conditions of employees, it is also reasonable to expect that the rate of receiving offers would be lower in Italy than in the U.S., other things equal. This supposition is supported by the macro time series evidence and by descriptive evidence at the individual level on the duration of unemployment spells and tenure at one's current employer. Thus we expect to find the rate of receiving offers both in the unemployed and the employed states to be lower in Italy.

Because of the centralized wage-setting process and the restrictions on varying employment intensity in Italy, it is natural to expect [and indeed is consistent with conventional wisdom] that both wage and earnings distributions would be more concentrated in Italy than in the U.S. Figures 1.a and 1.b provide some evidence on this issue. These panels contain histograms of hourly wages of young white male U.S. employees from the NLSY who were employed in January 1988 and the hourly wages from a same-age group from Lombardia employed in this same month. The U.S. wage distribution appears to be more right-skewed than the comparable distribution for Lombardia.⁴ In terms of the coefficient of variation associated with the distributions, the CV for the U.S. wage distribution is almost twice as large as the Italian CV. Observations like these have led some observers to claim that the regulatory system in Italy, while most probably leading to inefficiency in resource allocation, has led to a socially-beneficial equalization of labor market outcomes across individuals. This is contrasted with "unregulated" markets like that of the U.S., which while more efficient, come at the cost of large disparities in welfare outcomes.

Using a stationary search model, we shall show that the distribution of life-cycle welfare outcomes in the two countries is very different from the corresponding cross-sectional distributions. The distributions of ex post welfare outcomes based on simulations using the behavioral parameter estimates for Lombardia and the U.S. are presented in Figures 1.c and 1.d. The lifetime welfare distributions for both countries are less skewed than their corresponding distributions of wage rates sampled at a point in time. Moreover, the coefficient of variation of lifetime welfare is substantially less than the CV for sampled wage rates. Most striking is the fact that the CV of lifetime welfare for the U.S. is less than the CV of lifetime welfare for Lombardia.

⁴ The right-skew observed in the U.S. data would be even more pronounced if two extremely high wage observations were not first deleted from the U.S. sub-sample we extracted. Based on the demographic characteristics of the individuals involved, we decided that the wage draws were "unbelievably" high.

Figure 1.a
Hourly Wages
Italy

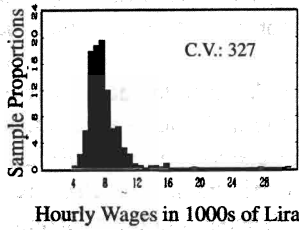


Figure 1.b
Hourly Wages
U.S.

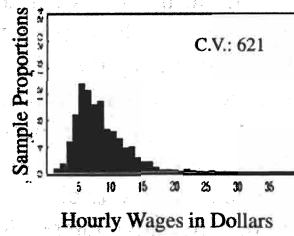


Figure 1.c
Ex Post Career Value
Italy

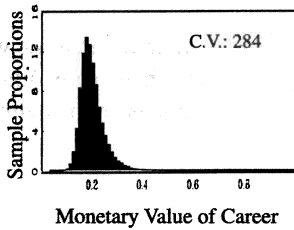


Figure 1.a
Ex Post Career Value
U.S.

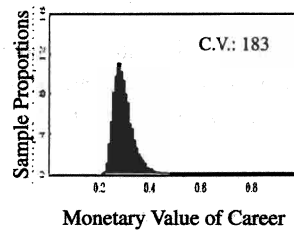


Figure 1.c
Steady State Wage Density
Italy

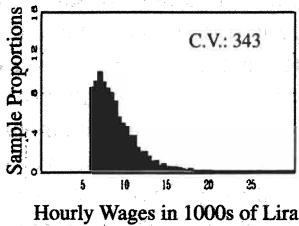
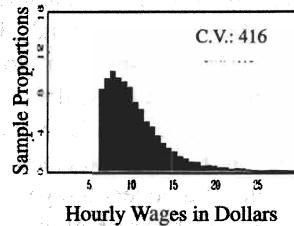


Figure 1.c
Steady State Wage Density
U.S.



Our model of lifetime welfare determination also carries implications for the cross-sectional distribution of wages, either in the steady state or outside of it. There exists a complex, but nonetheless relatively intuitive, mapping between the parameters which characterize the search environment and the distributional properties of steady state wages. In particular, the model illustrates the manner in which equality in life-cycle welfare realizations is consistent with inequality in the cross-sectional wage distribution. Figures 1.e and 1.f contain plots of the steady state wage densities, evaluated at the model estimates, for Italy and the U.S. Note that although life-cycle welfare outcomes are less disperse in the U.S., the CV of the steady state wage distribution is larger for the U.S. We will attempt to identify the key features of the Lombardian and U.S. labor markets and of our model which are responsible for producing these results.

The plan of the paper is as follows. In Section 2 we describe the data sources and provide descriptive evidence on labor market dynamics at the individual level in the two markets. Section 3 provides the theoretical structure for our analysis; in it we define the labor market environment and the decision rules used by participants. We also present some motivation for our substantive interest in the degree of dispersion in the realizations of life-cycle welfare. Section 4 contains a discussion of econometric issues and the maximum likelihood estimator we utilize. We present estimation and simulation results in Section 5, and end with a brief conclusion in Section 6.

2. Data description and descriptive statistics

In this section we describe the data to be used in all of the empirical analysis which follows, and then proceed to present some descriptive statistics which serve to illustrate the marked differences in transition rates between labor market states in these two environments.

2.1. Data description

We begin by describing the event-history data for the U.S. labor market, which is taken from the random sample component of the National Longitudinal Survey of Youth [NLSY]. This data set is very well-known to most labor market researchers, so only a brief discussion of the information we have extracted and the sample-selection criteria will be presented.

As we describe below, the Italian data utilized come from a random sample of households in the Lombardia region of Italy. For such a large industrial and commercial area [it includes Milan, the most important commercial center in Italy], it contains a relatively homogeneous labor market in the demographic sense. For purposes of comparison, we have restricted the U.S. sample to include only whites in an attempt to limit the amount of sample heterogeneity in background characteristics. Furthermore, since the observation period for the Lombardia data covers the period January 1988 through May 1989, we have constructed a continuous labor market history from the NLSY data which covers only this period as well. Since the NLSY sample members were 14-21 years of age in the initial survey year of 1979, they are [approximately] between the ages of 23-30 during the observation period. We have further restricted the samples to include only males, not only for the usual reason of not wishing to confront the problem of modelling the participation decisions of married women, but also because the patterns of female labor force participation are so different in the two environments.

For this sample of white males we have constructed event histories which include information on the duration of each job held, where a "job" is defined as an employment spell spent *with a particular employer*, and the hourly wage rate received on the job; we also have information on the length of time spent in each spell of nonemployment. We do not attempt to distinguish between periods spent actively searching and those spent out of the labor force - all are aggregated into the category "nonemployment" [though to avoid redundancy we shall sometimes call this state "unemployment"]. We have excluded from our final sample all individuals who reported never actively searching for employment or who did not hold a job over the observation period. Thus everyone in the final sample is a labor market participant in the formal sense of the term at some point in the 17 month period. We have also excluded individuals who reported being enrolled in school or in the military during the observation period, as well as all those remaining individuals who were not in the random sample component of the NLSY.

The Italian Statistical Institute [ISTAT] survey was conducted in Lombardia during the summer of 1989, and gathered information from a sample of households [as opposed to the NLSY, in which the sampling unit is the individual]. Respondents were asked to describe their labor market experiences rather completely over the 17 month period which spanned January 1988 through May 1989. In particular, respondents reported their labor market status on a month-by-month basis over this period. They were then asked to provide detailed information on up to two jobs held as a dependent worker and one job as an independent worker during this period. All duration information was given solely on a monthly basis.

Until recently, in Italy sample members of social surveys have seldom been asked for information relating to their earnings or income from other sources. Given that wage and/or earnings information is an essential requirement for estimating a structural search model, this is the only Italian survey information in existence which can be used for this purpose.⁵ Respondents were asked, as in the NLSY, to report a typical level of earnings for each job occurring during the observation period. Thus the duration and remuneration information collected is quite similar in the two surveys.

We conclude this discussion with three notes on the use of these samples for comparative analysis. First, labor market dynamics at the individual level, especially for young labor market participants, are likely to be quite sensitive to aggregate labor market conditions. In order to perform a comparative analysis which heavily relies on stationarity assumptions in order to extrapolate labor market dynamics from a brief historical period to the entire life-cycle, it is at a minimum necessary to ensure that the two economies were not at very different stages in the business cycle during our observation period. Fortunately for us, in 1988 and 1989 the U.S. and Italy [we don't have access to aggregate data specifically relating to Lombardia] seem to have been at similar stages in the business cycle. We arrived at this judgement by computing standardized values of annual Gross Domestic Product in constant local prices for Italy and the U.S. over the period 1980-1993.⁶ The values of this standardized measure in 1988 and 1989 for Italy were .501 and .828 in comparison with .568 and .828 in the U.S. When we computed these measure after first detrending each series, we arrived at scores of .720 and 1.170 for Italy and 1.221 and 1.359 for the U.S. Thus, on the basis of either set of values, both economies seem to have been in relatively "good" states, though neither was booming.

⁵ There are some administrative records from the Italian social security administration [IMPS] which could possibly be used, but it is difficult to ascribe earnings to a particular job using these data. In addition, only jobs covered by social security provision are included; for example, all public sector jobs are excluded.

⁶ That is, from each series we subtracted the mean over the period and divided the result by the standard deviation of the series. Thanks to Giuseppe Bertola for providing us with the series.

The second note concerns the absence of standard sets of "covariates" in the theoretical model and empirical work. We decided at an early stage not to attempt to perform this analysis conditional on observable characteristic of respondents [other than those used in the sample selection criteria]. While it is undoubtedly of interest to examine the manner in which environment-specific institutions serve to differentiate the population, with its resulting implications for lifetime welfare inequality, such an analysis is beset with problems at both the conceptual and practical levels. For example, while schooling information is ascertained in each of the two surveys, it is a difficult if not impossible task to develop a standard measure of educational attainment which can be used in the two environments.⁷ Because the focus of the current analysis is on labor market dynamics and life-cycle welfare issues very broadly defined, we decided to use a "homogeneous" population approach so as to sidestep these problems.

Thirdly, we are obliged to point out a problem of potential noncomparability with regards to the wage measures used in this study. The ISTAT survey in Lombardian explicitly asks respondents to report their *net* [i.e., after-tax] monthly earnings. Hourly wages are imputed by dividing this measure by the reported usual hours worked at the employer. On the other hand, in the NLSY individuals are asked to report how much they earned when they first began working at the employer. The respondent is not told to report gross [i.e., pre-tax] or net earnings, so that we can only speculate as to how respondents interpreted the question. Anecdotal evidence suggests Americans are more apt to report gross earnings when asked about compensation, while Italians are more likely to report net earnings. If members of the U.S. sample report gross earnings, how is this likely to affect the comparative analysis? The answer depends in part on the nature of the tax system. If it is proportional or progressive, the gross wage distribution will be more dispersed than the net wage distribution. In this case, cross-sectional differences in the wage distribution for Italy and the U.S. may be over-stated. Since the principle message of our analysis is that variation in welfare outcomes in the U.S. labor market may be smaller than those which exist in Lombardia, we would argue that having comparable measures [say net wages] would probably serve to further strengthen these findings.⁸ Unfortunately, we can't provide any further evidence on this issue with the data at our disposal.

2.2. Descriptive statistic

Table 1 contains some descriptive statistics for our samples of U.S. and Lombardian labor market participants. First note that there are approximately four times as many individuals in the U.S. sample. In terms of numbers of spells, the discrepancy is even larger since as we shall see transitions between labor market states are much more frequent in the NLSY than in the ISTAT sample.

⁷ Neither is it clear that it would be desirable to compare groups in the two environments which inhabit the same quantile, for example, in their environment's distribution of educational attainment. In fact, for either environment, but especially for Italy, it is probably quite misleading to think of educational attainment as being unidimensional since the curriculum followed in school has a strong effect on subsequent labor market opportunities.

⁸ It is tempting to argue that, since the coefficient of variation is invariant with respect to rescaling of the original variable, our results will be invariant with respect to changes in constant marginal tax rates. While this is true of the raw data on wages, it is unfortunately not the case with respect to results obtained from the simulations. This is because the structural parameter estimates from which the simulations are generated are not invariant with respect to a rescaling of wages.

Table 1 – Descriptive Statistics for Lombardia and U.S. Samples

CHARACTERISTIC	Lombardia	U.S.
Sample size	395	1521
Nonemployment		
Some	59 (.149)	432 (.284)
Always	14 (.035)	13 (.009)
Employment		
Same job all 17 months	324 (.820)	846 (.556)
Two or more employments spells	16 (.041)	535 (.352)
Number of employment spells with two or more consecutive jobs	12	402
Wage gain [i.e., $\omega(2) > \omega(1)$]	11 (.917)	286 (.711)

Nonemployment experiences were much more common in the U.S. than in Lombardia. In Lombardia, 15 percent of the sample experienced some nonemployment during the period as opposed to 28 percent in the U.S. However, less than 1 percent of the U.S. sample remained unemployed for the entire 17 month period, as opposed to 3.5 percent of the Lombardian sample. Job changes were much more common in the U.S. than in Italy; 82 percent of the total Italian sample worked at the same job in all 17 months while only 56 percent of the U.S. sample did not change job over the observation period. Only 4 percent of the Italians had two or more employment spells as opposed to 35 percent of the Americans. These figures highlight the very large differences in labor market transition rates between the two countries. The difference in job-changing rates is particularly striking and will be reflected in the structural parameter estimates and simulation results presented below.

Under the assumptions upon which the search model we estimate is built, in the absence of measurement error in wage rates, all job-to-job changes [with no intervening spell of nonemployment] should be accompanied by an increase in the wage rate. In the Italian sample, no individual reported having more than two jobs over the 17 month period. Of the 16 who reported having two jobs, 12 reported that the jobs were consecutive. Of these 12, the wage on the new job was greater than the wage on the first in 11 cases.

For the U.S. sample, 535 individuals reported having two or more jobs. Within this set of individuals, 389 individuals had two or more consecutive jobs, and some had multiple consecutive job spells within the 17 month period. The total number of employment spells containing two or more consecutive jobs is 402. Of these 402 spells, the second job in the spell paid a wage greater than that paid by the first job in 286 cases, yielding a "success rate" of .711. Thus in both samples, the tendency to have higher wages in the second job is clearly present, though especially in the U.S. sample, it is far from always the case. When we discuss estimation issues below we will consider ways in which the search model can be modified so as to account for these discrepancies.

3. The behavioral model

In this section we set out the behavioral model used in performing the empirical work below, which is one of off- and on-the-job search in a stationary labor market environment. As is common in the search literature, we assume that agents are risk-neutral expected-wealth maximizers. After describing the labor market environment and the characteristics of the decision rules used by labor market participants, we go on to briefly consider agents' attitude toward dispersion in the *ex post* value of labor market careers.

3.1 The stationary model of off- and on-the-job search under expected wealth maximization

We assume that agents are infinitely-lived,⁹ and that at each moment in time they occupy one of the following labor market states. They may be nonemployed and searching for work, a state denoted by n , or they may be employed in a job which pays a wage w – this state will be denoted $e(w)$. Job offers arrive to nonemployed individuals according to a Poisson process with parameter λ_n , and to employed individuals according to a Poisson process with parameter λ_e . For both employed and nonemployed searchers the value of the offer made is independent of the arrival process. Employment spells terminate either because (1) an employed searcher locates a better job or (2) an exogenous separation occurs. Exogenous separations arrive according to a Poisson process with parameter η ; note that the rate of “involuntary” separations is assumed to be independent of the wage paid on the job. The instantaneous rate of discount is ρ . Nonemployed searchers receive an instantaneous net benefit of b which is of unrestricted sign. The wage offer distribution is given by F ; the successive offers received by an individual are taken to be independently and identically distributed [i.i.d.] draws from F .

Consider an infinitesimally small period of time Δt . An agent who is currently employed at the instantaneous wage rate w may find himself in one of three states at the “end” of interval Δt . First, nothing may happen during the interval so that the individual remains employed at wage w . Second, the agent may be exogenously separated from his or her job and thus enter state n with associated value V_n . Finally, the agent may receive a new job offer with associated wage rate w' . Since the wage is the only characteristic of the job valued by the individual and there are no fixed costs associated with job-changing, the individual will change jobs if $w' > w$, in which case the end of “period” value will be $V_e(w')$; if $w' \leq w$ the agent will remain with his current employer and retain the match value $V_e(w)$. Thus we have the value of employment over the interval Δt is given by

$$V_e(w) = \frac{w\Delta t}{1 + \rho\Delta t} + \frac{1}{1 + \rho\Delta t} \{ \eta\Delta t V_n + (1 - \eta\Delta t - \lambda_e\Delta t) V_e(w) + \lambda_e\Delta t \int \max[V_e(w), V_e(x)] dF(x) \} + o(\Delta t), \quad [1]$$

where the term $(1 + \rho\Delta t)^{-1}$ is the “infinitesimal” discount factor associated with interval Δt , $\eta\Delta t$ is the approximate probability of being terminated from one's current job at the end of the interval Δt , $(1 + \eta\Delta t - \lambda_e\Delta t)$ is the approximate probability of not being terminated or

⁹ Alternatively we assume that agents are finitely-lived with unknown dates of death. For the stationarity of the model to be maintained it would be necessary to assume a constant risk of death. This mortality rate would solely alter the discount rate in the model explicated here.

receiving a new job offer, $\lambda_e \Delta t$ is the approximate probability of receiving a new job offer, V_n is the value of the nonemployment state, and $o(\Delta t)$ is a term which has the property that $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$. After collecting terms and taking the limit of [1] as $\Delta t \rightarrow 0$, we have

$$V_e(w) = \frac{w + \eta V_n + \lambda_e \int_w V_e(w') dF(w')}{\rho + \eta + \lambda_e \bar{F}(w)}. \quad [2]$$

The value of nonemployment is similarly derived. Over a small interval Δt

$$V_n = \frac{b \Delta t}{1 + \rho \Delta t} + \frac{1}{1 + \rho \Delta t} \{ \lambda_n \Delta t \int \max[V_n, V_e(w)] dF(w) + (1 - \lambda_n \Delta t) V_n \} + o(\Delta t) \quad [3]$$

where $\lambda_n \Delta t$ is the approximate probability of receiving one job offer over the interval Δt and $(1 - \lambda_n \Delta t)$ is the approximate probability of not receiving an offer; once again the period Δt is taken to be sufficiently small that the probability of two or more offers is of order $o(\Delta t)$. After rearranging terms and taking limits as $\Delta t \rightarrow 0$, we have

$$V_n = \frac{b + \lambda_n \int \max[V_n, V_e(w)] dF(w)}{\rho + \lambda_n}. \quad [4]$$

Given that an offer w is received by a nonemployed individual it will be accepted if and only if $V_e(w) \geq V_n$. Note that the value of search is a constant; given the easily demonstrated result that V_e is monotone increasing in its argument, it follows that there exists a unique reservation wage w^* such that $V_e(w^*) = V_n$. Then we can rewrite [4] as

$$V_n = \frac{b + \lambda_n \int_{w^*} V_e(w) dF(w)}{\rho + \lambda_n \bar{F}(w^*)}. \quad [5]$$

Note that w^* is solely a parameter which characterizes the nonemployed searches decision rule; it is a function of all the "primitive" parameters of the model: $\lambda_e, \lambda_n, \rho, \eta, b$, and F .

By way of summary, note that while the state valuation functions $V_e(w)$ and V_n are complicated functions of all the primitive parameters of the model, the decision rules themselves are very simply characterized as follows. If the agent is *nonemployed* and receives an offer w :

$$d_n^*(w) = \begin{cases} \text{accept offer } w & \Leftrightarrow w \geq w^* \\ \text{continue nonemployment} & \Leftrightarrow w < w^* \end{cases}. \quad [6]$$

If the agent is currently employed at some wage w [greater than w^* by definition] and receives an offer w' his rule is:

$$d_{e(w)}^*(w') = \begin{cases} \text{accept new job } w' & \Leftrightarrow w' > w \\ \text{continue current job } w & \Leftrightarrow w' \leq w \end{cases}. \quad [7]$$

These simple rules greatly facilitate estimation of the behavioral model, as we shall see below.

3.2. Valuation of the distribution of lifetime welfare

Much of the paper consists of comparing the CVs associated with various distributions of labor market outcomes. While we question the practice of drawing welfare inferences on the basis of cross-sectional remuneration distributions, one might reasonably ask if we are not committing a similar type of fallacy by focusing our attention on the dispersion present in the life-cycle welfare distribution. After all, risk-neutral agents of the kind assumed in our model should be indifferent with respect to variance in this distribution. Besides our own professional curiosity about dispersion in such distributions and political interest in such matters, we now consider some more substantive motivations for interest in these statistics.

To facilitate the discussion, it will be useful to restate the objective function of our risk-neutral agents in a slightly more transparent manner than was done [implicitly] within the previous subsection. Let $m(t)$ represent the monetarized utility flow associated with the labor market state occupied at moment t , where $m(t)$ equals the instantaneous wage rate if employed at t and equals b [a monetarized utility flow and/or rate of unemployment benefits] if nonemployed. This sample path (denoted by m) is generated by exogenous events which occur in an environment describe by some set of parameters Ω when the agent uses some strategy denoted by S . Let $Q(m|S, \Omega)$ denote the probability measure of sample paths of monetary values over the course of the career given environment Ω and strategy S . Then the expected monetary value of the career is given by

$$\begin{aligned} E[Y(\Omega, S)] &= \int \left[\int \exp[-\rho t] m(t) dt \right] dQ(m|\Omega, S) \\ &= \int Y(m) dQ(m|\Omega, S) \end{aligned}$$

where $Y(m)$ is the present monetary value of the career. The decision rules [6] and [7] which were derived above solve the problem

$$S^*(\Omega) = \arg \max_S E[Y(\Omega, S)].$$

The coefficient of variation for this distribution is $Var[Y(\Omega, S^*(\Omega))] / E[Y(\Omega, S^*(\Omega))]$.

Why be interested in $CV(\Omega, S^*(\Omega))$ if the agents don't care about $Var[Y(\Omega, S^*(\Omega))]$?

A simple motivation, and no doubt the most immediate, is that society members may dislike large amounts of welfare inequality. This concern may be generated solely from altruistic motives, or because of some negative consequences [to the individual] associated with large amounts of inequality, such as increased rates of crime and tax avoidance. In this case, the individual's preferences would be defined over $Y(m)$ and some function of the distribution of $Y(m)$, let us say the variance. While the individual controls [in a stochastic sense] his or her own $Y(m)$ process, the distribution of welfare values is a function of the strategies used by all the members of the environment, and so is not under the control of an individual acting in isolation. Without the possibility of strategic coordination, each agent will treat the variance in the society as predetermined and so will optimally choose to use the expected wealth-maximizing strategy him or herself.

Society members may address distributional issues indirectly by voting or through the installation of a social planner. While not altering individual decision rules, the social plan-

ner can affect dispersion in outcomes through the creation of an appropriate environment. We might think of the planner as facing the problem of choosing an institutional structure (Ω), given some cost function defined over the space of possible structures and given the reaction functions of labor market participants, with the objective of maximizing the expected value of welfare in the population, or

$$\Omega^* = \arg \max EY(\Omega, S^*(\Omega)) - \delta \text{Var}[Y(\Omega, S^*(\Omega))] - c(\Omega), \delta > 0.$$

In this case, risk neutral strategies are the relevant ones to examine, though dispersion in labor market outcomes is negatively related to the welfare of each agent. Agent's concerns about inequality are reflected in the design of the labor market environment, Ω , but not in the strategies they use conditional on this design.

A second possible motivation for concern with dispersion centers around the relaxation of the risk neutrality assumption. Consider a situation in which we generalize instantaneous utility to be of the form

$$U(x) = x - \beta x^2, \beta \geq 0. \quad [8]$$

where x denotes the rate of monetary compensation.¹⁰ When $\beta > 0$, the individual is risk averse.

In a world in which neither borrowing or saving is possible, the qualitative nature of the decision rules derived for the risk neutral case will be preserved, though the reservation wage associated with the nonemployment state will be reduced in value. The agent's maximization problem will be of the form:

$$\begin{aligned} S^* &= \arg \max_S \int \int \exp[-\rho t] [m(t) - \beta m(t)^2] dt dQ(m|\Omega, \beta, S) \\ &= \arg \max_S E[Y(\Omega, S)] - \beta \int \exp(-\rho t) m(t)^2 dt dQ(m|\Omega, \beta, S). \end{aligned} \quad [9]$$

From [9], we can see at least three problems with attempting to make welfare judgments on the basis of the dispersion observed in the monetary welfare distribution. First, this distribution is endogenous and is partially generated by the agent's choice of strategy. As such it is not possible to use as a basis for conducting comparative statics exercises. Secondly, variation in the value of labor market careers [i.e., variability in $Y(m)$] is not connected in any obvious way to the quantity $\int \exp(-\rho t) m(t)^2 dt dQ(m|\Omega, \beta, S)$. For example, careers could exhibit no variability in their total value while at the same time possessing significant moment-to-moment variation in the instantaneous value of labor market states. Third, the assumption of no capital markets is clearly counterfactual. Allowing for saving and borrowing would expand the strategy space of agents and make welfare comparisons much more difficult, but admittedly, much more interesting. Incorporating capital market institutions into such a comparative analysis is a task left to the future. For present purposes, we are forced simply to assume society members don't like inequality, but act in isolation like risk-neutral agents.

¹⁰ The specification in [8] might be thought of as either exact or as a first order approximation to a more general specification of instantaneous utility.

4. Econometric issues

We consider the formulation of a maximum likelihood estimator that allows us to retrieve all of the structural parameters of the model. These parameters are of interest in their own right as a basis for comparison of individual-level labor market dynamics in the two environments, and are perhaps even more useful when serving as a foundation for the generation of the life-cycle welfare simulations. Before turning to the mechanics of defining the estimator we must first consider two important preliminary issues - initial conditions, problems in the data and the modification of the model to account for job-to-job transitions associated with wage *decreases*. While these issues may appear strictly technical on the surface, they actually both raise a number of interesting substantive concerns and [hopefully] lead to some insights regarding the interpretation of the model and of cross-sectional remuneration distributions.

4.1. Initial conditions problems

The [limited] retrospective nature of the labor market information collected in the Italian data implies that the spell first sampled, that is the one on-going in January 1988, causes a nontrivial initial conditions problem when attempting to estimate a model of on-the-job search. If an individual was not employed at that time, no presample information was collected. If an individual was currently employed at the beginning of the sample period [January 1988], he was asked the date of the beginning of that particular job spell. The individual was not asked how many jobs he had in succession before the one occupied in January 1988.

The nature of the problem can be illustrated by referring to two hypothetical labor market histories, *A* and *B*¹¹. The individual with history *A* exited from nonemployment into employment a time a_1 . Under the assumptions of our model, he found a better job [in the sense of one paying a higher wage] at time a_2 . Thus the wage associated with his employment spell in progress at the beginning of the observation period is a random draw from the distribution $F_1(w)$, which is defined as the wage distribution of individuals whose previous labor market state was nonemployment [in our search model, $F_1(w) = F(w)/\tilde{F}(w^*)$, $w \geq w^*$, where F is the population wage offer distribution, F is the associated survivor function, and w^* is the reservation wage of nonemployed searchers]. While the data contains the individual's report of when the job spell in progress at the beginning of the observation period started [in this case a_1], the data do not contain information regarding his labor market state immediately prior to a_1 .

The individual with history *B* experienced a large number of job-to-job transitions prior to the onset of the observation period; in fact the individual is in his fourth [and last] job of the employment spell [which covers the period from b_1 through b_5] when the observation period begins. Unconditional on the wages associated with previous jobs in this employment spell, the wage associated with the sampled job spell will be a random draw from the distribution $F_4(w)$, which is defined as the wage distribution of individuals who have had exactly three consecutive jobs in the current employment spell prior to the current job. Assuming the wage offer distribution $F(w)$ is continuously differentiable, the distribution F_4 has an associated probability density function which is given by

$$f_4(w) = \int_{w^*}^w \int_{w_1}^w \int_{w_2}^w \frac{f(w)}{\tilde{F}(w_3)} \frac{f(w_3)}{\tilde{F}(w_2)} \frac{f(w_2)}{\tilde{F}(w_1)} \frac{f(w_1)}{\tilde{F}(w^*)} dw_3 dw_2 dw_1, w \geq w^*. \quad [10]$$

¹¹ Figure 2, displaying both histories, is not available for printing [note by the editor]

In Appendix 1 we derive an expression for the densities $f_i(w)$, $i = 2, 3, \dots$. These densities are only slightly special cases of standard order statistic density functions.

The form of the job-spell specific wage densities is particularly simple under the Pareto assumption used in the empirical work. In this case the densities are given by the closed-form expression:

$$f_j(w) = \begin{cases} f(w) & w \geq w^*, j = 1 \\ f(w) \frac{[\alpha(\ln(w) - \ln(w^*))]^{j-1}}{j-1} & w \geq w^*, j = 2, 3, \dots \end{cases} \quad [11]$$

Figure 3 contains plots of the wage p.d.f.s. associated with the first four job spells in any employment spell using structural parameter estimates for Lombardia and the U.S. Note the large differences in the densities across job spell, perhaps most notably with respect to the degree of symmetry exhibited. Given that these distributions are essentially those of order statistics, they can be ranked in terms of standard stochastic dominance criteria. The figure also serves to illustrate this point.

The transitions between job spells and the unemployment state leads to a steady-state distribution in terms of state occupancy - the probabilities of finding a randomly-sampled individual in nonemployment, in the first job of an employment spell, in the second job of an employment spell, etc. Let the conditional probability that an *employed* individual is in their i^{th} job in an employment spell be denoted by q_i . Then the steady state p.d.f. of wages is of the form

$$f^S(w) = \sum_{i=1}^{\infty} q_i f_i(w), \quad w \geq w^*, \quad [12]$$

$$\sum_{i=1}^{\infty} q_i = 1, \quad q_i \geq 0, \quad i = 1, 2, \dots$$

The probabilities q are derived in Appendix 1. It should be clear that they depend on all the structural parameters of the model since the transition rates between states are functions of the parameters characterizing exogenous labor market processes [such as the rates of receiving offers and dismissal] as well as the strategies of the labor market participants.

The steady state wage distribution has two potential uses with regards to this study. First, it is of substantive interest in that it provides a link between a stationary model of on-the-job search and the cross-sectional remuneration distribution that serves as a focus of much work on income inequality. In particular, it provides us with a good example of the interpretative problem to which Friedman was referring in the citation provided in the introduction. For consider increasing the rate of receiving offers while employed in an on-the-job search environment. This will increase the value of all states, n and $e(w)$. In terms of the weights q_i , the mass in this probability distribution will be shifted out to the higher-order spells. The result will be a shift in the steady state wage offer distribution that results in a higher mean wage and is consistent with increased variation¹². Thus increases in the CV of the cross-sectional distribution are consistent with welfare-improving changes in the search environment.

¹² We have not analytically established conditions under which increases in the CV must result from changes in the distribution of steady state spell orders [where the changes obey a stochastic dominance criteria]. However, we have never observed a decrease in the CV in any of the simulation exercises we have conducted.

Figure 3.a
Implied Wage Densities by Job Spell
Italy

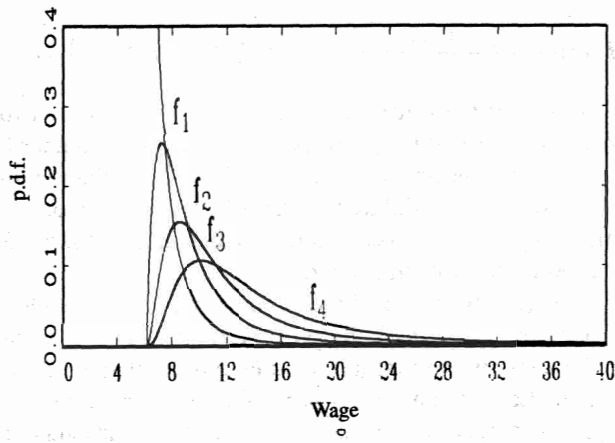
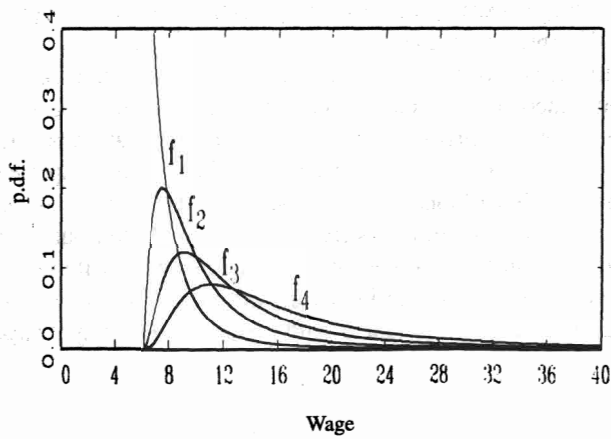


Figure 3.b
Implied Wage Densities by Job Spell
U.S.



A second potential use of the steady state wage distribution is as a solution to the initial conditions problem. If one is willing to assume that the data are generated by transitions consistent with the process when it has reached the steady state, the steady state density of wages can be used to describe the wage in any job in progress as of the beginning of the observation period. Unfortunately, this elegant solution is not very useful in this application since the cohorts we follow are very young during the sample period. Given the particularly low rate of movement between states observed in the Lombardian data, the time necessary to reach the steady state would reach well beyond the end of our observation period. Thus we have not used this distribution for estimation purposes in this study.

Our solution has been an agnostic one, which has the advantage of producing consistent estimators under a number of possible violations of modelling assumptions at the cost of discarding much sample information. It is to simply condition on the wage associated with any job in progress at the beginning of the sample period. While this method most certainly ignores a substantial amount of sample information, nevertheless we found most of the maximum likelihood estimates produced using it to be acceptably precise.

4.2. Accounting for job-to-job shifts involving wage losses

The strict model of wage-maximizing on-the-job search implies that any direct job-to-job transition should be associated with a wage gain. As we saw in Table 1, while this is in fact what is observed in the majority of sample cases, it is by no means true for all such movements. For the maximum likelihood estimator to be well-defined, some adjustment to model must be made to get rid of these probability-0 events.

There are essentially two ways to proceed. The most straightforward route, most always taken in the literature and the one followed here¹³, is to allow for measurement error of some simple type. The assumption made here is that:

$$\tilde{w} = w\varepsilon, \quad [13]$$

where w indicates the "true" wage rate associated with a spell, \tilde{w} is the observed wage, and ε is a random variable which is independently and identically log normally distributed, so that the density of ε is

$$m(\varepsilon) = \phi((\ln(\varepsilon) - \mu_\varepsilon)/\sigma_\varepsilon)/(\sigma_\varepsilon\varepsilon), \quad \varepsilon > 0, \quad [14]$$

where the parameter $\sigma_\varepsilon > 0$ and ϕ is the standard normal probability density function. We restrict the parameters μ_ε and σ_ε so that

$$\begin{aligned} E(\tilde{w}|w) &= w \\ \Rightarrow E(\varepsilon|w) &= 1 \quad \forall w \\ \Rightarrow \sigma_\varepsilon &= (-2\mu_\varepsilon)^{.5}, \end{aligned} \quad [15]$$

where the last line follows from the fact that the mean of ε is equal to $\exp(\mu_\varepsilon + .5\sigma_\varepsilon^2)$. This condition thus places an implicit restriction on μ_ε , namely, that it must be negative.

¹³ See, for example, the application of van den Berg and Ridder (1993) in an equilibrium on-the-job search context or Wolpin (1987) in a model of unemployed job search.

The advantage of this approach is that it is simple to implement, and the case for there being measurement error in survey reports on wages is compelling, to say the least¹⁴. The disadvantages are that, especially within a highly nonlinear estimation problem, functional form assumptions regarding the distributional properties of the measurement error are required and are ultimately arbitrary.

More ambitiously, one might consider modifying the structure of the behavioral model in nontrivial ways. For example, it is clearly the case that wages are not the only utility-determining aspect of jobs. If we allowed for a second utility-influencing attribute, ξ say, then we would be looking for job changes in which $U(w', \xi') > U(w, \xi)$, with U denoting the instantaneous utility yield associated with a job paying w and with other characteristic ξ . It will generally be the case that there will exist jobs which dominate the current one (w, ξ) in terms of utility yield yet pay a lower wage $w' < w$. This would apparently resolve our problem in a more elegant manner.

This response has several severe limitations. The first is the problem of observing, or for that matter even identifying, the characteristic ξ . If this characteristic is assumed to be unobservable, then to estimate the model we must assume a joint distribution for (w, ξ) . This assumption will be as arbitrary as any measurement error distributional assumption, so there would seem to be little to choose between the two approaches. On the other hand, if the characteristic is assumed to be observable, for any choice of U we are again likely to run into the problem that certain job-to-job transitions will be associated with a utility loss.

The main reason we have not pursued the option of introducing a characteristic ξ is more substantive. Without a clear conception of what a variable(s) represents, its presence would merely serve to cloud the comparative analysis we are attempting. Thus we have settled for the addition of measurement error as the most direct and uncontroversial way to make the model consistent with the data.

4.3. The maximum likelihood estimator

Under the recoverability condition defined by Flinn and Heckman (1982), which essentially requires that the analyst fix the lower support of the wage offer distribution when the wage offer distribution is assumed to belong to some given parametric family, all the parameters of the on-the-job search model are identified with the exception of the discount rate ρ and the cost of search b [which are not jointly identified]. If ρ is fixed, all remaining parameters are then identified.

For purposes of writing down the likelihood in a succinct fashion, it will make sense to break the data for each individual into components we will refer to as "cycles"¹⁵. A cycle is demarcated by the occurrence of a nonemployment spell. Cycles are defined over the observation period as follows. If an individual begins the sample period in the nonemployment state, he remains in that cycle until such time as he leaves the initial spell of nonemployment and enters a new one. Since every nonemployment spell ends in employment, the individual remains in the same cycle through all successive jobs he holds after the initial nonemployment spell [given that all job-to-job transitions are direct]. If an individual begins the observation period in a job, he remains in the original cycle until such time as he experiences a nonemployment spell. In theory an individual can experience an indefinitely

¹⁴ Duncan and Hill (1985) provide some disturbing evidence on the unreliability of self-reports of wages in the Panel Study of Income Dynamics. There is no reason to believe that self-reports of wages in the NLSY or the ISTAT supplemental labor force survey in Lombardia are any more reliable.

¹⁵ This follows Wolpin (1992).

large number of cycles over any finite-length sample period. We will let C_i denote the number of cycles experienced by sample member i over the sample period.

For reasons of computational tractability we will utilize duration and wage information only from the first two jobs in each cycle. This will not affect the consistency properties of the estimators, though throwing away information will lead to an efficiency loss. The likelihood function will incorporate information pertaining to whether a second job spell completed during the observation period is followed by another job spell or nonemployment.

In defining the likelihood function we will utilize the following notation [the individual subscript i has been dropped for notational simplicity]:

- $X_{n,c}$ equal to 1 if there is a nonemployment spell in cycle c
- $X_{1,c}$ equal to 1 if there is a first job in cycle c
- $X_{2,c}$ equal to 1 if there is a second job in cycle c
- $X_{3,c}$ equal to 1 if there is a third job in cycle c
- $t_{n,c}$ duration of nonemployment spell in cycle c
- $t_{1,c}$ duration of first job in cycle c
- $t_{2,c}$ duration of second job in cycle c
- $\tilde{w}_{1,c}$ observed wage in first job in cycle c
- $w_{2,c}$ observed wage in second job in cycle c
- $r_{n,c}$ equal to 1 if the nonemployment spell in cycle c is censored [16]
- $r_{1,c}$ equal to 1 if the first job spell in cycle c is censored
- $r_{2,c}$ equal to 1 if the second job spell in cycle c is censored

Whenever a variable is undefined in a particular cycle [for example, if there is no second job spell in the cycle then $t_{2,c}$, $w_{2,c}$, and $r_{2,c}$ are undefined] we set it equal to zero by convention. Note that the indicator variables for the presence of right-censoring need only be defined for the last cycle in the observation period [C_i for each i].

In the absence of initial conditions problems, the likelihood contribution for a given individual can be written as:

$$\begin{aligned}
 l = & \prod_{c=1}^C \int_{w^*} \int_{w_1} \left\{ h_n^{1-r_{n,c}} \exp(-h_n t_{n,c}) \right\}^{\lambda_{n,c}} \\
 & \times \left\{ \exp(-D(w_1)t_{1,c}) \left[(\lambda_e \tilde{F}(w_1))^{\chi_{2,c}} \eta^{1-\chi_{2,c}} \right]^{1-r_{1,c}} m(\tilde{w}_{1,c}/w_1)/w_1 \right\}^{\chi_{1,c}} \\
 & \times \left\{ \exp(-D(w_2)t_{2,c}) \left[(\lambda_e \tilde{F}(w_2))^{\chi_{3,c}} \eta^{1-\chi_{3,c}} \right]^{1-r_{2,c}} m(\tilde{w}_{2,c}/w_2)/w_2 \right\}^{\chi_{2,c}} \\
 & \times \frac{f(w_2)}{\tilde{F}(w_1)} \frac{f(w_1)}{\tilde{F}(w^*)} dw_2 dw_1,
 \end{aligned}
 \tag{17}$$

where $D(w) \equiv \eta + \lambda_e \tilde{F}(w)$ and $h_n \equiv \lambda_n F(w^*)$. Note that $m(\tilde{w}_{i,c}/w_j)/w_j$ is the density of the observed wage in the j th job in the cycle under the measurement error specification [13] and [14]; the term w_j^{-1} is the Jacobian of the transformation.

As was pointed out above, we do not have the information available which would allow us to use [17] to estimate model parameters because of the initial conditions problem. We have elected to treat the wage in any employment spell in progress at the beginning of our observation period as data. However, we don't observe the true wage in this case, but only have access to a noisy reading on it. Under our measurement error assumptions, we do know that the true wage is related to the observed wage as follows:

$$w = \tilde{w}/\varepsilon. \quad [19]$$

Then the density of the "true" wage in the sampled job spell is

$$\frac{m(\tilde{w}_s/w_s)\tilde{w}_s/w_s^2}{\Gamma(\tilde{w}_s)}, \quad w_s > w^*, \quad [20]$$

where the s subscript denotes sampled spell, the term \tilde{w}_s/w_s^2 is the Jacobian of the transformation, and $\Gamma(\tilde{w}_s)$ is a normalizing constant which ensures that the density integrates to unity.

Given the true wage in the sampled spell, the distribution of the wage in any immediately successive spell is $f(w')/\tilde{F}(w_s)$, $w' > w_s$. Conditional on the true wage associated with any sampled employment spell the duration distribution of the sampled employment spell from the sampling time until the completion of the spell $[t]$ is given by

$$D(w_s) \exp(-D(w_s)\tilde{t}), \quad [21]$$

which is identical to the population density of conditional [on the wage rate] job spell durations since the population distribution is exponential. If the sampled job spell is not completed by the end of the observation period, the probability of this event conditional on the wage is given by $\exp(-D(w_s)T)$, where T is the length of the observation period. For a sample *nonemployment* spell which is completed before T the likelihood contribution is $\exp(-h_n T)$. Due to the stationarity of the model, the densities and survivor functions associated with the forward recurrence times of the sampled spell are exactly the same as their population counterparts. Thus the only change in the likelihood function [17] which is required is the substitution of the sampled wage density [20] for the population density of the wage associated with the first job in the first cycle in the observation period when the first cycle begins with a job spell.

In terms of the parameterization of the likelihood function, we have directly estimated the set $\{\lambda_n, \lambda_e, w^*, \alpha(F), \eta, \mu\}$. Thus the reservation wage characterizing the decision rule of nonemployed individuals is treated as a parameter. As was done in Flinn and Heckman (1982), given estimates of this set of parameters and an assumption concerning the instantaneous discount factor ρ , the point estimate of b is found as follows¹⁶. For notational simplicity define $A(w) = \int_w V_e(\chi) dF(\chi)$ for $w \geq w^*$. Now we can write the value of search as

$$V_n = \frac{b + \lambda_n A(w^*)}{\rho + \lambda_n \tilde{F}(w^*)}, \quad [22]$$

or alternatively we can express it as

$$\begin{aligned} V_e(w^*) &= V_n \\ \Rightarrow V_n &= \frac{w^* + \eta V_n + \lambda_e A(w^*)}{\rho + \eta + \lambda_e \tilde{F}(w^*)} \\ \Rightarrow V_n &= \frac{w^* + \lambda_e A(w^*)}{\rho + \lambda_e \tilde{F}(w^*)}. \end{aligned} \quad [23]$$

¹⁶ It should be noted that the model estimated in Flinn and Heckman made no allowance for on-the-job search [that is, $\lambda_e = 0$], so that the explicit expression for b was substantially different.

Note that in [23], given the reservation wage the value of nonemployment is not a function of b . We can substitute [23] into [2] to get

$$V_e(w) = \frac{w + \eta \left[\frac{w^* + \lambda_e A(w^*)}{\rho + \lambda_e \tilde{F}(w^*)} \right] + \lambda_e A(w)}{\rho + \eta + \lambda_e \tilde{F}(w)}. \quad [24]$$

Call the right hand side [RHS] of [24] $S(V_e, w^*)$. Then we can solve the integral equation $V_e = S(V_e, w^*)$ using m.l. point estimates of all required structural parameters, the m.l. estimate of w^* , and the assumed value of ρ ; call the solution of the integral equation \hat{V}_e . Finally, we can solve for b using

$$b = \left[\frac{\rho + \lambda_n \tilde{F}(w^*)}{\rho + \lambda_e \tilde{F}(w^*)} \right] [w^* + \lambda_e A(w^*)] - \lambda_n A(w^*). \quad [25]$$

The smoothness of the RHS (25) insures that the invariance property of maximum likelihood estimators holds. Thus the estimate of b obtained by substituting m.l. point estimates into [25], including the estimated function \hat{V}_e , is consistent and has a \sqrt{N} asymptotic normal distribution.

5. Structural parameter estimates and simulation exercises

In all of the empirical work conducted we have assumed that the offer distribution is Pareto. As was shown in Flinn and Heckman (1982), with access only to data containing accepted wage offers and the duration of time spent in labor market states, parametric assumptions are required to estimate a structural search model. Though the Pareto is not "recoverable" in the sense of Flinn and Heckman, we are able to use it after making the additional assumption that the lower bound of the support of the distribution is equal to the reservation wage. The reader should bear in mind that inferences drawn from the model are likely to be sensitive to assumptions regarding the wage offer distribution.

Table 2 contains estimates of the search model for the ISTAT and NLSY samples. The estimates of the rates of receiving offers in the nonemployed and employed states and the rates of dismissal provide a good summary of the purely descriptive evidence presented in Table 1. The rates of receiving offers when unemployed are about three times greater in the NLSY sample, and the rate of receiving job offers when in the employed states is more than an order of magnitude greater in the U.S. than in Lombardia. On the other hand, the rate of dismissals is over an order of magnitude larger in the U.S. These estimates are very much in accord with conventional wisdom regarding the functioning of the markets in the two countries. The Italian market is characterized as one in which individuals spend long periods of time searching for their first job, but once they find it, tend to keep it for a long time, often their entire working life. The U.S. labor market is thought to be one with rapid rates of turnover between jobs, both for "voluntary" and "involuntary" reasons. All of these gross generalizations are consistent with the evidence presented in Table 2.

Table 2 – Search Model Estimates

PARAMETER	Italy	U.S.
λ_n	.073 (.011)	.237 (.010)
λ_e	.006 (.002)	.089 (.006)
w^*	5.874 (.246)	6.190 (.222)
$\alpha(F)$	4.884 (.956)	4.169 (.369)
η	.0016 (.0005)	.019 (.001)
μ_e	-.102 (.004)	-.108 (.007)
b	-3.895	2.573
L	-424.890	8433.196
$E(w)$	7.386	8.143
$SD(w)$	1.533	2.002

Monetary quantities are not directly comparable across the two samples, but partly due to our choice of units the wage offer distributions in the two environments appear to be relatively similar [the units chosen were 1000s of lire per hour for the ISTAT sample and dollars per hour for the NLSY]. The reservation wage, which is the lower bound of the wage offer distribution, is similar in the two samples, as is the shape of the Pareto. The estimated mean and standard deviation of the wage offer distribution are presented at the bottom of the table. The mean and standard deviation are similar across the two samples, though both are a bit larger for the U.S.

We now turn to a consideration of the implications of these estimates for dispersion in lifetime welfare outcomes. The Italian labor market has been characterized as one in which welfare outcomes are “compressed” relative to what is experienced by participants in more competitive markets like that of the U.S. This view is commonly supported by appealing to cross-sectional evidence on wages or earnings like that presented in Figures 1a. and 1.b. As we hope has been made clear, such evidence is very difficult to interpret in a life-cycle setting. Our simulation exercise utilizes point estimates of the search models to compute welfare levels associated with a large number of pseudo-labor market careers.

The simulation exercise itself is simple [it is described in detail in Appendix 2]. The procedure begins by fixing values for all the structural parameters of the model [the m.l. point estimates are used for this purpose]. Each “individual” enters the labor market in the nonemployment state - then a random number is generated which determines the duration of time he spends in the state until finding a job. The wage associated with the first job is determined by another random number draw and the population distribution of [first job] accepted wage offers. Other random numbers are drawn to determine when new offers are received, their values, and the times of exogenous terminations. The process is repeated until the labor market career has lasted at least 45 years [540 months]. For each labor market environment, 100,000 sample histories were created. The coeffi-

coefficients of variation reported for the lifetime welfare distribution and the steady state wage distribution are computed from these sample distributions¹⁷.

Table 3 contains estimates of the coefficient of variation for a number of distributions we have singled out for attention throughout the analysis. The first line reproduces information displayed in Figure 1. We noted there the CV of wages for people employed in the first week of January 1988 was about twice as large in the NLSY sample than in the ISTAT sample.

Table 3 – Coefficients of Variation for Wage and Welfare Distributions

REMUNERATION MEASURE	Italy	U.S.
Observed Wage (Jan 88)	.327	.621
Wage Offer	.208	.246
Lifetime Welfare	.284	.181
Steady State Wage	.343	.416

The second line of the table contains the coefficient of variation associated with the wage offer distributions in the two environments. This CV is computed by using the point estimates of the Pareto distribution parameters from Table 2. Here the CVs associated with the two environments are much more similar, with the CV for the U.S. distribution about 25% larger than the CV for the Lombardian distribution.

The single most striking result we have to report regards the comparison of the CVs of the lifetime welfare distributions. The CV for Lombardia is over 50% greater than the CV for the U.S. distribution. We can provide a heuristic explanation for this finding, which is hopefully intuitive. Within this model, the probability of an employed individual leaving his current job for a better one is $\lambda_e \tilde{G}(w) / (\lambda_e \tilde{G}(w) + \eta)$. Since this probability is decreasing in w , individuals with low wages are relatively more likely to leave their jobs for higher paying ones than are individuals currently receiving high wages. Individuals at the high end of the wage distribution are effectively limited in the enjoyment of their good fortune by the dismissal process. When the dismissal rate is high, as in the U.S., this tends to reduce the probability mass in the right tail of the lifetime wealth distribution by limiting the length of time jobs paying high wages are kept. Conversely, individuals with low current wage draws are relatively likely to move up the wage distribution. For example, individuals working at the lowest wage job in each market [i.e., one paying the reservation wage] have a probability of ending the job by finding a better one of .824 in the U.S. and .789 in Lombardia. This type of selection mechanism tends to push low wage earners up the wage distribution in each country, thus reducing lifetime inequality. The big difference in the two markets is the speed with which the mechanism operates. Consider again individuals earning the reservation wage. In the U.S., such an individual has a probability of .726 of leaving this job within one year, while in Lombardia the probability of leaving such a job within a year is only .087. Thus the speed with which people are pushed up the wage distribution from low wage jobs or are simply forced to take new draws [which according to the law of large numbers would tend to concentrate the lifetime welfare distribution in and of itself] seems to account for the smaller amount of dispersion in the U.S. distribution.

The bottom line of the table gives the CV associated with the steady state wage distribution. As noted previously, the CV for the U.S. is noticeably larger. The principal reason is

¹⁷ The CV for the steady state wage distribution is computed as follows. We let the simulation process run for 10000 months for each labor market participant, and we compute the CV from the distribution of wages in the 10000th month, which should be a more than adequate approximation to the steady state distribution.

that in the steady state, more individuals are in higher-order job spells in the U.S. [recall that the wage offer distributions themselves were quite similar]. Since being in a higher-order job spell is better than being in a lower-order one in a stochastic sense, substantial variation in the cross-sectional distribution is a "good" thing, even when members of the society value low variation in lifetime welfare outcomes.

Some of these results are quite striking, and it is natural to question whether the model which generates them has any claim to be taken seriously. We can use our simulated labor market histories to partially address this question. In particular, we look at the manner in which the mean and standard deviation of wages change over the life cycle. Even though the model is highly stylized, and neglects such admittedly important phenomena as within job spell wage change and business cycle effects on the parameters characterizing the labor market environment, Figure 4 provides some evidence that the model predictions are broadly consistent with evidence obtained using panel data to estimate flexible models of earning dynamics.

Life-cycle phenomena are produced in a stationary environment as the process moves from initial conditions to the steady state. In our case, all labor market participants are assumed to share the same initial condition, which is nonemployment. The length of time it takes for the distribution of state occupancies to be well-approximated by the steady state distribution depends on the transition rates which characterize the process. Moreover, the distributions of state occupancies as indexed by labor market age of the cohort will move in a systematic [e.g., monotonic] fashion since all individuals begin their careers unemployed. We expect to see the mean wage of the employed increase as a function of labor market wage; predictions regarding trends in dispersion seem to be dependent on the functional form of the wage offer distribution.

In Figure 4.a, we see for Lombardia the model predicts that unemployment will be heavily concentrated in the early years of labor market participation, which is consistent with substantial amounts of empirical evidence for Italy. After the first few years in the labor market, however, the Italian employment rate is predicted to be a few points higher than that of the U.S., which is graphed in Figure 4.b. Because the rate of receiving offers is so much greater in the U.S., and because all offers are accepted under our modelling assumptions, the steady-state employment rate is achieved within a few years in the U.S., in comparison with 5 or 6 for Italy. There is no pronounced concentration of unemployment in the early years of labor market participation for the U.S.¹⁸

Figures 4.c and 4.d contain plots of mean wage *growth* for the U.S. and Lombardian samples. We see that growth is always positive in the two environments, though the rate of growth is dramatically greater in the U.S. over the first 8 to 10 years of labor market participation. The growth rate in Italy is higher in the later years of life, simply because the U.S. distribution has already converged to its steady state value [in which there is no growth due to our assumption of a time-invariant wage offer distribution], while the Italian distribution will not converge to the steady state distribution until long after a normal working lifetime is completed. Of course, this life-cycle growth in wages is built into our model, so as such it is not a surprising finding. What is interesting is the predicted relative growth rates in mean earnings for the two countries.

Figures 4.e and 4.f complete our brief look at life-cycle movements with plots of the standard deviation of the wage distribution as a function of labor market age. We see that in both countries wage inequality grows monotonically as the population ages¹⁹. This observation is consistent with findings from many panel data studies of earnings dynamics. The

¹⁸ Recall that these data pertain to white males only. We would expect to see much longer spells of unemployment and more concentration of unemployment among youth if the sample were extended to include other racial groups.

¹⁹ The plot for the standard deviation displays more "wiggles" because higher order moments are estimated less precisely than lower order moments. The negative values of the growth rate plotted are solely due to sampling variability, which in this case could be reduced by producing larger numbers of simulated histories.

Figure 4.a
Employment Rates
Italy



Figure 4.b
Employment Rates
U.S.

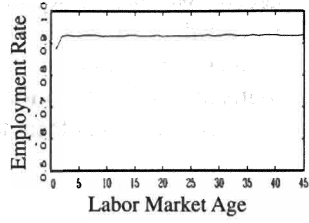


Figure 4.c
Mean Wage Growth
Italy

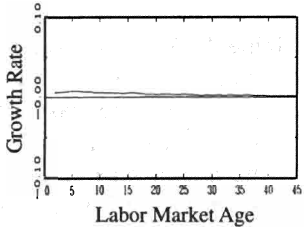


Figure 4.d
Mean Wage Growth
U.S.



Figure 4.e
Growth in S.D. of Wages
Italy

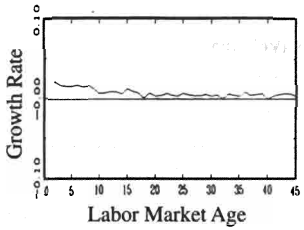


Figure 4.f
Growth in S.D. of Wages
U.S.



rate of convergence to the steady state value of the standard deviation is similar to what was observed in the case of the mean values.

We conclude our discussion of model implications by looking at the effects of changes in the some of [the more comparable] structural parameters of the model on both the mean and the CV of the lifetime welfare distribution. Since we have not estimated an equilibrium model, this exercise can not be construed as representing any type of policy evaluation, but is merely intended to indicate the sensitivity of distributional parameters of interest to changes in the structure of the environment, everything else equal.

We conducted the exercise by increasing each of the parameters considered [one at a time] by 10 percent. We then re-solved the model for the new reservation wage²⁰. Using this rule, we generated a new collection of labor market careers. From this new collection, we compute summary statistics of the life-time welfare distribution, and compared the new value of each statistic to its value prior to the perturbation of the parameters. The [numerical] estimate of the elasticity was then determined by

$$\nu_{X,i} = \frac{\left(\frac{X(\theta_{-i}) - X(\theta)}{X(\theta)} \right)}{.1},$$

where X is either the mean or the coefficient of variation of the welfare distribution and θ denotes the parameter vector which characterizes the search model; $X(\theta)$ denotes that the sample statistic is computed using the distribution generated from the original model estimates and $X(\theta_{-i})$ denotes that the sample statistic is computed using the distribution generated after increasing the i^{th} element in θ by 10 percent, keeping the value of all other elements fixed.

The results of this exercise are summarized in Table 4. We first discuss the mean welfare elasticities. As expected, increases in the rates of receiving offers and in the utility flow of nonemployment increase mean welfare, but nowhere is the elasticity large. A small change in the rate of receiving offers in the nonemployed state has a larger effect on mean welfare in Italy than does a small change in the rate of receiving offers while employed; this ordering is reversed in the U.S. The difference probably results from the very different distributions of unemployment duration in the two countries. Reducing the lengthy time spent waiting for the first job in Lombardia has a bigger impact on welfare than does reducing the relatively short waiting-time for employment in the U.S. Similarly, increases in the dismissal rate have a relatively large [negative] effect on welfare in the U.S., because the dismissal rate is already high there. Conversely, an increased rate of dismissal has a smaller effect in Italy because the rate there is so low to begin with.

Table 4 – Elasticities of the Mean and the CV of Lifetime Welfare

PARAMETER	MEAN WELFARE		WELFARE CV	
	Italy	U.S.	Italy	U.S.
λ_n	.165	.074	.042	-.121
λ_e	.070	.134	-.063	.270
η	-.117	-.215	-.281	-.772
b	.027	.015	-.334	-.309

²⁰ Increasing some parameters, such as the dismissal rate, decreases the value of search and hence the reservation wage. In our framework, this has no effect on the probability of accepting an offer given receipt of one, since by construction the lower bound of the support of the distribution of wage offers was equal to the original reservation wage. When the value of search is an increasing function of a parameter, the reservation wage will increase and the probability of acceptance conditional on receipt of an offer will decrease.

Interpreting the effects of parameter shifts on the CV is much more difficult, since changes in this statistic reflect both movements in the mean and standard deviation of the welfare distribution. For the most part, the absolute size of the elasticities is again small, with a few possible exceptions. Small increases in the rates of arrival of offers have almost no effect on the CV for Italy. The same is not as true for the U.S., where we find that a small increase in the rate of arrival of offers when unemployed results in less dispersed welfare distribution [in the sense of the CV]. Conversely, increases in λ_e tend to increase inequality given the U.S. parameter values.

The dismissal rate parameter has a relatively large contribution to make in the sense of reducing lifetime inequality. It is not difficult to understand why. As we increase the dismissal rate, we increase the number of jobs which end in dismissal. In the limit, when all jobs end due to dismissal, a career of jobs is just a sequence of i.i.d. draws from the distribution F . If the rate of discounting is very low, all careers will have virtually identical values. In the U.S., which starts from a high dismissal rate, an increase gets us even closer to the range in which a law of large numbers phenomenon is strongly exhibited. We are less close to this range when using the estimates from Lombardia.

The elasticity associated with the utility flow from nonemployment is interesting. This elasticity is very similar for the U.S. and Lombardia both for the case of the mean and the CV. For neither market does an increase in b have much of an effect on mean welfare, but in both it has a nonnegligible negative effect on the dispersion in the distribution. This suggests that the lower tail of the lifetime welfare distribution may be primarily composed of individuals who have had unusually long unemployment spells. Increasing compensation in this state reduces the mass in the left tail area in a manner which has no large effect on mean welfare, but does have a substantial effect on the standard deviation of the distribution.

6. Conclusion

We provide evidence that the cautionary tale told by Friedman is well-worth heeding when comparing levels of inequality in very different institutional environments. Environments which produce more "equitable" distributions of remuneration in the cross-section may be associated with more "inequitable" distributions of lifetime welfare. We showed that this was exactly the case with respect to a comparison between a very large region in Italy and the U.S., using samples composed of young male labor market participants.

The analysis is based on a standard model of off-and on-the-job search. The search framework serves three purposes. First, it gives us the ability to use the data at our disposal, which is based on a very brief observation period of 17 months, to estimate parameters which can generate outcomes over the entire labor market career of participants. Second, it provides a metric, however limited, for making welfare comparisons. Third, it provides a way to associate characteristics of the labor market with the distribution of welfare outcomes. We would argue that even when data on remuneration is available for a substantial fraction of each sample member's labor market career, the use of a model [not necessarily one of on-the-job search] greatly facilitates the conversion of a remuneration history into a summary welfare measure and provides a readily interpretable way to map characteristics of the environment into moments of the welfare distribution.

While the evidence we have presented might lead us to label Italy [or its stand-in here, Lombardia] as the status society, there are obviously a number of qualifications to make. As we have remarked previously, the model is very stylized, and is particularly suspect due to its omissions of business cycle variability in the search environment and on-the-job wage vari-

ability. Superior data could be used to rectify these problems at the cost of considerable increases in model complexity. It is not clear how the neglect of these features in both environments would tend to bias our comparisons of welfare inequality, but clearly our inferences could easily change.

A more direct measurement issue concerns the reporting of labor market activity in the two environments. It is well-known that there is a significant “underground” economy in the Italy. Because this unofficial labor market is illegal, little is known about its exact size or nature, though we can assume that there is nothing of analogous importance in the U.S. [in particular for the demographic group of which our sample is composed]. We can not be certain as to whether respondents to a government-financed survey would be reluctant to report holding an “illegal” job, though it is reasonable to suppose that they might. If so, non reporting of such jobs might seriously affect our estimates of transition rates between labor market states in Lombardia, with resulting repercussions for our comparisons of welfare inequality. There seem little of a constructive nature to do about this problem beyond being aware of its existence.

Finally, we believe a complete analysis of the issues addressed here requires a careful consideration of differences in capital markets. While capital markets in Lombardia and the U.S. seem to be relatively similar now, there still exist substantial differences in requirements for loans and the types of savings instruments available. Given any realization of a lifetime stream of earnings, different capital market environments can produce grossly different lifetime welfare outcomes. We hope our analysis can be serve as a first step in this direction.

Appendix 1. Derivation of the Steady-State Wage Distribution

In this section we explicitly derive the mapping between the dynamic behavioral model and the steady state cross-sectional earnings distribution. Before proceeding to the details of the derivation, we present a heuristic explanation of this mapping.

When individuals are sampled at some random point in time, agents who happen to be employed at the sampling time can be distinguished by the number of the job they are currently holding in the on-going employment spell. For example, in Figure 2 individual *A* was found in his first job in the employment spell while individual *B* was found to be in his fourth. The individual who has had three previous jobs in the current spell will on average have a higher wage than individual *A* since he has the maximum of four draws from the acceptable wage offer distribution while individual *A* has had only one draw.

To compute the steady state wage distribution requires us to first find the wage distribution by job order, and then to find the steady state distribution of job orders in on-going employment spells. The first task is relatively straight-forward, being essentially an exercise involving order statistics. To compute the steady state job order distribution, we first compute the steady state distribution of the number of jobs in an employment spell. Conditional on there being k jobs in a completed employment spell, we then derive the probability that an individual will be sampled while in his j^{th} job, $j = 1, \dots, k$. Then the probability that an individual will be found in his first job in an employment spell will be the steady state probability of having a one-job employment spell plus the conditional probability of being in the first job when sampled randomly given the employment spell ends after two jobs multiplied by the probability that the employment spell ends after two jobs, etc. We now proceed to a formal derivation of these distributions.

By the structure of the model all nonemployed individuals accept any draw from the

distribution F which exceeds the reservation wage associated with the nonemployment state, w^* . Then the density of wages in the first job in an employment spell is

$$f_1(w_1) = \frac{f(w_1)}{\tilde{F}(w^*)} \chi[w_1 \geq w^*], \quad [26]$$

where $\chi[A]$ takes the value 1 if A is true and equals 0 if A is false.

To derive the marginal density of the wages associated with the second job in an employment spell, we first write the joint density of first and second job wages,

$$f_{1,2}(w_1, w_2) = f_{2|1}(w_2|w_1)f_1(w_1) \quad [27]$$

$$= \frac{f(w_2)}{\tilde{F}(w_1)} \chi[w_2 > w_1] \frac{f(w_1)}{\tilde{F}(w^*)} \chi[w_1 > w^*] \quad [28]$$

$$= \frac{f(w_2)}{\tilde{F}(w^*)} \frac{f(w_1)}{\tilde{F}(w_1)} \chi[w_2 > w_1 > w^*], \quad [29]$$

where the form of the conditional density $f_{2|1}$ follows from the fact that the second job is a random draw from f truncated from below at the first job wage w_1 . Integrating over w_1 to get the marginal density of w_2 we have

$$f_2(w_2) = \frac{f(w_2)}{\tilde{F}(w^*)} \left[\int_{w^*}^{w_2} h(w_1) dw_1 \right] \chi[w_2 > w^*], \quad [30]$$

$$= \frac{f(w_2)}{\tilde{F}(w^*)} AF_2(w_2) \chi[w_2 > w^*], \quad [31]$$

where $AF_2(w_2)$ [for "adjustment factor for the second job wage density evaluated at w_2 "] is the term in brackets in [30] and $h(x)$ is the hazard function associated with the distribution F evaluated at x , or $h(x) \equiv f(x)/\tilde{F}(x)$.

By induction we can construct the marginal density of wages associated with the k^{th} job in an employment spell as

$$f_k(w_k) = \frac{f(w_k)}{\tilde{F}(w^*)} AF_k(w_k) \chi[w_k > w^*], \quad [32]$$

where the adjustment factor $AF_k(w_k)$ is defined recursively by

$$AF_k(w_k) = \int_{w^*}^{w_k} AF_{k-1}(x) h(x) dx. \quad [33]$$

To derive the weights q_1, q_2, \dots referred to in the text, we proceed as follows. Consider the *conditional* [on employment] steady state distribution of spell orders, Q , where $Q(k)$ gives the probability that a randomly sampled employed individual is in his k^{th} job spell. We derive this distribution in two steps. First we determine the probability that a randomly-selected employment spell ends after exactly k jobs, and we denote this proba-

bility distribution by R . An individual in his first job spell who is employed at wage w_1 exits into another job at rate $\lambda_e \tilde{F}(w_1)$ and exits into nonemployment [i.e., is "involuntarily" separated from this job] at rate η . Then conditional on his wage, the likelihood that employment spell ends after first job is the likelihood that the first job ends due to a dismissal, which is $\eta/(\eta + \lambda_e \tilde{F}(w_1))$. Then the unconditional probability of an employment spell ending after one job is

$$R(1) = \int_{w^*} \frac{\eta}{\eta + \lambda_e \tilde{F}(w_1)} f_1(w_1) dw_1. \quad [34]$$

Conditional on not having exited an employment spell before the k^{th} job, we have by extension that the probability that the employment spell ends after the k^{th} job [which we will denote $\hat{R}(k)$] is

$$\hat{R}(k) = \int_{w^*} \frac{\eta}{\eta + \lambda_e \tilde{F}(w_k)} f_k(w_k) dw_k, \quad k = 2, 3, \dots \quad [35]$$

Then the unconditional distribution of numbers of jobs in completed employment spells is

$$R(k) = \prod_{j=1}^{k-1} (1 - \hat{R}(j)) \hat{R}(k), \quad k = 2, 3, \dots, \quad [36]$$

with $R(1)$ defined in [34] and $\hat{R}(1) \equiv R(1)$.

The second step in the process involves the computation of the conditional probabilities of being found in job J when sampled given that the employment spell contains $k \geq J$ jobs. Denote this generic conditional probability by π_{jk} , and define the matrix P [which is a square upper triangular matrix of countably infinite dimension] by

$$P = \begin{bmatrix} 1 & \pi_{12} & \pi_{13} & \cdots \\ 0 & \pi_{22} & \pi_{23} & \cdots \\ 0 & 0 & \pi_{33} & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{bmatrix}. \quad [37]$$

The fact that $\pi_{11} = 1$ is definitional.

Consider the determination of the elements in the second column of P . Now the total time spent in employment spells which end after two jobs [denoted $t^{(2)}$] is the sum of the first job spell duration conditional on the first job ending in a quit and the second job spell duration conditional on the second job ending in a dismissal. Conditional on the wages associated with these two jobs, the spell durations are independently distributed. The conditional expectation of the duration of a two-job employment spell is then

$$Et^{(2)}(w_1, w_2) = \int_0^\infty \int_0^\infty (t_1 + t_2) \lambda_e \tilde{F}(w_1) \exp(-D(w_1)t_1) \times \eta \exp(-D(w_2)t_2) dt_1 dt_2 \quad [38]$$

$$= \int_0^\infty t_1 \lambda_e \tilde{F}(w_1) \exp(-D(w_1)t_1) dt_1 \quad [39]$$

$$+ \int_0^\infty t_2 \eta \exp(-D(w_2)t_2) dt_2 \quad [40]$$

$$= \lambda_e \tilde{F}(w_1) D(w_1)^{-2} + \eta D(w_2)^{-2},$$

where $D(w)$ is defined as $\lambda_e \tilde{F}(w) + \eta$. Then the unconditional expectation of the duration of a two-job spell is

$$Et^{(2)} = \int_{w^*} \int_{w^*} [\lambda_e \tilde{F}(w_1) D(w_1)^{-2} + \eta D(w_2)^{-2}] f_{1,2}(w_1, w_2) dw_1 dw_2 \quad [41]$$

$$= \int_{w^*} \lambda_e \tilde{F}(w_1) D(w_1)^{-2} f_1(w_1) dw_1 \quad [42]$$

$$+ \int_{w^*} \eta D(w_2)^{-2} f_2(w_2) dw_2. \quad [43]$$

To simplify notation, define

$$\mathcal{E}_k \equiv \int_{w^*} \lambda_e \tilde{F}(w_k) D(w_k)^{-2} f_k(w_k) dw_k \quad [44]$$

$$\mathcal{W}_k \equiv \int_{w^*} \eta D(w_k)^{-2} f_k(w_k) dw_k. \quad [45]$$

Using standard ergodic arguments, the probability of finding an individual in the first job of a two job spell under random sampling is

$$\pi_{12} = \frac{\mathcal{E}_1}{\mathcal{E}_1 + \mathcal{W}_2} \quad [46]$$

and of course $\pi_{22} = 1 - \pi_{12}$. In general, the expected duration of an employment spell which ends after the k jobs is

$$Et^{(k)} = \mathcal{E}_1 + \mathcal{E}_2 + \dots + \mathcal{E}_{k-1} + \mathcal{W}_k \quad [47]$$

$$= T_{k-1} + \mathcal{W}_k, \quad [48]$$

where $T_{k-1} = \sum_{i=1}^{k-1} \mathcal{E}_i$. Since the probability of finding an individual in job spell j of an employment spell which contains a total of k jobs is equal to the average duration of job

spell j in a k job employment spell divided by the expected duration of a k job employment spell, we have

$$\pi_{jk} = \begin{cases} \frac{\varepsilon_j}{T_{k-1} + W_k} & \Leftrightarrow j = 1, \dots, k-1; k = 2, 3, \dots \\ \frac{W_k}{T_{k-1} + W_k} & \Leftrightarrow j = k; k = 1, 2, \dots \\ 0 & \Leftrightarrow j > k; k = 1, 2, \dots \end{cases} \quad [49]$$

Now we can define the probability distribution for finding an individual in the j^{th} job of an employment spell given that they are employed at the random sampling time as

$$Q = PR. \quad [50]$$

Appendix 2. Simulation methods

To compute the welfare values $\omega(1), \dots, \omega(N)$ corresponding to a given set of values for the structural parameters, the procedure used was as follows. All "hypothetical" careers begin in the unemployment state. The discount rate was set to $\rho/12$ [recall that this is a monthly rate], where we set $\rho = .04$. The generic spell is indexed by i , and it begins at time $t_i [t_1 = 0]$. The labor market career is terminated at the conclusion of the first spell with a termination date after the 540th of the career, which corresponds to the individual spending approximately 45 years in the labor market. The instantaneous remuneration rate attached to spell i is denoted r_i , and the type of spell is denoted d_i , where $d_i = 1$ if spell i is an employment spell and is equal to 0 if it is an unemployment spell, and the total duration of the spell is denoted t_i .

If spell i was an unemployment spell [so that $d_i = 0$], we first generated a draw t_i from an exponential distribution with parameter $\lambda_n \tilde{F}(w^*)$. We then generated a wage draw w_{i+1} [since this wage is associated with the $i + 1^{\text{st}}$ spell] from the *accepted* wage offer distribution $F(w|w \geq w^*)^{21}$. The contribution of spell i to lifetime welfare is given by

$$V_i = \exp(-\rho t_i) \int_0^{t_i} b \exp(-\rho u) du \quad [51]$$

$$= \rho^{-1} \exp(-\rho t_i) [1 - \exp(-\rho t_i)] b \quad [52]$$

The next [employment] spell would then begin at calendar time $\tau_{i+1} = \tau_i + t_i$ at the wage w_{i+1} . If spell i is an employment spell the procedure is essentially the same except for the fact that there are two ways to exit the spell, either through an exogenous separation or a quit into a higher-paying job. Let the wage associated with the job be given by w_i . We first took a draw from an exponential distribution with parameter $\eta + \lambda_n \tilde{F}(w_i)$, which gave us the duration of the employment spell, t_i . We then generated a draw x from a uniform distribution on the unit interval. If $x < \eta / (\eta + \lambda_n \tilde{F}(w_i))$ the spell was considered to have ended in a dismissal

²¹ Under our assumption that the lower bound of the support of the Pareto is equal to the reservation wage, this distribution is the same as F .

so that spell $i+1$ is a nonemployment spell; conversely, if $x \geq \eta / (\eta + \lambda e^{\tilde{F}(w_i)})$ the next spell was an employment spell. When spell $i+1$ was an employment spell, a wage rate was generated from the distribution $F(w|w > w_i)$. Given spell i is an employment spell, its contribution to lifetime welfare is

$$V_i = \rho^{-1} \exp(-\rho\tau_i) [1 - \exp(-\rho t_i)] w_i. \quad [53]$$

Let N denote the number of spells which commenced prior to the 540th month. Say that this history has been generated for "individual" J . Then individual J 's labor market career has value

$$\omega(j) = \sum_{i=1}^N V_i \quad [54]$$

under the particular set of structural parameter values utilized in the experiment.

References

- AABERGE, R., A. BJÖRKLUND, M. JÄNTTI, M. PALME, P. PEDERSEN, N. SMITH, AND T. WENNEMO, "Income Inequality and Income Mobility in the Scandinavian Countries Compared to the United States", Discussion Paper 168, Statistics Norway, March 1996.
- ATKINSON, A., F. BOURGUIGNON, AND C. MORRISON, *Empirical Studies of Earnings Mobility*, Harwood Publishers, 1992.
- ATKINSON, A., SMEEDING, T., AND L. RAINWATER, "Income Distribution in the OECD Countries: The Evidence from the Luxembourg Income Study", *Social Policy Studies*, Volume 18, Paris, OECD, 1995.
- BERTOLA, G. AND A. ICHINO, "Wage Inequality and Unemployment: U.S. vs. Europe" *NBER Macroeconomic Annual 1995*.
- BERTOLA, G. AND A. ICHINO, "Institutions and Labor Reallocation", Forthcoming, *European Economic Review* (1997).
- BJÖRKLUND, A., "A Comparison Between Actual Distributions of Annual and Lifetime Income: Sweden 1951-1989", *Review of Income and Wealth* 39 (1993), 377-386.
- BLANCHFLOWER, D., L. KATZ, AND G. LOVEMAN, "A Comparison of Changes in the Structure of Wages in Four OECD Countries", N.B.E.R. W.P. 4297, 1993.
- BLOMQUIST, S., "A Comparison of Distributions of Annual and Lifetime Income: Sweden Around 1970", *Review of Income and Wealth* 2 (1981), 243-264.
- BRANDOLINI, A. AND P. SESTITO, "Cyclical and Trend Changes in Inequality in Italy, 1977-1991", Banca d'Italia, Servizio Studi, July 1994.
- BRANDOLINI, A. AND P. SESTITO, "Earnings Dispersion in Italy, 1977-1995", Mimeo, Banca d'Italia Research Department, 1996.
- DEL BOCA, D., "Labor Adjustment and Industrial Relations in Italy and the U.S., 1970-1985", unpublished Ph.D. Dissertation, University of Wisconsin-Madison, 1988.
- DUNCAN, G. AND D. HILL, "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data", *Journal of Labor Economics* 3 (1985), 508-531.
- DEMEKAS, D., "Labor Market Institutions and Flexibility in Italy: A Critical Evaluation and some International Comparisons", International Monetary Fund W.P. 94130, 1994.
- FLINN, C., "Wage and Job Mobility of Young Workers", *Journal of Political Economy* 94 (1986), S88-S110.
- FLINN, C., "Equilibrium Wage and Dismissal Processes", *Journal of Business and Economic Statistics* (1997).
- FLINN, C., AND J. HECKMAN, "New Methods for Analyzing Structural Models of Labor Force Dynamics", *Journal of Econometrics* 18 (1982), 115-68.
- FRIEDMAN, M., *Capitalism and Freedom*, Chicago, University of Chicago Press, 1962.
- GEWEKE, J., R. MARSHALL AND G. ZARKIN, "Mobility Indices in Continuous Time Markov Chain", *Econometrica* 54 (1986), 1407-1423.

- GOTTSCHALK, PETER, "Changes in Inequality of Family Income in Seven Industrialized Countries", *American Economic Review* 83 (1993), 136-142.
- GOTTSCHALK, P., AND T. SMEEEDING, "Cross National Comparisons of Levels and Trends in Inequality", Mimeo, Boston College, May 1995.
- GRUBB, D. AND W. WELLS, "Employment Regulation and Patterns of Work in E.C. Countries", *OECD Economic Studies* 21 (1993), 7-58.
- JOVANOVIC, B., "Job Matching and the Theory of Turnover", *Journal of Political Economy* 87 (1979), 972-990.
- LEVY, F. AND R. MURNAME, "U.S. Earnings Levels and Earnings Inequality: A Review of the Recent Trends and Proposed Explanations", *Journal of Economic Literature* 30 (1992), 1333-1381.
- MILLER, R., "Job Matching and Occupational Choice", *Journal of Political Economy* 92 (1984), 1086-1120.
- PRAIS, S., "Measuring Social Mobility", *Journal of the Royal Statistical Society, Series A*, part I 118 (1955), 56-66.
- SHORROCKS, A., "Income Inequality and Income Mobility", *Journal of Economic Theory* 19 (1978), 376-393.
- SHORROCKS, A., "The Measurement of Mobility", *Econometrica* 46 (1978), 1013-1024.
- TRIVELLATO, U., "Forze di Lavoro: Disegno dell'Indagine ed Analisi Strutturali" [in Italian], *Annali di Statistica*, Serie IX, Volume 11, ISTAT, 1991.
- VAN DEN BERG, G. AND G. RIDDER, "An Empirical Equilibrium Search Model of the Labour Market", Mimeo, Free University [Amsterdam], 1993. Forthcoming, *Econometrica*.
- WOLPIN, K., "Estimating a Structural Search Model: The Transition from School to Work", *Econometrica* 55 (1987), 801-817.
- WOLPIN, K., "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth", *Journal of Political Economy* 100 (1992), 535-560.

LA MISURA DELLA DISOCCUPAZIONE E LA MODELLAZIONE DELL'OFFERTA DI LAVORO: DEFINIZIONI A PRIORI E STIME DIPENDENTI DA MODELLI A CONFRONTO

Enrico Rettore e Ugo Trivellato

Dipartimento di Scienze Statistiche, Università di Padova

1. Introduzione e sommario

La misura della disoccupazione è questione che ha una lunga storia, ed è tuttora controversa. Alcune delle basilari difficoltà sono state nitidamente messe in luce, oltre una ventina d'anni fa, da Shiskin (1976), e fondamentalmente hanno a vedere col fatto che si cerca di misurare "il buco" piuttosto che "la ciambella". Ulteriori difficoltà sono poi venute emergendo in conseguenza del dilatarsi di politiche di welfare e/o di politiche attive del lavoro, che tendono ad espandere situazioni ai confini tra gli usuali stati che descrivono la condizione rispetto al lavoro - occupato, disoccupato, inattivo (Malinvaud, 1986).

In letteratura, sono riconoscibili almeno tre diverse linee di attacco all'argomento. Un primo approccio, il più tradizionale e consolidato, focalizza l'attenzione su definizioni e classificazioni a priori, e sul confronto tra informazioni disponibili da differenti fonti (per una recente rassegna, vedi Green, 1995). Un secondo approccio, ricollegandosi alla scuola francese dell'analisi dei dati, muove dalle risposte fornite a tipici insiemi di domande sulla partecipazione al lavoro e/o sulla ricerca di lavoro e conduce analisi di classificazione esplorativa, con lo scopo di individuare insiemi di persone omogenei quanto ad *attachment* al mercato del lavoro (vedi, ad esempio, Rettore, Torelli, and Trivellato, 1990). Vi è, infine, un approccio *model-based*, che affronta il tema della rilevanza della distinzione fra stati in termini di diversità di comportamenti strutturali, eventualmente contaminata da errori di classificazione.

Questo lavoro si colloca nell'ambito dell'approccio *model-based*, ma nel contempo si caratterizza per un'attenzione non superficiale alle modalità con cui viene correntemente rilevata la condizione rispetto al lavoro, nelle indagini sulle forze di lavoro o comunque in *household surveys*, e alle convenzioni - e alle questioni - definitorie che le accompagnano.

Innanzitutto, conviene definire i termini in cui si affronta il problema. Il dibattito sulla misura della condizione rispetto al lavoro - in particolare sulla distinzione tra disoccupazione e inattività -, e sui problemi per l'analisi che derivano da errori nella classificazione della condizione, è formalizzato come un problema di inferenza sui parametri di un modello a

Lavoro svolto nell'ambito del progetto strategico di ricerca CNR "La disoccupazione e il basso livello di attività in Italia: cause specifiche e rimedi adeguati" (contributo n.95.04495.ST74). Ringraziamo Giorgio Brunello, Tito Boeri, Bruno Contini e Renzo Orsi per commenti su una versione preliminare

stati latenti discreti in presenza di indicatori fallibili degli stati latenti. In altre parole, incentriamo l'attenzione sulla situazione in cui in generale non è possibile risalire univocamente agli stati dalla determinazione degli indicatori.

In quest'ambito, guardiamo in particolare:

- (a) alle conseguenze per la stima del modello e di grandezze derivabili dallo stesso, segnatamente i tassi specifici di attività e di disoccupazione, dell'ignorare gli errori di classificazione;
- (b) a condizioni per l'identificazione dei parametri del modello e delle probabilità di classificazione.

Inizialmente, l'argomento viene sviluppato ammettendo semplicemente errore di classificazione nello stato riportato da ciascun individuo (sez. 2). Successivamente, perdendo in qualche modo in generalità ma acquistando in pertinenza, si restringe l'attenzione al tipico pattern di acquisizione delle informazioni sulla partecipazione al lavoro, costituito dalle indagini sulle forze di lavoro, e se ne utilizzano le peculiarità, in termini di disponibilità di risposte ad un più ampio insieme di domande (combinata con plausibili restrizioni circa l'ambito delle incertezze nella classificazione), per la specificazione di un modello dell'offerta di lavoro a stati discreti e con indicatori in parte fallibili, introdotto in Rettore e Trivellato (1993) (sez. 3).

Alla luce di questo quadro, ripresentiamo e discutiamo le evidenze empiriche dell'applicazione del modello ad un campione di donne sposate, dall'indagine sulle forze di lavoro per la Lombardia e il Veneto nel secondo trimestre 1984. In particolare, valutiamo le conseguenze di errori di classificazione sui parametri di interesse e le indicazioni che l'esercizio fornisce in merito al tema della misura della disoccupazione (sez. 4). Ovviamente, la portata sostanziale dei risultati è strettamente riferibile allo specifico ambito di applicazione, e quindi limitata ad un sottoinsieme della popolazione e circoscritta nello spazio e nel tempo. Quale naturale sviluppo, nella conclusiva sez. 5 discutiamo brevemente di prospettive per nuove analisi empiriche, vuoi utilizzando ancora dati dell'indagine sulle forze di lavoro (previa la presa in conto delle consistenti modifiche introdotte dal 1992, segnatamente nel questionario e nelle definizioni), vuoi utilizzando i dati dell'*European Community Household Panel*.

2. Modellazione della partecipazione al lavoro e della disoccupazione in presenza di errori di classificazione

Si intende modellare la relazione tra tassi di attività e di disoccupazione e le caratteristiche x dell'individuo (includendovi eventualmente le condizioni macro in cui l'individuo opera).

Siano O e D variabili binarie definite nel modo seguente:

$$O = \begin{cases} 1 & \text{se l'individuo è occupato} \\ 0 & \text{altrimenti} \end{cases} \quad [1]$$

$$D = \begin{cases} 1 & \text{se l'individuo è disoccupato} \\ 0 & \text{altrimenti} \end{cases} \quad [2]$$

I tassi specifici di attività e di disoccupazione secondo le caratteristiche x sono definiti, rispettivamente, come segue:

$$TA(x) = \Pr(O+D=1|x), \quad [3]$$

$$TD(x) = \Pr(D=1|x, O+D=1) = \Pr(D=1|x) / \Pr(O+D=1|x). \quad [4]$$

La condizione rispetto al lavoro è osservata a meno di errori di classificazione. Nel seguito, consideriamo gli effetti sulla stima di $TA(x)$ e di $TD(x)$ di due distinti processi generatori degli errori di classificazione. Nel primo caso ci si concentra soltanto sulla partecipazione al lavoro, per cui gli errori di classificazione consistono semplicemente nell'inversione della condizione di attivo/inattivo per una frazione di individui. Nel secondo caso sono prese in considerazione le tre condizioni - occupato, disoccupato e inattivo: si muove da una formulazione generale in cui si ammette che l'errore di classificazione possa spostare individui da qualsiasi condizione a qualsiasi altra condizione; ci si concentra infine su un caso particolare ma, come si vedrà nella sezione successiva, rilevante per il dibattito in corso, caso in cui la condizione di occupazione è rilevata senza errore mentre possono avere luogo erronee classificazioni di disoccupati tra gli inattivi e viceversa.

Con riferimento ad entrambi i casi si discutono, nell'ordine, (i) gli effetti sui tassi della presenza di errori di classificazione e (ii) alcune condizioni utili all'identificabilità dei tassi e del processo generatore degli errori di classificazione.

2.1. Modellazione della probabilità di partecipazione al lavoro

La variabile $A=O+D$ è osservabile con errore. Sia \bar{A} la variabile binaria affetta da errore, legata ad A dalla seguente relazione:

$$\Pr(\bar{A}=1|A) = \begin{cases} p & \text{se } A = 1 \\ 1 - q & \text{se } A = 0 \end{cases} \quad [5]$$

In altre parole, si assume che (i) la probabilità di corretta classificazione dipenda solo dalla vera condizione, e non dalle caratteristiche x , e che (ii) tale probabilità sia pari a p (q) per gli individui attivi (inattivi).

Stante il processo [5] di generazione degli errori di classificazione, segue che il tasso specifico di attività osservato $\bar{TA}(x)$ è pari a:

$$\bar{TA}(x) = \Pr(A=1|x) p + \Pr(A=0|x) (1-q) = (p+q-1) \Pr(A=1|x) + (1-q). \quad [6]$$

Utilizzando la [6], è immediato ricavare la seguente diseuguaglianza (per i casi in cui almeno uno tra p e q è inferiore a 1):

$$\bar{TA}(x) > TA(x) \text{ se e solo se } TA(x) < (1-q)/(2-p-q). \quad [7]$$

Ciò equivale a dire che i tassi specifici di attività osservati sovrastimano i veri tassi se questi ultimi sono inferiori ad una certa soglia (che dipende solo da p e q) e li sottostimano se sono superiori a tale soglia. Ne risulta quindi una attenuazione dell'eterogeneità di tali tassi. In particolare:

- (a) se $p=q$, vale a dire se la probabilità di corretta classificazione non dipende dallo stato, la soglia in questione è pari a 0.5;
- (b) se $q=1$, $p<1$, cioè se tutti gli inattivi sono classificati correttamente, ma alcuni attivi sono classificati erroneamente, la soglia è pari a zero, vale a dire che tutti i tassi specifici di attività sono sottostimati;
- (c) se $p=1$, $q<1$, cioè se tutti gli attivi sono classificati correttamente, ma alcuni inattivi sono classificati erroneamente, la soglia è pari a 1, vale a dire che tutti i tassi specifici di attività sono sovrastimati.

Utilizzando la [6], è pure immediato ricavare che:

$$\partial \bar{TA}(x) / \partial x = (p+q-1) \partial TA(x) / \partial x, \quad [8]$$

con $(p+q-1)$ in modulo inferiore (o al più eguale) a 1, positivo se la somma delle probabilità di corretta classificazione è superiore a 1, altrimenti negativo. La [8] dice che: (i) l'effetto osservato di una variazione di x sul tasso specifico di attività risulta attenuato rispetto al vero effetto (salvo, ovviamente, il caso in cui $p=q=1$, cioè se non vi sono errori di classificazione); (ii) il segno dell'effetto viene mantenuto se $(p+q) > 1$, altrimenti si inverte; (iii) il rapporto tra gli effetti osservati della variazione di due distinte variabili x_i e x_j sul tasso specifico di attività risulta pari al vero rapporto (se si esclude il caso $p+q=1$ nel quale le derivate [8] risultano nulle).

Se è ragionevole ritenere che vi siano individui nella popolazione con caratteristiche x_0 tali per cui il loro tasso di attività è approssimativamente nullo ($TA(x_0) \cong 0$) ed altri individui con caratteristiche x_1 tali per cui il loro tasso di attività è approssimativamente pari a 1 ($TA(x_1) \cong 1$), allora:

$$\bar{TA}(x_0) \cong 1-q, \quad [9.1]$$

$$\bar{TA}(x_1) \cong p, \quad [9.2]$$

vale a dire che le probabilità di corretta classificazione emergono dai tassi di attività osservati per gli individui con caratteristiche x_0 e x_1 .

Si noti che per ottenere l'identificabilità di p e q non è necessario conoscere x_0 e x_1 . È sufficiente che esistano nella popolazione di riferimento. In tale caso infatti, valgono le seguenti identità:

$$q = 1 - \min_x \bar{TA}(x), \quad [10.1]$$

$$p = \max_x \bar{TA}(x). \quad [10.2]$$

Dalla [6], si ricava immediatamente che l'identificabilità di p e q comporta l'identificabilità del tasso di attività specifico.

Nella modellazione della partecipazione al lavoro, si adotta frequentemente la seguente parametrizzazione per $TA(x)$:

$$\bar{TA}(x) = F(\beta' x), \quad [11]$$

dove β sono i coefficienti della regressione binomiale della variabile binaria A sulle variabili x e $F(\cdot)$ è una funzione di ripartizione. Tale parametrizzazione impone restrizioni su $TA(x)$ (il rapporto $\partial TA(x)/\partial x_i / \partial TA(x)/\partial x_j$, risulta pari a β_i/β_j in ogni punto del dominio di x , per ogni coppia i, j) ed è giustificabile economicamente, ad esempio, se si interpreta la decisione riguardo alla partecipazione al lavoro come il risultato del confronto operato da ogni potenziale partecipante tra saggio salariale di riserva e saggio salariale di mercato. Se vale la [11], la [6] diventa:

$$\bar{TA}(x) = (p+q-1) F(\beta' x) + (1-q). \quad [12]$$

Il problema di inferenza sui parametri ignoti della [12] (p, q, β) è discusso in Hausman e Scott Morton (1993). Tra i risultati conseguiti da tali autori, segnaliamo i seguenti (in alcuni casi riformulati per guadagnare in semplicità espositiva):

(a) Quanto all'identificabilità dei parametri ignoti (p, q, β) e della funzione $F(\cdot)$, se sono soddisfatte le condizioni per l'identificabilità del modello di regressione binomiale sta-

bilita in Manski (1988), allora sono identificabili β e $(p+q-1)F(\cdot) + (1-q)$. Facendo tendere $a \rightarrow +\infty$ l'argomento della funzione $F(\cdot)$ si ottiene p , facendolo tendere $a \rightarrow -\infty$ si ottiene $1-q$, per cui risultano identificabili sia p , che q , che $F(\cdot)$.

- (b) Se si adotta un approccio parametrico e non si commettono errori di specificazione nella scelta di $F(\cdot)$, la funzione di verosimiglianza per i parametri p , q , β , derivata facendo uso dell'espressione [12], dà luogo a stimatori consistenti di tutti i parametri ignoti.
- (c) È possibile stimare in modo semiparametrico, vale a dire senza imporre assunzioni troppo stringenti su $F(\cdot)$, sia β che p e q . Ciò è vero perché $(p+q-1)F(\beta'x) + (1-q)$ è una trasformazione monotona, continua e derivabile (se $F(\cdot)$ lo è) di $\beta'x$, per cui β è stimabile semiparametricamente in modo consistente (a meno di un parametro di scala; vedi ad esempio Manski, 1985).
Data una stima consistente di $\hat{\beta}$, è possibile stimare consistentemente in modo non parametrico la regressione di \hat{A} , la condizione affetta da errore di classificazione, su $\hat{\beta}'x$, vale a dire $(p+q-1)F(\hat{\beta}'x) + (1-q)$. Il valore massimo e (1 meno) il valore minimo di tale regressione forniscono una stima di p e q .
- (d) Se vi sono errori di classificazione (p e q sono minori di 1), le stime di massima verosimiglianza di β ottenute ignorando l'esistenza di tali errori sono inconsistenti e la distorsione asintotica può essere anche molto grave.

Vale la pena di esplicitare che in presenza di errori di classificazione, mentre è sempre possibile (se sono soddisfatte opportune condizioni di regolarità) stimare consistentemente in modo semiparametrico β , la possibilità di stimare consistentemente p , q ed $F(\cdot)$, e conseguentemente la possibilità di stimare $TA(x)$, poggia crucialmente sull'ipotesi che nella popolazione di riferimento il tasso di attività specifico $TA(x)$ assuma anche i valori estremi 0 e 1.

2.2. Modellazione congiunta della probabilità di partecipazione al lavoro e di disoccupazione

Siano $O, D, (1-O-D)$, le variabili indicatrici della vera condizione rispetto al lavoro e $\bar{O}, \bar{D}, (1-\bar{O}-\bar{D})$ le loro controparti affette da errore di classificazione. Sia la seguente relazione a collegare la condizione osservata alla vera condizione:

$$\Pr(\bar{O} = 1|y) = \begin{cases} p(olo) & \text{se } y=O=1 \\ p(old) & \text{se } Y=D=1 \\ p(oli) & \text{se } y=1-O-D=1 \end{cases} \quad [13]$$

Quindi, $p(alb)$ è la probabilità che un individuo nella condizione b sia classificato nella condizione a : correttamente se $a=b$, erroneamente se $a \neq b$. In modo analogo sono definite le probabilità $p(dlo)$, $p(dld)$, $p(dli)$, e $p(ilo)$, $p(ild)$, $p(ili)$. Si noti che, come nella sez. 2.1, si assume che le probabilità di corretta e errata classificazione dipendano solo dalla vera condizione e non dalle caratteristiche x .

Stante il processo [13] di generazione degli errori di classificazione, segue che il tasso specifico di attività osservato $\bar{TA}(x)$ è pari a:

$$\begin{aligned} \bar{TA}(x) &= \Pr(O=1|x) (p(olo) + p(dlo)) + \Pr(D=1|x) (p(old) + p(dld)) + \\ &\quad \Pr(O=D=0|x) (p(oli) + p(dli)) \\ &= \Pr(O=1|x) (p(olo) + p(dlo) - p(oli) - p(dli)) + \\ &\quad \Pr(D=1|x) (p(old) + p(dld) - p(oli) - p(dli)) + (p(oli) + p(dli)). \end{aligned} \quad [14]$$

In questo contesto, l'analisi degli effetti degli errori di classificazione sul tasso di attività è più complessa. Si può però procedere all'analisi di casi particolari. Tra questi, è specificamente rilevante per i dibattiti in corso sulla misura della disoccupazione il caso in cui:

$$p(olo)=1, \quad [15.1]$$

$$p(old)=p(oli)=0; \quad [15.2]$$

vale a dire quello in cui l'insieme degli occupati osservato include tutti e solo i veri occupati, mentre alcuni disoccupati e alcuni inattivi sono erroneamente inclusi rispettivamente tra gli inattivi e tra i disoccupati.

In tale caso la [14] diventa:

$$\begin{aligned} \bar{TA}(x) &= \Pr(O=1|x) + \Pr(D=1|x) p(dld) + \Pr(O=D=1|x) p(dli) \\ &= p(ili) TA(x) - p(ild) \Pr(D=1|x) + p(dli). \end{aligned} \quad [16]$$

È immediato ricavare la seguente diseuguaglianza:

$$\bar{TA}(x) > TA(x) \text{ se e solo se } \frac{\Pr(D=1|x)}{\Pr(O=D=0|x)} < \frac{p(dli)}{p(ild)}. \quad [17]$$

Ciò equivale a dire che i tassi di attività specifici osservati sovrastimano i corrispondenti veri tassi nei casi (e solo nei casi) in cui il numero di veri disoccupati in rapporto al numero di veri inattivi è inferiore ad una certa soglia (che dipende solo dalle probabilità di errata classificazione dei disoccupati e degli inattivi).

Utilizzando la [16], è pure immediato ricavare che:

$$\partial \bar{TA}(x)/\partial x = p(ili) \partial TA(x)/\partial x - p(ild) \partial \Pr(D=1|x)/\partial x. \quad [18]$$

Ne segue che: (i) se $\partial TA(x)/\partial x$ e $\partial \Pr(D=1|x)/\partial x$ hanno lo stesso segno, allora $\partial \bar{TA}(x)/\partial x$ risulterà attenuata o con il segno cambiato rispetto a $\partial TA(x)/\partial x$, secondo i casi; (ii) se $\partial TA(x)/\partial x$ e $\partial \Pr(D=1|x)/\partial x$ hanno segno opposto, il risultato è incerto, dato che la moltiplicazione di $\partial TA(x)/\partial x$ per $p(ili)$ e l'aggiunta di $-p(ild) \partial \Pr(D=1|x)/\partial x$ agiscono in senso l'uno contrario all'altro (tuttavia, essendo plausibilmente $p(ili)$ prossimo a 1, è verosimile che prevalga l'effetto del secondo termine e ne risulti quindi una sovrastima del valore assoluto di $\partial TA(x)/\partial x$).

Quanto al tasso specifico di disoccupazione osservato, esso risulta pari a:

$$\bar{TD}(x) = [\Pr(O=1|x) p(dlo) + \Pr(D=1|x) p(dld) + \Pr(O=D=0|x) p(dli)]/TA(x). \quad [19]$$

Se valgono le condizioni [15], la [19] diventa:

$$\bar{TD}(x) = [\Pr(D=1|x) p(dld) + \Pr(O=D=0|x) p(dli)] / [\Pr(O=1|x) + \Pr(D=1|x) p(dld) + \Pr(O=D=1|x) p(dli)]. \quad [20]$$

L'effetto degli errori di classificazione sul tasso di disoccupazione non è immediatamente visibile. I suoi effetti principali possono però essere colti ricavando lo sviluppo in serie di Taylor della [20] rispetto a $p(dld)$ e $p(dli)$, nel punto $p(dld) = 1$ e $p(dli) = 0$ (che corrisponde al caso in cui non vi sono errori di classificazione):

$$\bar{TD}(x) \cong TD(x) + \frac{\Pr(O=1|x)}{TA(x)^2} [\Pr(D=1|x) (p(dld)-1) + \Pr(O=D=1|x) p(dli)], \quad [21]$$

dal quale si ricava immediatamente che in prima approssimazione vale la seguente condizione:

$$\bar{TD}(x) > TD(x) \text{ se e solo se } \frac{\Pr(D=1|x)}{\Pr(O=D=0|x)} < \frac{p(dli)}{1-p(ild)}, \quad [22]$$

vale a dire che il tasso di disoccupazione specifico risulta sovrastimato nei casi (e solo nei casi) in cui il numero di veri disoccupati in rapporto al numero di veri inattivi è inferiore ad una certa soglia (che dipende solo dalle probabilità di errata classificazione dei disoccupati e degli inattivi). Dato che la soglia in questione coincide con quella che compare nella [17], ne risulta che i tassi di disoccupazione specifici sono sovrastimati in tutti e soli i casi in cui sono sovrastimati anche i tassi di attività specifici.

Quanto agli effetti degli errori di classificazione sulle derivate di $TD(x)$, in generale, non sembrano esservi risultati ovvi.

Per quel che riguarda l'identificabilità delle probabilità di corretta e errata classificazione e, conseguentemente, l'identificabilità dei tassi di attività e di disoccupazione specifici, le restrizioni proposte da Hausman e Scott Morton (1993), viste in precedenza, necessitano di alcuni adattamenti per risultare utili nel caso in questione.

Dalla relazione che collega \bar{O} , la variabile indicatrice della condizione di occupazione affetta da errore, a $O, D, 1-O-D$, le tre variabili indicatrici della vera condizione:

$$\Pr(\bar{O}=1|x) = \Pr(O=1|x) p(olo) + \Pr(D=1|x) p(old) + \Pr(O=D=0|x) p(oli), \quad [23]$$

e dalle due analoghe relazioni relative a \bar{D} e $1-\bar{O}-\bar{D}$, si ricava quanto segue:

- (a) Se nella popolazione vi sono individui con caratteristiche x_0 (x_D, x_I) tali che la loro probabilità di occupazione (disoccupazione, inattività) è pari a 1 e tali caratteristiche sono note, allora risultano identificabili $p(olo)$, $p(dlo)$ e $p(ilo)$ ($p(old)$, $p(dld)$ e $p(ild)$; $p(oli)$, $p(dli)$ e $p(ili)$). Infatti, se è soddisfatta tale condizione, valgono le identità:

$$\begin{aligned} \Pr(\bar{O}=1|x_0) &= p(olo), \\ \Pr(\bar{D}=1|x_0) &= p(dlo), \\ \Pr(1-\bar{O}-\bar{D}=1|x_0) &= p(ilo), \end{aligned} \quad [24]$$

(analoghe identità forniscono le altre probabilità di corretta e scorretta classificazione).

- (b) L'informazione a priori sulla nullità di $\Pr(O=1|x)$ ($\Pr(D=1|x)$, $\Pr(O=D=0|x)$) in corrispondenza a certi valori di x , non consente l'identificabilità di alcun parametro di interesse. Ad esempio, la nullità di $\Pr(O=1|x)$ nella [23], comporta che vi sono infiniti modi di scegliere $\Pr(D=1|x)$, $\Pr(O=D=0|x)$, $p(old)$ e $p(oli)$ (soddisfacenti il vincolo $\Pr(D=1|x) + \Pr(O=D=0|x)=1$) compatibili con il valore osservato $\Pr(\bar{O}=1|x)$.
- (c) Se valgono le condizioni dette al punto (a), ma le caratteristiche x_0 (x_D, x_I) non sono note, in assenza di ulteriori condizioni i parametri non sono identificabili. Infatti, la soluzione del problema:

$$\max_x \Pr(\bar{O}=1|x)$$

analogo ai [10], non identifica alcunché nel caso in esame: il massimo della media ponderata [23] ha luogo in corrispondenza al valore di x che attribuisce peso pari a 1 alla maggiore delle tre probabilità $p(olo)$, $p(old)$ e $p(oli)$, vale a dire che ha luogo in x_0 , oppure in x_D , oppure in x_I .

Nelle situazioni in cui la seguente restrizione è ragionevole:

$$p(o|o) > p(o|d) \quad [25.1]$$

$$p(o|o) > p(o|i), \quad [25.2]$$

vale a dire se gli occupati correttamente classificati sono mediamente più numerosi sia dei disoccupati che degli inattivi erroneamente classificati tra gli occupati, i parametri risultano identificabili. In tale caso infatti, $\Pr(\bar{O}=1|x)$ assume il suo valore massimo, pari a $p(o|o)$, in corrispondenza a x_o , che risulta in tale modo a sua volta identificabile. Essendo identificabile x_o , per quanto detto al punto (a) risultano identificabili anche $p(d|o)$ e $p(i|o)$.

Se vi sono ragioni per assumere che anche per $p(d|d)$ e $p(i|i)$ valgano condizioni analoghe alle [25], si ricava l'identificabilità delle altre probabilità di corretta ed errata classificazione.

3. Misura e modellazione della partecipazione al lavoro e della disoccupazione utilizzando molteplici indicatori fallibili

Nell'approccio delineato sinora, la possibilità di stima consistente dei parametri di interesse β e delle probabilità di corretta/errata classificazione, e quindi del tasso di attività (e, nel caso considerato nella sez. 2.2, anche del tasso di disoccupazione) poggia, al meglio, sull'assunto che nella popolazione di riferimento il tasso specifico di attività assuma (anche) i valori estremi 0 e 1 (e, come si è appena visto nel caso di modellazione congiunta delle probabilità di partecipazione al lavoro e di disoccupazione, su assunti ancora più stringenti su $TA(x)$ e $TD(x)$ o sulle probabilità di corretta/errata classificazione).

Ora, la ragionevolezza di tali assunti è palesemente dubbia: perlomeno discutibile per il tasso di attività e affatto implausibile per il tasso di disoccupazione, anche quando si disponga di un insieme sufficientemente esteso di caratteristiche individuali. D'altra parte, v'è da considerare che le tipiche fonti per la misura della partecipazione al lavoro, cioè a dire le indagini sulle forze di lavoro, non rilevano direttamente lo stato dell'individuo rispetto al lavoro - occupato o disoccupato o inattivo - ma giungono a classificarlo in uno dei tre stati sulla scorta delle risposte ad un più articolato insieme di domande (Husmanns, Merhan e Verma, 1990). Ancora, v'è da aggiungere che, tanto in letteratura quanto nella pratica degli istituti statistici nazionali, la classificazione degli individui negli stati è per larga parte dei casi non controversa, mentre incertezza e dissenso si concentrano in alcune "aree grigie" ai confini fra i tre stati e segnatamente tra disoccupazione e inattività (vedi, tra gli altri, Shiskin, 1976; Malinvaud, 1986; Rettore, Torelli e Trivellato, 1990).

Tutto ciò suggerisce che è opportuno affrontare le questioni di misura e di modellazione della partecipazione al lavoro in un contesto maggiormente circostanziato. Il prezzo che si paga in termini di assunzioni aggiuntive sul processo di misura non è alto, appunto perchè si tratta di proposizioni non particolarmente forti e, soprattutto, condivise. I vantaggi che se ne hanno sono d'altronde evidenti: i parametri d'interesse del modello di partecipazione al lavoro risultano identificabili (date queste restrizioni sul processo di misura, sulle quali torneremo tra poco), qualunque sia il campo di variazione dei tassi di attività e di disoccupazione; le varianti nelle definizioni a priori in tema di misura della disoccupazione, sulle quali vi è dissenso, risultano essere restrizioni di sovraidentificazione, e possono quindi essere sottoposte a verifica.

Per procedere, conviene prendere le mosse dalle caratteristiche salienti delle convenzioni adottate internazionalmente per la classificazione delle persone secondo la condizione

rispetto al lavoro e i modi in cui tali convenzioni sono rese operative. Com'è noto, i criteri basilari si ritrovano nelle raccomandazioni dell'ILO (vedi Hussmanns, Merhan e Verma, 1990). In sintesi, con riguardo ad un periodo di riferimento - generalmente una settimana - sono considerate:

- (a) occupate le persone che hanno (dichiarato di aver) svolto un lavoro produttivo anche per una sola ora;
- (b) disoccupate le persone che rispondono a tre criteri, cioè a dire sono (o meglio, dichiarano di essere): (b1) senza lavoro; (b2) disponibili a lavorare; (b3) alla ricerca attiva di un lavoro, nel senso che hanno compiuto specifici passi nel corso di un determinato periodo recente per cercare lavoro;
- (c) inattive le persone non incluse tra gli occupati e i disoccupati.

In merito a queste definizioni e criteri, vanno evidenziati almeno tre punti. Innanzitutto, è palese che le condizioni di occupazione, disoccupazione e inattività sono mutuamente esclusive e sono disposte gerarchicamente. In secondo luogo, come già anticipato, la classificazione delle persone non avviene sulla scorta di uno spazio delle risposte che corrisponde allo spazio degli stati - e che poggia grosso modo su una "autoclassificazione" - bensì in base alle risposte ad un più articolato insieme di domande (sul lavoro, le ore lavorate, la disponibilità a lavorare, l'attività di ricerca di lavoro; domande che, tra l'altro, mirano a rilevare fatti piuttosto che a registrare l'opinione dell'individuo circa la sua condizione rispetto al lavoro). Infine, le raccomandazioni dell'ILO sono naturalmente espresse in termini generali, e sono variamente interpretate e adattate nei diversi contesti nazionali.

È proprio in sede di specificazione delle raccomandazioni a livello nazionale, in termini sia di formulazione dei quesiti sia (e soprattutto) di definizioni operative, che emergono aree di dibattito e di potenziale dissenso, e che si hanno conseguentemente scelte discutibili. Tipicamente, le "aree grigie", cioè a dire gli insiemi di persone di problematica collocazione, si situano ai confini fra occupati e disoccupati e fra disoccupati e inattivi. Alcune situazioni in prossimità di queste frontiere, che hanno rilievo a fini tanto interpretativi che di politiche, sono costituite per un verso dalla sottoccupazione visibile e dalle condizioni di occupazione-formazione (vedi, ad esempio, Malinvaud, 1986) e per un altro verso da non occupati con vario grado di attachment al mercato del lavoro, sino ai cosiddetti "lavoratori scoraggiati" (vedi, tra gli altri, Flaim, 1984; OECD, 1987; Rettore Torelli e Trivellato, 1990 e 1993).

In questa sede, fermiamo l'attenzione sulle "aree grigie" ai confini fra disoccupazione e inattività, sia per la particolare rilevanza che anche recentemente hanno assunto nel dibattito sul mercato del lavoro italiano, sia per la cospicua variante definitoria introdotta da ottobre 1992 nelle stime della disoccupazione dall'indagine sulla forze di lavoro, a seguito dell'armonizzazione dell'indagine alle direttive comunitarie (vedi, tra gli altri, Sestito, 1989; Trivellato, 1993; Casavola e Sestito, 1994).

In sintesi, le questioni aperte sono due.

- (a) La prima ruota attorno all'interrogativo di quanto debba essere stringente il criterio/test sulla ricerca di lavoro per qualificare una persona come disoccupata. Al riguardo, è utile riferirsi alla polarizzazione delle posizioni che emerge dalle scelte definitorie adottate dall'Istat rispettivamente prima del e dall'ottobre 1992 (scelte che, implicitamente ma palesemente, rimandano a schemi interpretativi, qui soltanto evocati, di un'offerta sovrabbondante dominata tout court dalla domanda e, all'opposto, di non insignificanti capacità di scegliere dei singoli rispetto al lavoro).
 - (a1) Fino a giugno 1992 è stato adottato un criterio piuttosto lasco: se una persona dichiarava di cercare lavoro e di aver compiuto azioni di ricerca, tanto bastava per ritenerla disoccupata, indipendentemente dal periodo - anche piuttosto remoto - al quale risaliva l'ultima azione di ricerca compiuta.

- (a2) Da ottobre 1992, in conformità con una direttiva dell'UE, è accolto un criterio parecchio più restrittivo, che accerta la volontà/disponibilità a lavorare tramite l'effettuazione di una specifica azione di ricerca di lavoro nel corso dell'ultimo mese.
- (b) La seconda questione attiene al trattamento dei "lavoratori scoraggiati", cioè, in prima approssimazione, delle persone disponibili a lavorare, ma che (dichiarano che) non cercano un lavoro o che hanno smesso di cercarlo convinte di non poterlo trovare. La loro abituale inclusione tra gli inattivi risulta per l'appunto dal mancato soddisfacimento del requisito riguardante l'essere (meglio, il dichiararsi) alla ricerca di lavoro. Vi è peraltro un'ampia letteratura in merito ai possibili criteri di rilevazione dello scoraggiamento, alla pertinenza della classificazione degli scoraggiati fra gli inattivi (e non fra i disoccupati), allo studio del fenomeno (vedi, tra gli altri, Flaim, 1984; OECD, 1987; Rettore, Torelli e Trivellato, 1990).

Rispetto a queste questioni, una strategia ragionevole appare quella di ammettere che per alcuni insiemi di persone, che si collocano appunto in queste "aree grigie", la classificazione in uno stato o nell'altro, sulla base delle risposte ai singoli quesiti del questionario, sia incerta. Ciò equivale appunto ad ammettere che per un sottoinsieme del campione la condizione rispetto al lavoro sia colta da indicatori fallibili, e motiva l'approccio che combina un modello bivariato dell'offerta di lavoro formulato in termini di stati discreti parzialmente non osservati con un modello di misura il quale utilizza le risposte ai singoli quesiti. Di tale approccio, proposto in Rettore e Trivellato (1993), si richiama, in termini semplificati, la formulazione che attiene al modello di misura.

Si assume che, a fronte della variabile y , di volta in volta coincidente con una delle tre variabili indicatrici O , D e $1-O-D$ come in [13], parzialmente non osservabile (nel modo in cui si dirà tra poco), si dispone per ogni unità campionaria della determinazione di una variabile casuale discreta R , dipendente stocasticamente da y , costruita a partire dalle risposte ai singoli quesiti del questionario, la cui funzione di probabilità condizionata è:

$$\Pr(R=l | y, x) = \Pr(R=l | y), l = 1, r \text{ con } r \geq 3. \quad [26]$$

Si noti che: (i) si continua ad assumere, come in [5] e in [13], che R , l'indicatore dello stato, dipenda stocasticamente soltanto dalla vera condizione rispetto al lavoro, e non anche da x^1 ; (ii) in generale, le modalità che assume la variabile R sono in numero superiore a quello delle condizioni rispetto al lavoro ($r \geq 3$), proprio perché si utilizzano le risposte all'insieme dei quesiti del questionario sulla partecipazione al/ricerca di lavoro; (iii) R è interpretabile come un indicatore fallibile della condizione non osservata, dove "fallibile" significa che in generale non è possibile risalire ad y dalla determinazione di R , e che è possibile risalire soltanto ad una distribuzione di probabilità condizionata non degenerare $\Pr(y | R=l, x)$ sulle tre condizioni.

Entro questo quadro, è agevole incorporare assunzioni circa le relazioni fra condizione e indicatore fallibile:

- (a) ad esempio, del tipo:

$$p(l|y) > 0 \text{ soltanto se } y=O=1, \quad [27]$$

la quale implica che la l -esima modalità di R può essere fornita soltanto da individui che appartengono alla condizione di occupato (ma, si noti, non esclude la possibilità che un individuo occupato dia luogo ad una determinazione di R diversa da l);

¹ Tale assunzione è mantenuta per semplificare l'analisi. Si ammetta che sia le probabilità [26], $\Pr(R|y, x)$ che le probabilità di appartenenza ai vari stati, $\Pr(y|x)$, dipendano da caratteristiche osservabili x degli individui. Una condizione sufficiente per l'identificabilità è che vi sia almeno una variabile in x esclusa da $\Pr(R|y, x)$ (è una conseguenza immediata dei risultati ottenuti in Rettore, Trivellato, 1993).

(b) oppure del tipo:

$$p(1|y) > 0 \text{ se } y=O=1 \text{ oppure } y=D=1, \quad [28]$$

nella quale l'osservazione della modalità l -esima restringe a 2 l'insieme delle condizioni alle quali l'individuo può appartenere, ma rimane comunque ambigua perché non individua univocamente la condizione;

(c) o ancora del tipo che esclude errori di misura, caratterizzato dal fatto che, mediante opportuni accorpamenti delle modalità, $r=3$ e ogni modalità di R è univocamente associata ad una e una sola condizione.

Nel caso specifico, tenendo presenti le questioni aperte appena individuate e muovendo dalle risposte ai vari quesiti, una ragionevole specificazione delle modalità di R , è la seguente:

- 1.OCC: numero di ore lavorate nella settimana maggiore di zero;
- 2.RIC1: numero di ore lavorate nella settimana pari a zero; in cerca di lavoro; immediatamente disponibile a lavorare; un'azione di ricerca compiuta nell'ultimo mese;
- 3.RIC2: numero di ore lavorate nella settimana pari a zero; in cerca di lavoro; immediatamente disponibile a lavorare; ultima azione di ricerca compiuta da uno a sei mesi prima;
- 4.RIC3: numero di ore lavorate nella settimana pari a zero; in cerca di lavoro; immediatamente disponibile a lavorare; ultima azione di ricerca compiuta più di sei mesi prima;
- 5.RIC4: numero di ore lavorate nella settimana pari a zero; in cerca di lavoro; immediatamente disponibile a lavorare; azioni concrete di ricerca non ancora iniziate;
- 6.NORIC1: numero di ore lavorate nella settimana pari a zero; non in cerca di lavoro per i seguenti motivi: (i) vana ricerca nel passato, oppure (ii) convinzione di non disporre di sufficiente preparazione professionale, oppure (iii) è considerato troppo giovane o troppo vecchio dai datori di lavoro;
- 7.NORIC2: numero di ore lavorate nella settimana pari a zero; non in cerca di lavoro per motivi diversi da quelli di NORIC1.

Alla luce della discussione precedente sulle aree di incertezza classificatoria, è poi affatto plausibile imporre le seguenti restrizioni:

$$p(1|y) = \begin{cases} 1 & \text{se } y=O=1 \\ 0 & \text{altrimenti} \end{cases} \quad [29.1]$$

$$p(2|y) \neq 0 \text{ solo se } y=D=1 \quad [29.2]$$

$$p(7|y) \neq 0 \text{ solo se } y=1=O=D=1 \quad [29.3]$$

Palesemente, l'assunto [29.1] stabilisce una corrispondenza biunivoca fra la condizione di occupato e la modalità OCC. Per parte sua, l'assunto [29.2] afferma che fra coloro per i quali R assume la modalità RIC1 vi sono soltanto disoccupati. Infine l'assunto [29.3] stabilisce che fra coloro per i quali R assume la modalità NORIC2 vi sono soltanto inattivi. Di conseguenza, coloro per i quali R assume le modalità RIC2, RIC3, RIC4 o NORIC1 possono essere disoccupati oppure inattivi².

Siano poi $\Pr(O=1|x;\beta)$, $\Pr(D=1|x;\beta)$ e $\Pr(O=D=0|x;\beta)$ le probabilità di sperimentare le tre condizioni, delle quali si è esplicitata la dipendenza da certi parametri ignoti β , oggetto del problema di inferenza al pari delle probabilità di classificazione.

Con riferimento a questa specificazione del modello, si tratta innanzitutto di verificare se le restrizioni imposte sulla matrice delle probabilità di classificazione:

$$P = \begin{bmatrix} p(1|O=1) & p(1|D=1) & p(1|1-O-D=1) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ p(7|O=1) & p(7|D=1) & p(7|1-O-D=1) \end{bmatrix} \quad [30]$$

assicurano l'identificabilità dei parametri ignoti coinvolti nel modello.

Risultati generali sull'identificabilità di un modello con risposte discrete latenti e indicatori fallibili della risposta sono in Rettore e Trivellato (1993, Appendice A), che studiano le condizioni per l'identificabilità in presenza di un insieme di restrizioni lineari sulle probabilità di classificazione. (Incidentalmente, vale la pena di notare che si considerano soltanto restrizioni sul modello di misura, e non si esamina se assunzioni sul modello di interesse possono costituire una fonte ulteriore di restrizioni di identificabilità.)

Dai risultati generali, è agevole ricavare che le restrizioni [29] soddisfano le condizioni necessarie per l'identificabilità (il soddisfacimento della condizione sufficiente dipende dai veri e ignoti valori assunti dai parametri liberi). E' quindi possibile procedere a stimare congiuntamente il modello di offerta di lavoro di interesse e il modello di misura, ottenendo stime consistenti tanto dei parametri β quanto degli elementi non ristretti di P , e quindi dei tassi di attività e di disoccupazione.

Nel caso in esame, grazie al fatto che le restrizioni a priori vincolano a zero certe probabilità di errata classificazione, il problema di stima si presta ad essere scomposto in due parti, autonome l'una dall'altra. Tale modo di procedere dà luogo a stime che, pur essendo consistenti, sono meno efficienti delle stime di massima verosimiglianza; esso serve esclusivamente ad illustrare in modo semplice perché le restrizioni proposte consentono l'identificabilità dei parametri (un modo computazionalmente conveniente per ottenere le stime di massima verosimiglianza è in Rettore e Trivellato, 1993, sez. 5).

Si procede innanzitutto alla stima di β , mediante la massimizzazione di una funzione di verosimiglianza nella quale non compaiono le probabilità di classificazione. Infatti, le restrizioni imposte a priori comportano che di ogni unità campionaria, secondo le determinazioni assunte dall'indicatore R, è possibile dire la condizione in cui è, o, alla peggio, la condizione in cui non è: le unità per le quali R=OCC contribuiscono alla verosimiglianza con $\Pr(O=1|x;\beta)$, le unità per le quali R=RIC1 con $\Pr(D=1|x;\beta)$, le unità per le quali

² In linea di principio l'assunto di esatta corrispondenza fra la condizione di occupazione e la modalità OCC dell'indicatore R può essere rilasciata ammettendo che alcune delle persone che si dichiarano alla ricerca di una occupazione siano in realtà occupate. In tal modo, peraltro, si mette in discussione uno degli assunti - base dell'impianto definitorio dell'ILO, vale a dire che sia possibile identificare operativamente coloro che svolgono almeno un'ora di lavoro nella settimana di riferimento.

$R=NORIC2$ con $\Pr(O=D=0|x;\beta)$; le rimanenti unità, delle quali non si dice a priori se siano disoccupati o inattivi, contribuiscono con $\Pr(O=0|x;\beta)$, vale a dire la probabilità di non essere occupati. Ne risulta la verosimiglianza seguente:

$$L(\beta) = \prod_{i=1}^n \Pr(O=1|x_i;\beta)^{I(R_i=OCC)} \Pr(D=1|x_i;\beta)^{I(R_i=RIC1)} \Pr(O=D=0|x_i;\beta)^{I(R_i=NORIC2)} \Pr(O=0|x_i;\beta)^{I(R_i=RIC2, \dots, NORIC1)}, \quad [31]$$

dove $I(A)$ è la variabile binaria che assume il valore 1 se l'evento A è verificato e il valore 0 altrimenti. Al divergere di n la massimizzazione della [31] fornisce il vero valore di β .

Ottenute nel modo detto stime consistenti di β , si procede alla stima delle probabilità di classificazione. Si consideri il caso di $p(RIC2|D=1)$ e $p(RIC2|O=D=0)$. Si consideri la media di $I(R=RIC2)$ condizionata ad x nella popolazione. Grazie agli assunti [26] e [29], vale l'identità:

$$E\{I(R=RIC2)|x\} = \Pr(R=RIC2|x) = \Pr(D=1|x;\beta) p(RIC2|D=1) + \Pr(O=D=0|x;\beta) p(RIC2|O=D=0) = E\{I(R=RIC2) | \Pr(D=1|x;\beta), \Pr(O=D=0|x;\beta)\}, \quad [32]$$

vale a dire che (i) nella regressione della variabile binaria $I(R=RIC2)$ su x , le variabili x compaiono solo tramite le due variabili esplicative $\Pr(D=1|x;\beta)$, $\Pr(O=D=0|x;\beta)$, dipendenti dal parametro ignoto β , e (ii) le due probabilità $p(RIC2|D=1)$, $p(RIC2|O=D=0)$ compaiono nella regressione come coefficienti delle due esplicative dette. Ne segue che sostituendo all'ignoto β la sua stima consistente ottenuta massimizzando la [31] e stimando la regressione [32] si ottengono stime consistenti di $p(RIC2|D=1)$, $p(RIC2|O=D=0)$.

In modo analogo si ottengono stime delle probabilità $p(RIC3|D=1)$, $p(RIC3|O=D=0)$, $p(RIC4|D=1)$, $p(RIC4|O=D=0)$, $p(NORIC1|D=1)$, $p(NORIC1|O=D=0)$, e, per complemento a 1, di $p(RIC1|D=1)$, $p(NORIC2|O=D=0)$.

L'interpretazione di tale stimatore è piuttosto semplice. Se le condizioni fossero note, $p(RIC2|D=1) = E\{I(R=RIC2)|D=1\}$ si otterrebbe regredendo $I(R=RIC2)$ su $I(D=1)$, vale a dire rapportando il numero di disoccupati per i quali $R=RIC2$ al numero totale di disoccupati (e analogamente si procederebbe per ottenere le altre probabilità). Essendo ignote le condizioni di certe unità campionarie (in particolare, essendo ignoti sia il numero di disoccupati per i quali $R=RIC2$ che il numero totale di disoccupati), le stime di β ottenute dalla [31] sono utilizzate per stimare la condizione di tali unità, mediante le probabilità $\Pr(D=1|x;\beta)$, $\Pr(O=D=0|x;\beta)$ (oltre che le analoghe probabilità per le unità delle quali la condizione è nota). La regressione di $I(R=RIC2)$ sulle condizioni stimate fornisce i parametri cercati.

Riassumendo, il metodo proposto fornisce informazioni sulla condizione di coloro che appartengono all'"area grigia" tra disoccupazione e inattività nel modo seguente:

- (a) Utilizzando restrizioni a priori che, almeno al momento, non sono in discussione in quanto comuni a tutti i criteri di classificazione in competizione ($OCC \Leftrightarrow$ occupazione, $RIC1 \Rightarrow$ disoccupazione, $NORIC2 \Rightarrow$ inattività), si ottengono stime delle probabilità di appartenenza alle varie condizioni.
- (b) Le stime delle probabilità di appartenenza alle varie condizioni sono utilizzate per "predire" la condizione di coloro circa la cui condizione non vi è accordo ($RIC2$, $RIC3$, $RIC4$, $NORIC1$). Per il modo in cui sono ottenute, tali condizioni "previste" rappresentano le condizioni più plausibili alla luce delle restrizioni a priori (e dell'evidenza empirica).

- (c) Sulla base delle condizioni "previste", è possibile stabilire quale criterio di classificazione tra quelli in competizione è più coerente con le restrizioni a priori condivise da tutti i criteri. In questo senso, il metodo può essere visto come uno strumento per accertare il grado di coerenza interna dei vari criteri di classificazione.

4. I risultati di una verifica empirica

Retto e Trivellato (1993) specificano e stimano un modello della partecipazione al lavoro e della disoccupazione ammettendo che vi sia incertezza nella reale condizione di una parte degli individui inclusi nel campione. I dati utilizzati vengono dall'indagine trimestrale sulle forze di lavoro di aprile 1984. Il campione è costituito da donne sposate, residenti nel Veneto o in Lombardia, di età non superiore a 60 anni, il cui marito sia di età non superiore a 65 anni.

Le probabilità di appartenenza alle tre condizioni di occupazione, disoccupazione ed inattività risultano da una versione stilizzata del modello *Double-Hurdle* dell'offerta di lavoro proposto da Blundell, Ham e Meghir (1987). Il modello è costituito da due equazioni. La prima fornisce le ore di lavoro desiderate H^* , come combinazione lineare delle variabili osservabili x e della variabile non osservabile ϵ . La seconda equazione determina, mediante i valori assunti dalla variabile latente E^* , quali individui tra coloro che desiderano lavorare ($H^* > 0$) otterranno una occupazione e quali invece rimarranno disoccupati; E^* risulta dalla combinazione lineare di certe variabili osservabili z e della non osservabile v . Il modello che ne risulta è il seguente:

$$\begin{aligned} H^* &= x' \beta + \epsilon \\ E^* &= z' \gamma + v. \end{aligned} \quad [33]$$

Le variabili indicatrici O e D definite in [1] e [2] sono legate alle due variabili latenti H^* ed E^* nel modo seguente:

$$O = \begin{cases} 1 & \text{se } H^* = x' \beta + \epsilon > 0 \text{ e } E^* = z' \gamma + v > 0 \\ 0 & \text{altrimenti} \end{cases} \quad [34]$$

$$D = \begin{cases} 1 & \text{se } H^* = x' \beta + \epsilon > 0 \text{ e } E^* = z' \gamma + v \leq 0 \\ 0 & \text{altrimenti} \end{cases}$$

Le probabilità di essere occupato, disoccupato ed inattivo, condizionatamente alle caratteristiche x e z risultano quindi pari, rispettivamente, a:

$$\begin{aligned} \Pr(O=1|x,z) &= \Pr(H^* > 0, E^* > 0|x,z), \\ \Pr(D=1|x,z) &= \Pr(H^* > 0, E^* \leq 0|x,z), \\ \Pr(O=D=0|x,z) &= \Pr(H^* \leq 0|x). \end{aligned} \quad [35]$$

Nell'applicazione, le variabili incluse tra le x e le z sono elencate nella Tabella 1. Inoltre, le caratteristiche non osservabili (ϵ , v) sono assunte essere determinazioni di una normale bivariata di media nulla e matrice di covarianza Σ , non scalare. Il modello che ne risulta è quindi un Probit bivariato.

Una semplice interpretazione del modello nel contesto italiano è la seguente. I saggi salariali di mercato w sono determinati istituzionalmente. I potenziali lavoratori confrontano il loro saggio salariale di riserva w_r con w e partecipano al mercato del lavoro se e solo se $w_r < w$; in tale caso $H^* > 0$. I datori di lavoro stabiliscono il massimo saggio salariale w_m che sono disponibili a pagare al potenziale lavoratore, sulla base della sua produttività: tale lavoratore trova una occupazione se e solo se $w_m > w$. In tale caso $E^* > 0$. A parità di altre condizioni, un aumento di w produce un aumento del tasso di attività (aumenta il numero di individui per i quali vale $w_r < w$), mentre gli effetti sul tasso di disoccupazione sono incerti: il tasso di disoccupazione aumenta se tra i nuovi attivi, vale a dire coloro che sono indotti alla partecipazione dall'aumento di w , il numero di individui che trovano lavoro è sufficientemente piccolo rispetto al numero di nuovi attivi che non trovano lavoro e di già occupati che perdono il lavoro a seguito dell'aumento di w .

A tale modello della partecipazione al lavoro e della disoccupazione, è sovrapposto un modello di misura che collega i tre stati di occupazione, disoccupazione e inattività, in generale non osservabili, all'indicatore fallibile R introdotto nella sezione precedente, mediante la relazione [26]. Differenti insiemi di restrizioni imposte a priori su tale relazione danno luogo a specificazioni diverse del modello.

Nella Tabella 1 sono presentati i risultati della stima di tre distinte specificazioni del modello, denotate rispettivamente (a), (b) e (c). Per le ragioni dette nella sezione precedente, le tre specificazioni poggiano sulla comune assunzione che la condizione di occupazione sia osservata senza errore. Le prime due specificazioni corrispondono alla stima standard del Probit bivariato, sotto due diversi assunti di perfetta osservabilità degli stati, mentre la terza specificazione ammette indicatori fallibili. Più precisamente:

- (a) Nella specificazione (a) i disoccupati (e per residuo gli inattivi) sono definiti in accordo ad una interpretazione rigida dei criteri ILO: sono classificati tali tutti e solo coloro per i quali vale $R=RIC1$ (cioè a dire, coloro che dichiarano di aver svolto concrete azioni di ricerca negli ultimi trenta giorni precedenti l'intervista).
- (b) Nella specificazione (b), all'estremo opposto, si adotta una definizione della disoccupazione assai lasca: sono definiti disoccupati tutti coloro che dichiarano di cercare un lavoro ($RIC1, RIC2, RIC3, RIC4$) e i lavoratori scoraggiati ($NORIC1$).
- (c) La specificazione (c) introduce alcuni elementi di incertezza sulla reale condizione, dati dalla specificazione [29] del modello di misura. Come già detto, con essa si assume che tutti coloro per i quali $R=RIC1$ siano disoccupati, tutti coloro per i quali vale $R=NORIC2$ siano inattivi. Quanto ai rimanenti non occupati ($R=RIC2, RIC3, RIC4, NORIC1$), nulla si assume riguardo la loro condizione (eccetto l'assunto preliminare che non sono occupati).

Si noti che le specificazioni (a) e (b) sono annidate nella (c). Quindi, da un lato le restrizioni che è necessario imporre sulla (c) per ottenerle sono sottoponibili a verifica d'ipotesi, dall'altro l'erronea imposizione di tali restrizioni dà luogo, in generale, a stime inconsistenti dei parametri liberi.

I risultati documentati nella Tabella 1, e alcune loro implicazioni documentate nella Tabella 2, si prestano a numerosi commenti, sia per quel che riguarda i parametri del modello *Double-Hurdle*, e quindi le caratteristiche dei tassi specifici di attività e disoccupazione, sia per quel che riguarda i parametri del modello di misura.

Tabella 1 – Stime di massima verosimiglianza dei parametri del modello di partecipazione al lavoro/disoccupazione, in tre diverse specificazioni: (a) classificazione a priori ILO rigida; (b) classificazione a priori con disoccupazione lasca; (c) modello con stati latenti e indicatori fallibili (n=2480; standard errors tra parentesi)

		Specificaz. (a)	Specificaz. (b)	Specificaz. (c)
Equazione di partecipazione /non-partecipazione				
Intercetta		-.5929 (.5210)	.4981 (.6468)	.3942 (.6106)
Figlio più giov.	0-3	-.2791 (.1278)	-.4335 (.1462)	-.4624 (.1432)
“ “ “	4-6	-.2312 (.1247)	-.3337 (.1388)	-.3572 (.1369)
“ “ “	7-11	-.1889 (.1073)	-.2412 (.1133)	-.2724 (.1129)
“ “ “	12-19	-.0748 (.0997)	-.0662 (.1024)	-.0972 (.1019)
Numero di figli		-.1640 (.0389)	-.2096 (.0389)	-.1970 (.0388)
Età		.1341 (.0345)	.0977 (.0345)	.1128 (.0352)
Età ²		-.0019 (.0004)	-.0016 (.0004)	-.0018 (.0004)
Almeno licenza media		.3316 (.0769)	.3300 (.0844)	.3430 (.0808)
Almeno diploma sc. sup.		.2592 (.1009)	.3259 (.1138)	.3238 (.1104)
Laurea		.6836 (.2104)	.7658 (.2383)	.7742 (.2339)
Età del marito		-.0678 (.0305)	-.0674 (.0349)	-.0743 (.0319)
Età del marito ²		.0006 (.0003)	.0006 (.0004)	.0007 (.0003)
Al.no licenza media mar.		.0107 (.0593)	.0734 (.0773)	.0413 (.0636)
Al.no dipl. sc.sup. mar.		.0606 (.0782)	-.1645 (.1050)	-.1169 (.0882)
Laurea marito		-.0606 (.1591)	-.1639 (.1876)	-.2084 (.1730)
Equazione di occupazione/disoccupazione				
Intercetta		1.1506 (.4107)	1.5208 (.1547)	1.2139 (.2641)
Figlio più giov.	0-3	-.1874 (.1444)	-.1137 (.1082)	-.1992 (.1339)
“ “ “	4-6	-.1951 (.1440)	-.1322 (.1067)	-.2099 (.1299)
“ “ “	7-11	.0035 (.1375)	-.0358 (.0987)	-.0568 (.1230)
“ “ “	12-19	.1834 (.1339)	-.0156 (.0904)	.0324 (.1173)
Numero di figli		-.1138 (.0658)	.0512 (.0520)	-.0441 (.0678)
30 ≤ Età < 41		.2238 (.0998)	.2093 (.0764)	.1970 (.0942)
41 ≤ Età < 51		.1877 (.1572)	.3044 (.1021)	.1777 (.1553)
Età ≥ 51		.2461 (.2944)	.4694 (.1615)	.3308 (.2833)
Almeno licenza media		.3986 (.0947)	.2201 (.0879)	.3643 (.0930)
Almeno diploma sc.sup.		.2050 (.1152)	.1227 (.0855)	.1933 (.1006)
Laurea		.4866 (.2651)	.3380 (.2010)	.4619 (.2167)
Tasso di disocc. prov		-.0260 (.0163)	-.0312 (.0128)	-.0388 (.0142)
Coeff correlazione		.5167 (.2599)	-.3550 (.2533)	.3346 (.2582)
Modello di misura				
Pr(RIC2(D(1)				.3389 (.0312)
Pr(RIC3(D(1)				.0845 (.0181)
Pr(RIC4(D(1)				.0660 (.0177)
Pr(NORIC1(D(1)				.0086 (.0162)
Pr(RIC2(O(D(0)				2.9E-13 (.0017)
Pr(RIC3(O(D(0)				.0009 (.0006)
Pr(RIC4(O(D(0)				.0003 (.0006)
Pr(NORIC1(O(D(0)				.0092 (.0011)

Nota: Standard errors tra parentesi

Tabella 2 – Distribuzione delle unità campionarie secondo le modalità dell'indicatore fallibile, e classificazioni secondo la condizione risultanti da criteri stabiliti a priori e dalle stime del modello

INDICATORI	Frequenze (a)	Classificazioni a priori dei disocc.		Disoccupati implicati dal modello stimato	
		ILO rigida	disoccup. lasca	(b)	(c)
		OCC	710	-	-
RIC1	243	243	243	243 (d)	243 (d)
RIC2	164	-	164	164	164
RIC3	52	-	52	41	49
RIC4	36	-	36	32	36
NORIC1	112	-	112	4	0
NORIC2	1,163	-	-	0 (d)	0 (d)
Tasso di partecipazione		37.97	39.86	39.22	39.36
Tasso di disoccupazione		3.31	7.88	6.38	6.48

(a) Il campione utilizzato è stato estratto dal file originale campionando con probabilità .1 le unità per le quali R=OCC,NORIC2, e con probabilità 1 le unità per le quali R=RIC1, RIC2, RIC3, RIC4, NORIC1

(b) Ottenuta minimizzando il quadrato dell'errore di stima della condizione

(c) Ottenuta minimizzando il valore assoluto dell'errore di stima della condizione

(d) Questi valori risultano dalle restrizioni imposte per l'identificabilità del modello

- (1) Innanzitutto vi è evidenza che lo spostamento dell'insieme di non occupati corrispondenti a R=RIC2, RIC3, RIC4, NORIC1 dall'insieme degli inattivi (specificazione [a]) all'insieme dei disoccupati (specificazione [b]) modifica in modo sensibile le stime dei parametri del modello *Double-Hurdle*. Vi sono parecchi parametri che risultano significativi in una specificazione e non nell'altra; i valori assoluti delle stime si modificano in modo a volte tutt'altro che trascurabile. Vi è da notare che nel passaggio dalla specificazione (a) alla (b), il valore assoluto di molti dei parametri dell'equazione di partecipazione aumenta. Alla luce della discussione svolta nella sez. 2 (si vedano i commenti alla [8]), questa evidenza indica che la specificazione (b) è più prossima della (a) alla corretta specificazione (ma non dice alcunché circa l'eventuale presenza di errori nella specificazione [b]).
- (2) Quanto alla specificazione (c), dal confronto con i risultati ottenuti in (a) e (b) si ricava che queste ultime attenuano gli effetti delle variabili esplicative sulle risposte: con poche eccezioni, i due sottoinsiemi di parametri che risultano significativi nelle specificazioni (a) e (b) sono sottoinsiemi dell'insieme dei parametri che risultano significativi nella specificazione (c); in generale i valori assoluti delle stime in (a) e (b) sono inferiori a quelli in (c).
- (3) In accordo con quanto osservato alla conclusione del punto (1), i valori assoluti delle stime della specificazione (c) sono molto più prossimi a quelle della specificazione (b) che non a quelle della specificazione (a).

Le stime dei parametri del modello *Double-Hurdle* sono ragionevoli. L'età e il numero dei figli, l'età della donna e la sua istruzione influenzano la decisione di partecipazione nel modo atteso; l'età del marito ha un ruolo autonomo, mentre la sua istruzione risulta non significativa.

Quanto alla probabilità di occupazione, le variabili che contano sono l'età, con un effetto positivo (la classe omessa è "età < 30"); l'istruzione, con un effetto positivo (la classe omessa corrisponde alla licenza elementare [al più]); il tasso di disoccupazione provinciale, con un effetto negativo.

Il coefficiente di correlazione tra le non osservabili nelle due equazioni risulta statisticamente nullo.

(4) Quanto al modello di misura, i risultati sono univoci. La probabilità che un inattivo presenti $R=RIC2, RIC3, RIC4$ e che un disoccupato presenti $R=NORIC1$ è pari a zero agli usuali livelli di significatività.

Ciò equivale a dire che i disoccupati per i quali $R=RIC1$ costituiscono circa il 50% del totale dei disoccupati implicati dalle stime ottenute; il rimanente 50% è costituito da individui per i quali $R=RIC2, RIC3, RIC4$. Al contrario, stando alle stime ottenute, gli individui per i quali $R=NORIC1$ sono da includere tra gli inattivi.

Nella Tabella 2 sono presentate le conseguenze di quanto appena esposto. Utilizzando due diversi modi di stimare il vero e ignoto stato, condizionatamente a z, x ed R (in un caso si minimizza il quadrato dell'errore di stima, nell'altro il valore assoluto dell'errore di stima), si è ottenuta una stima dell'ammontare degli aggregati dei disoccupati e degli inattivi. Ne risulta che, con minime variazioni tra i due metodi di stima dello stato, quasi tutti gli individui per i quali $R=RIC2, RIC3, RIC4$ sono inclusi tra i disoccupati e quasi tutti gli individui per i quali $R=NORIC1$ sono inclusi tra gli inattivi.

Le conseguenze per i tassi di attività e di disoccupazione sono di rilievo. Passando dalla definizione più rigida a quella più lasca dell'aggregato dei disoccupati, il tasso di attività e, soprattutto, il tasso di disoccupazione si modificano in modo sostanziale (rispettivamente, dal 37.97 al 39.86 e dal 3.31 al 7.88%). Sulla base dei risultati ottenuti col modello a stati latenti e indicatori fallibili, poi, i due tassi sono molto più prossimi agli estremi superiori che agli estremi inferiori: rispettivamente, leggermente superiori al 39% e al 6%.

Formulato l'ovvio invito alla cautela nel trarre conclusioni, suggerito dal fatto che i risultati ottenuti sono circoscritti nel tempo, nello spazio e ad un particolare sottoinsieme della popolazione, vale infine la pena di notare che tali risultati sono in completo accordo con il criterio di classificazione dei disoccupati adottato dall'Istat fino a pochi anni fa, abbandonato in favore dell'attuale (che corrisponde a quello che abbiamo denominato 'ILO rigido') per uniformarsi ai criteri vigenti nell'Unione Europea.

Ci sembra che le linee di attacco alla validità di tali risultati possano essere essenzialmente due. Come illustrato alla fine della sez. 3, il metodo adottato consiste nel 'prevedere' la condizione di coloro circa la cui condizione *non vi è accordo* tra i vari criteri di classificazione, mediante un modello stimato poggiando sulle restrizioni a priori sulle quali *vi è accordo* tra i vari criteri di classificazione. La stima del modello è possibile grazie al fatto che le restrizioni a priori sulle quali *vi è accordo* sono sufficienti per l'identificabilità. Ne segue che:

(a) La validità dei risultati dipende dalla capacità del modello di prevedere la condizione di individui non inclusi nel campione utilizzato per la stima del modello stesso. I risultati ottenuti (e non riportati in questa sede) mostrano che il modello si comporta bene *nel campione*.

Ne segue che per difendere la tesi che il modello è incapace di prevedere *fuori dal campione* è necessario difendere anche la tesi che la condizione di coloro per i quali $R=RIC2, RIC3, RIC4, NORIC1$ è determinata da un processo diverso da quello che determina la condizione di coloro per i quali $R=OCC, RIC1, NORIC2$.

(b) Se anche gli assunti in comune ai vari criteri fossero errati, i risultati ottenuti, in generale, perderebbero la loro validità.

5. Prospettive per nuove analisi empiriche

Alla luce dei risultati analitici e delle evidenze empiriche presentate sinora, riteniamo di qualche interesse segnalare brevemente alcune prospettive per nuove analisi empiriche,

che consentano di approfondire le questioni di classificazione/comportamento della popolazione rispetto al lavoro. Tre sono le linee di ricerca che ci paiono promettenti.

Innanzitutto, è utile replicare analisi del tipo di quelle riassunte nella sez. 4 in altri contesti: per altre sottopopolazioni di interesse (ad esempio, per la sottopopolazione dei giovani) e/o per ambiti territoriali differenti (ad esempio, per il Mezzogiorno o per l'intero Paese) e/o con riferimento a tempi diversi - vuoi per disporre di evidenze empiriche più aggiornate vuoi per vagliare l'eventuale reattività delle stime dei parametri (di comportamento e di risposta) a differenti fasi cicliche. Un insieme di analisi di tal fatta consentirebbe, plausibilmente, di individuare i tratti di stabilità (o, all'opposto, di variabilità) nei risultati, in ispecie in quelli attinenti al modello di misura, e in definitiva di giudicare della 'robustezza' delle indicazioni in tema di classificazione delle 'aree grigie' in una delle tre condizioni.

In secondo luogo, è importante cercare di utilizzare al meglio le opportunità offerte dalla nuova versione dell'indagine sulle forze di lavoro, adottata dall'Istat dall'ottobre 1992. Non ci riferiamo tanto all'ovvio adattamento ai nuovi confini definitivi per forze di lavoro e per occupati (tra i quali spicca l'elevamento a 15 anni del limite inferiore per definire la popolazione in età lavorativa; per maggiori dettagli, vedi Casavola e Sestito, 1994, e Trivellato, 1997), quanto ad alcune possibilità specificamente rilevanti per gettare maggiore luce sulla questione della classificazione delle 'aree grigie'.

Ciò vale senz'altro per l'area dei lavoratori scoraggiati. Al riguardo, la nuova indagine ne consente una caratterizzazione più precisa - o, se si vuole, più articolata. Infatti, accanto alle usuali informazioni (i) sull'assenza di ore di lavoro nella settimana di riferimento e (ii) sulla mancata ricerca di lavoro "perché ritiene di non riuscire a trovare un lavoro" (informazioni che grosso modo corrispondono a quelle che portano a collocare un individuo nella modalità NORIC1), (iii) per chi non cerca lavoro (a meno degli occupati o sospesi dal lavoro o in Cassa Integrazione Guadagni) la nuova indagine rileva anche se "sarebbe comunque disposto a lavorare" o meno. Le opportunità che ne vengono sono palesi:

- (a) per un verso, si può circoscrivere con maggior precisione l'insieme dei lavoratori scoraggiati, includendo in NORIC1 soltanto coloro che soddisfano all'ulteriore criterio di essere disponibili a lavorare (in ciò conformandosi con la prassi più diffusa: vedi OECD, 1987);
- (b) alternativamente, si può articolare maggiormente l'insieme R degli indicatori, bipartendo NORIC1 (ed eventualmente NORIC2) sulla base della disponibilità o meno a lavorare, e si possono poi valutare i risultati delle stime del modello a stati latenti e indicatori fallibili.

Ci si può interrogare poi sulla possibilità di utilizzare un ulteriore insieme di domande poste a quanti hanno dichiarato di essere in cerca di lavoro e di aver effettuato azioni di ricerca, domande che, in sostanza, mirano a cogliere tratti di *choosiness* nell'offerta di lavoro ("Qual è il tipo di occupazione prevalentemente cercato?"; "Con quale orario vorrebbe lavorare?"; "In quale luogo sarebbe disposto a lavorare?"; "A quale livello di retribuzione minimo mensile sarebbe disposto a lavorare avendo presenti il tipo, l'orario e il luogo di lavoro indicati nei precedenti quesiti?"). L'utilizzazione di questi quesiti, in relazione alle tematiche di comportamento/classificazione dell'offerta di lavoro qui affrontate, appare peraltro piuttosto problematica, per un duplice ordine di considerazioni: da un lato, l'impianto lessicografico delle domande - e delle correlate modalità di risposta - rende difficile (forse, addirittura preclude) l'identificazione di preferenze utili per l'analisi di comportamenti; dall'altro lato, l'evidenziazione di un sottoinsieme di non occupati in cerca contraddistinti da una marcata *choosiness* non è facilmente compatibile con il mantenimento di uno spazio degli stati latenti circoscritto a tre condizioni - occupato, disoccupato, inattivo.

In terzo luogo, e per finire, vanno considerate le possibilità di analisi che vengono dalla *European Community Household Panel* (ECHP). Restando alle opportunità strettamente pertinenti all'approccio adottato in questo lavoro, e trascurando quindi gli aspetti di maggior interesse dell'ECHP - quelli longitudinali -, due sono i tratti meritevoli di attenzione: (i) il fatto che l'indagine 'mimi' piuttosto accuratamente il tipico insieme di quesiti adottati nelle indagini sulle forze di lavoro per rilevare la condizione (ma con la differenza che parecchi quesiti sulla disponibilità a lavorare e sul lavoro cercato attengono ad "un lavoro *adatto*": qualificazione questa potenzialmente non trascurabile, e della quale è comunque pressoché impossibile cogliere le implicazioni); (ii) la disponibilità di un insieme estremamente più ricco di informazioni su variabili individuali e familiari, che possono permettere un più penetrante trattamento dell'eterogeneità.

References

- BLUNDELL R., J. HAM E C. MEGHIR (1987), "Unemployment and female labour supply", *The Economic Journal*, 97 (Conference), pp. 44-64.
- CASAVOLA P. E P. SESTITO (1994), "L'indagine Istat sulle forze di lavoro", *Lavoro e Relazioni Industriali*, 1, pp. 179-195.
- FLAUM P. (1984), "Discouraged workers: how strong are their links to the job market?", *Monthly Labor Review*, (8), pp. 8-11.
- GREEN A.E. (1995), "A comparison of alternative measures of unemployment", *Environment and Planning A*, 27, pp. 535-556.
- HAUSMAN J.A. E F. M. SCOTT MORTON (1993), "Misclassification of a dependent variable in a qualitative response setting", Cambridge (MA), MIT and NBER, mimeo.
- HUSSMANN R., F. MERHAN E S.M. VERMA (1990), *Surveys of economically active population, employment, unemployment and underemployment: an ILO manual on concepts and definitions*, Geneva, ILO.
- MALINVAUD E. (1986), *Sur les statistiques de l'emploi et du chômage*, Paris, La Documentation Française.
- MANSKI C. (1985), "Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator", *Journal of Econometrics*, 27, pp. 313-333.
- MANSKI C. (1988), "Identification of binary response models", *Journal of the American Statistical Association*, 83 (403), pp. 729-738.
- OECD (1987), *Employment outlook*, Parigi.
- RETTORE E., N. TORELLI E U. TRIVELLATO (1990), "Unemployment and search for work: exploratory analyses of labour market attachment using CPS-type data", *Labour*, 4 (3), pp. 161-190.
- RETTORE E., N. TORELLI E U. TRIVELLATO (1993), "La misura della disoccupazione mediante indagini campionarie sulle famiglie", in U. Trivellato (a cura di), *Norme e metodi sul mercato del lavoro*, vol. 1, Roma, CNEL, 1993, pp. 183-215.
- RETTORE E. E U. TRIVELLATO (1993), "A Double-Hurdle labour supply model with fallible indicators of labour force state", *Statistica*, 53 (3), pp. 341-367.
- SESTITO P. (1989), *Misurazione dell'offerta di lavoro e tasso di disoccupazione*, Temi di discussione n. 132, Roma, Banca d'Italia.
- SHISKIN J. (1976), "Employment and unemployment: The doughnut or the hole?", *Monthly Labor Review*, 99 (Febr.), pp. 3-10.
- TRIVELLATO U. (1993), "La conta dei disoccupati: evidenze ed interrogativi dalle nuove stime della rilevazione sulle forze di lavoro", *Rassegna di Statistiche del Lavoro*, (1), pp. 86-93.
- TRIVELLATO U. (1997), "Statistiche armonizzate su occupazione e disoccupazione", in U. Trivellato (a cura di), *Norme e metodi sul mercato del lavoro*, vol. IV, Roma, CNEL, in corso di pubblicazione.

SELECTION AND MEASUREMENT ERROR PROBLEMS IN THE EUROPEAN COMMUNITY HOUSEHOLD PANEL SURVEY

Franco Peracchi

Università di Roma Tor Vergata, Italy

1. Introduction

The European Community Household Panel (ECHP) is a standardized multi-purpose annual longitudinal survey carried out at the level of the European Union (EU). It is centrally designed and coordinated by the Statistical Office of the European Communities (Eurostat), and covers demographics, labor force behavior, income, health, education and training, housing, migration, etc.

The ECHP project was launched "to yield information on the social dimension of the European Union. It will be used in the formulation and monitoring of specific policies by the Commission in a variety of areas and will serve the need of research both inside and outside the Commission. It should help in the revision of the Treaty in 1996--97" [Eurostat (1996)].

The objective of the ECHP is to represent the population of the EU at the level of households and individuals. It also aims at being both cross-sectionally and longitudinally representative, with changes in the population over time reflected by the continuous evolution of the sample. In this respect, the ECHP is patterned after the Panel Study of Income Dynamics (PSID) conducted by the Survey Research Center at the University of Michigan.¹

The ECHP was initially conceived as a 3-wave panel. The first wave, conducted in 1994, includes all current members of the EU except Austria, Finland and Sweden. The second one, conducted in 1995, also includes Austria. The third one, conducted in 1996, further includes Finland. Recently, funding for 3 additional waves has been approved by the European Commission.

The purpose of this paper is to discuss some of the issues that an applied economist potentially interested in the ECHP may face when trying to deal with these data. First, we describe the survey in some detail. Section 2 describes its organization (coverage, sampling design, follow-up rules, questionnaire and fieldwork). Section 3 describes the con-

¹ The PSID is a large and widely used longitudinal survey of the U.S. population and the families in which they reside. The study is designed to provide a continuous representation of families and of individuals, and contains detailed information on socio-economic and demographic variables. It has been ongoing since 1968. Data are collected annually and the publicly released data files, now available on CD-ROM, contain the full span of the information collected over the course of the study. For an introduction to the PSID, see Hill (1992).

I thank seminar participants at CEMFI, ISTAT, New York University and Texas A&M University for their comments. I also thank Manuel Arellano, Samuel Bentolila, Juan Dolado, Chris Flinn, Andrea Regoli, Paloma Seoane and Finis Welch for helpful discussions. Financial support from CICYT and Fundación BBV is gratefully acknowledged.

e-mail: franco.peracchi@uniroma2.it

tent of the survey and the basic definitions adopted, focusing on demographics, economic activity status, income, and education and training. Section 4 describes how the databases are organized and may be accessed. The substantive problems of measurement errors and selection in the ECHP are addressed in Sections 5-8. Specifically, Section 5 discusses some problems related to measurement errors. Section 6 discusses the issue of nonresponse, distinguishing between unit nonresponse, item nonresponse, sample attrition and new entry. Section 7 discusses the weighting procedures with which Eurostat tries to compensate for differential nonresponse and attrition rates. Section 8 discusses the imputation methods adopted by Eurostat in order to deal with item nonresponse. Finally, Section 9 offers some conclusions.

2. Organization of the survey

The ECHP is carried out by National Data Collection Units (NDU), with Eurostat providing centralized support and coordination. The NDUs are the National Statistical Institutes in eight of the countries, namely Finland, France, Germany, Greece, Italy, Netherlands, Portugal and Spain, and other public bodies or private organizations in the remaining countries. In Belgium and the Netherlands, the ECHP is linked to already existing national panels.

The NDUs are responsible for sample selection, adaptation of the questionnaire, fieldwork, basic data processing and editing, and initial weighting of the data. Although Eurostat sets general guidelines to ensure comparability of survey results, the NDUs largely rely on their normal rules and routines. Unfortunately, national reporting on the survey organization is lacking, making it somewhat difficult to assess data quality.

2.1. Coverage and sampling design

The target population includes all private households throughout the national territory of each country.² Although the same criteria are adopted to define a household, namely "sharing the same dwelling" and "common living arrangements", the precise definition is not completely standardized across countries.

All national samples are selected through probability sampling. Sampling frames and sampling procedures are not standardized across countries, however, and each NDU relies on its own methods. In most countries, the sampling frame is either the population register or a master-sample created from the latest census of the population. Netherlands and the UK use the postal address register, whereas Ireland uses the electoral roll. The choice of sampling frame implies noncoverage of small portions of the target population, such as households recently arrived in a country (Ireland, Italy) or nonresidents unable to speak the national language (Greece, Netherlands).

The most common sampling procedure is two-stage sampling, with geographical areas (usually the municipalities) as primary sampling units and households or addresses as secondary sampling units. Denmark and Luxembourg rely instead on simple random sampling of households, whereas three-stage sampling is adopted in Portugal and part of France.

² Some countries exclude small parts of their national territory. For example, France excludes overseas departments and territory, Italy excludes some small islands, Spain excludes Ceuta and Melilla.

Originally, in 1992, Eurostat planned a minimum sample size of 1000 household per country and an enlarged sample size of 2000 households for two groups of countries: the four largest ones (Germany, France, UK and Italy) and the four economically least developed (Spain, Portugal, Greece and Ireland). The planned sample size was later expanded by a factor of four to include about 83,000 households in the 12 countries. The achieved sample size in wave 1 is nearly 61,000 households [Table 1].

Table 1 – Population in 1993 (thousand persons) and sample size (planned and achieved number of households) in wave 1, by country

COUNTRY	Population		Sample			
	1993	%	Planned	%	Achieved	%
Belgium	10084	2.9	4912	5.9	4189	6.9
Denmark	5189	1.5	5580	6.7	3482	5.7
France	57655	16.6	10915	13.1	7344	12.1
Germany	81179	23.3	10572	12.7	5054	8.3
Greece	10380	3.0	6131	7.4	5523	9.1
Ireland	3563	1.0	7252	8.7	4048	6.7
Italy	57049	16.4	7841	9.4	7115	11.7
Luxembourg	398	.1	2485	3.0	1011	1.7
Netherlands	15290	4.4	5926	7.1	5187	8.5
Portuga	19876	2.8	5492	6.6	4881	8.0
Spain	39083	11.2	7992	9.6	7206	11.8
UK	58191	16.7	8104	9.7	5779	9.5
Total	347937	100.0	83202	100.0	60819	100.0

Within each household, the ECHP distinguishes between sample and nonsample persons. A sample person is anybody in the initial (wave 1) sample who is still alive, plus any child born, subsequently to wave 1, to a sample women. A nonsample person is any other person residing in a private household containing one or more sample persons. Sample and nonsample persons are eligible for personal interview if they are "adult", that is, 16 or older on January 1 of the survey year.

Households and persons are identified through fixed and unique identification numbers. Personal identification numbers are made up of three components: 1) the original fixed identification number of the household where the person is first listed, 2) the split number of the household at the time the person is first listed, 3) a person number within the household, unique across the whole duration of its existence.

2.2. Follow-up rules

The households which are passed on wave $t > 1$ for follow-up are those interviewed in wave $t - 1$ and those not interviewed in wave $t - 1$ because of non-contact, physical incapacity or inability to respond, or refusal which is considered initial or less than final. To the follow-ups are added the newly formed households between the two waves and any additional household added to the sample.

Excluded from follow-up are the households which gave a final refusal in wave $t - 1$ and those which have moved to a country outside the EU or have been entirely institutionalized. These households are only "traced" for possible future follow-up in case they return to the study population. Sample addresses or households found to be ineligible or nonexistent

are dropped from the survey. Starting from wave 3, households not interviewed in two consecutive waves are also dropped.

Turning to persons, sample members are always followed-up until they die or become ineligible. Sample persons who move to a country outside the EU or have been institutionalized are not interviewed but are "traced" for possible future follow-up if and when they return to the study population. Like the PSID, nonsample persons are not followed if they move into a household not containing sample persons.

Enlarging or supplementing the initial ECHP sample is currently being discussed. The reasons are panel attrition, the need of more reliable or more extensive information on particular social groups and particular issues, and the low response rates in wave 1 for some countries.

2.3. Questionnaire

The questionnaire is centrally designed by Eurostat but must be adapted by the NDUs. Adaptation is not a trivial task and has been a source of problems.

The questionnaire consists of i) a household register, mainly for record keeping and control of the sample, ii) a household questionnaire submitted to a "reference person" (usually the household head or the spouse/partner of the head), and iii) a personal questionnaire submitted to all eligible household members, which are therefore all treated symmetrically.³

There is a single respondent for the household interview, who is typically the head of the household or the spouse/partner of the head. In addition, all household members aged 16 or more, present or temporarily absent, are eligible for personal interview.

Although Eurostat recognizes the importance of motivating the respondents, rewards are only "symbolic". Eurostat guideline is: "Let us be nice to respondents but refrain from paying them". For wave 1, the basic instrument has been a letter of information about the objectives of the survey and the requirements asked from the respondents. A few months after the interview, a thank-you letter was generally sent to the respondent households.

Verma (1995a) provides some data on mean duration of the household interview in wave 1 for Denmark and the UK. Mean duration is about 20 minutes in both countries, with over 90 percent of the household interviews requiring not more than 30 minutes. Regoli (1996) reports a similar value for mean duration in Italy, namely 18 minutes. However, whereas in Denmark no interview exceeds 50 minutes, in the UK a small proportion (1.2 percent) took from 50 minutes to a maximum of 8 hours. Turning to the personal interview, Regoli (1996) reports a mean duration of 20 minutes in Italy, which is only slightly higher than mean duration of a household interview.

Subjective and sensitive wording questions are asked often. Examples include the following: "To what extent are housing costs [or debt] a burden to you?", "Thinking of your household's total monthly income, is your household able to make ends meet?", "In your opinion what is the very lowest net monthly income that your household would have to have in order to make ends meet?", "How satisfied are you with your present situation in the following areas: work or main activity, financial situation, housing situation, amount of available leisure time?"

The reference period varies considerably depending on the particular questions being asked. For personal characteristics, it is the day of the interview. For economic activity

³ This is unlike the PSID, where the head (or the spouse/partner of the head) serves as the sole respondent and the most detailed information is collected for the head and, since the late 1970s, for the spouse/partner.

(occupational status, etc.), it is the reference week, defined as the last 7 days before the interview. For absence from work, it is the last 4 working weeks before the interview. For income and income sources, it is the last calendar year. For cash and non-cash social assistance, it is each month in the last calendar year.

The questionnaire makes extensive use of branching and skip patterns, which may create problems if interviewers are not sufficiently well trained.

2.4. Fieldwork

Since part of the information collected refers to the last calendar year, the original plan was to carry out the fieldwork early each year. Fieldwork for wave 1, however, did not conform to this standard. In Belgium, Denmark, Ireland, Italy, Netherlands, and Portugal, it began in Spring 1994 and was completed by October. In Germany, fieldwork began in May for six Länders and in September for the remaining ones except Hamburg, where it only started in January 1995. In France, Greece, Luxembourg, Spain and the UK, fieldwork was entirely conducted in the last quarter of 1994. Lack of sincronization persists in the subsequent waves. Fieldwork for waves 2 and 3 has been carried out in the first part of the year in all countries except France, Greece, Italy and Spain, where the survey has taken place in the second part.

The interviewing method recommended by Eurostat is face-to-face personal interviewing. Other interviewing methods have been used, however, such as self-completed by the respondent, telephone, and proxy (for personal interviews only). Eurostat recommends the use of proxy interviewing as "last resort in a restricted range of cases", such as temporary absence, illness or language problems. This recommendation has not been followed very closely. In Italy, for example, proxy respondents account for 15.3 percent of the individual interviews in wave 1 and 11.9 percent in wave 2 [Regoli (1996)].

In four countries, namely Greece, Portugal, Netherlands and the UK, interviews are carried out, at least partly, using computer assisted personal interviewing (CAPI). All other countries rely instead on the conventional "paper and pencil" method. Eurostat has been rather lukewarm about CAPI, fearing it could restrict the flexibility of interviewing, thereby reducing response rates and full participation of all household members. On the other hand, as stressed for example by Olsen (1992), "the demand of modern behavioral modelling in the social sciences together with the costly and error-prone process of editing data from paper and pencil methods argue for the use of modern, computer-assisted methods of data collection".

Little is known about selection of the interviewers, and the training and instructions that they receive, especially training in order to get valid responses on sensitive items and instructions in case of refusal to respond to certain questions. Although Eurostat recommends monitoring of interviewers through post-interview verification, it is not known to what extent this is done. In the Italian case, a total of 480 interviewers were employed in the first two waves, with an average load of about 16 household interviews each [Regoli (1996)]. Of the total number of interviewers employed, 72.7 percent were common to wave 1 and wave 2, whereas 72.5 percent of the households with completed interviews were interviewed by the same interviewer at both waves.

In order to reduce inter-wave inconsistencies, Eurostat recommends feeding forward a certain amount of information, mainly demographic, from one wave to the next. Responds should verify this information and update it, if necessary. Because of fear of suppressing genuine changes, Eurostat decided not to extend forward feeding to include parts of labor force and income details.

3. Content of the survey

The ECHS contains detailed information on demographic characteristics, main activity status, income, health, education and training, housing, migration, and other variables, at both the household and the personal level. In what follows, I shall focus on the information collected by the Community version of wave 1 questionnaire on demographic characteristics (Section 3.1), economic activity (Section 3.2), income (Section 3.3), and education and training (Section 3.4).

3.1. Demographic characteristics

The household interview provides information on household size, including temporarily absent persons still considered household members, and household dynamics between successive waves of the survey.⁴

The personal interview provides information on a person's month and year of birth, sex, relation to the head of the household, residential status,⁵ reasons for temporary absence or why no longer a household member, own or adopted children, and marital status.

3.2. Economic activity

This information is from the personal interview. Economic activity is broken down into two mutually exclusive categories: 1) "with a job", defined as working in, or temporarily absent from, a job or business normally involving at least 15 hours a week; and 2) "without a job", defined as being in education or training, unemployed, retired, doing housework, looking after children or other persons, in community or military service, otherwise economically inactive.

Those with a job are asked about the type⁶ and the characteristics of the present job,⁷ the normal hours of work per week in the main job, the main reason for working less than 30 hours a week, gross and net normal monthly earnings from the main job, the existence of an employer provided supplementary pension scheme, the existence of employer-provided free or subsidized services or benefits,⁸ the degree of satisfaction with present job,⁹ the number of days of absence from work (for illness or other reasons) in the last four working weeks, additional job or business and its type, hours of work per week on the additional job or business, and whether they are looking for any work, either instead or in addition to present work.

Those without a job are asked whether they have done any work done during the past 7 days, whether they are seeking work and, if not, the main reason for not seeking work.¹⁰

If they have done some work in the past 7 days, they are asked about the characteristics of such work,¹¹ the main reasons for working part-time,¹² and gross and net monthly

⁴ Persons moving out, deaths, persons returned, persons moving in, new births.

⁵ Currently living in the household, absent but still considered household member, past household member.

⁶ Paid employment or apprenticeship (full- or part-time), self-employment, and unpaid work in family enterprise.

⁷ Occupation, job-related education or training received, type and size of business or organization, foreign languages used at work, position in the business or organization.

⁸ Child minding, health care, education and training, sports and leisure, housing.

⁹ In terms of earnings, job security, type of work, number of working hours, working times and conditions, distance to job.

¹⁰ Engaged in other activities (study, training, community or military service, other), housework or looking after children or other persons, retirement, personal illness or disability, believe no suitable work is available, has already found work to start in the future, awaiting outcome of other possibilities, other.

¹¹ Kind of work, principal activity performed, kind of business or organization, weekly hours worked.

¹² Undergoing education or training, housework or looking after children or other persons, personal illness or disability, want but cannot find a full-time job, do not want a full-time job, etc.

earnings received. If they are seeking full-time or part-time work, they are asked how many hours per week they would like to work, the minimum monthly income in order to work such hours, the chances of finding the kind of job they are looking for within a year, whether they would be able to start working within the next 2 weeks if they found a suitable job, whether they receive unemployment benefits, whether they are registered at a public employment office, the job offers accepted or received during last 4 weeks, and the reasons for not accepting the most recent job offer.¹³

Although most of the questions refer to current labor force status, the survey also collects some retrospective information on labor force behavior. First, respondents are asked to report any changes in activity status in the year before the survey. They are also asked to report main activity status in the year before the survey and separately for each month.

Respondents are then asked about age at first job and unemployment experience over the past five years. In addition, those who have a job are asked when the present job began, how it was found, whether they were unemployed directly before the present job, whether they ever worked before, when they stopped working in the previous job, the main reason why they stopped working in the previous job,¹⁴ and whether the current job is better than the previous one.

Those who do not have a job and are not seeking work are asked whether they ever worked in a job involving at least 15 hours per week, when they stopped working in the last such job, the characteristics of last such job and the business or organization, the main reason why they stopped working, distinguishing between job related reasons,¹⁵ sale or closure of own or family business, and personal or other reasons.¹⁶

3.3. Income

Gathering information on household and personal income is one of the main goals of the survey. The following income sources are distinguished: 1) wages or salaries, 2) income from self-employment or farming, 3) pensions, 4) any types of social benefits and grants other than pensions, 5) income from investment, savings or property, and 6) income from other sources. Usually, people are asked to report income amounts both gross and net of tax and contributions to social insurance. Although there is no topcoding, the questionnaire gives respondents some freedom in the way income is reported, such as the specification of ranges instead of actual amounts.

The household respondent is first asked to indicate what income sources are present and which is the largest. The respondent is then asked to report the household's total net income per month (after tax and social insurance contributions) or, if he does not know it, to provide an approximate range.

Details on the components of household income are asked in the personal interviews. Nevertheless, the household respondent is asked to report whether, in the year before the survey, anyone in the household received social assistance payments and non cash-assistance with detail for each month; anyone in the household received income from renting

¹³ Rate of pay, temporary or insecure job, type of work, number of working hours, working times (day time, night time, shifts, etc.), working conditions or environment, distance to job or commuting, could not start the job at the required time, not yet decided but still considering the offer.

¹⁴ Obtained a better or more suitable job, obliged to stop by employer (business closure, early retirement, dismissal), end of contract or temporary job, sale or closure of own or family business, personal or other reasons.

¹⁵ Retirement at normal age, obliged to stop by employer (business closure, early retirement, or dismissal), end of contract or temporary job.

¹⁶ Marriage, child birth or need to look after children, looking after old, sick or disabled persons, partner's job required move to another place, study, national service, own illness or disability, wanted to retire or live off private means, etc.

property with the indication of the amount received net of mortgage, repairs, maintenance and insurance; anyone in the household inherited any property or capital, or received gift or lottery winnings worth 2,000 ECU or more; consuming food from own agriculture or gardening, consuming other goods from own business or do-it-yourself activities resulted in a saving of at least 1,000 ECU in annual household expenditures.

Each person eligible for interview is asked to provide information on three types of income: earnings, income and benefits from sources other than work, and income from capital or investment. Whereas the information on the first two types of income is very detailed, information on income from capital and investment is not. Except for the latter, the survey usually asks the average monthly income amount from each source as well as the number of months this income was received.

Those who undertook any paid work, either as a self-employed or as an employee, are asked about their gross and net "normal" earnings,¹⁷ the total earnings received during the year if their work is irregular, extra payments for overtime work or commissions or tips, other kinds of extra payments,¹⁸ income (pre-tax profits) from self-employment (own business, profession or farm) and from other work (agricultural, secondary or casual job).

Income and benefits from sources other than work exclude housing allowances and social assistance, which are recorded separately in the household questionnaire, but include benefits related to unemployment, job creation or training;¹⁹ survivor's pension or benefits;²⁰ pension or benefit relating to old-age or retirement;²¹ family-related benefits, including maternity and single-parent benefits;²² benefits relating to sickness or invalidity;²³ other benefits or assistance;²⁴ and financial support or maintenance from relatives, friends or other persons.

3.4 Education and training

This information is from the personal interview. Everybody is asked about the highest level of education attained,²⁵ the third level of education completed (if any),²⁶ and the age when this education was completed. Those who completed or undertook vocational education or training are also asked about their level of qualification (e.g. technical college), specific vocational training lasting one year or longer,²⁷ training or apprenticeship lasting less than one year, other vocational training, and the age when this vocational education or training was completed. Everybody is then asked about the degree of satisfaction with the education or training received.

Those who report education or training as their main activity are asked to specify what kind of education or training it is.

Those who have been in any education or training (including any part-time or short courses), at any time in the year before the survey, are asked whether this is still going on or

¹⁷ If more than one job at the same time, all jobs are considered; if different jobs during the year, the longest one is considered.

¹⁸ 13th or 14th salary, holiday pay, profit sharing or bonus, other lump sum payments, company shares, etc.

¹⁹ Unemployment insurance, unemployment assistance, training or retraining allowance, placement or resettling or rehabilitation benefits, etc.

²⁰ Basic and supplementary schemes (public and private), personal schemes, means tested welfare schemes, other widow's benefits, orphan's pension or allowance.

²¹ Basic and supplementary schemes (public and private), personal schemes, means tested welfare schemes, early retirement schemes, etc.

²² Allowances for children, care of invalid dependants, maternity, birth, unmarried mother, deserted wife, etc.

²³ Income maintenance benefits in case of sickness or injury, other sickness benefits, compensation for occupational accidents and diseases, invalidity pension, etc.

²⁴ Scholarships, study grants, etc.

²⁵ Recognized 3rd level, secondary stage of secondary level, first stage of secondary level, still at school.

²⁶ Higher university degree or post-graduate, initial university degree or equivalent, 3rd level of education other than university degree

²⁷ At a vocational school or college, in a working environment, "dual" system including apprenticeship.

is finished, the kind of education or training, whether the course was paid for by employer, whether it was an attendance- or a correspondence-course and, if attendance, whether it was full- or part-time, the overall duration, whether improving skills or job prospects was one of the reasons for taking the course, and the usefulness of the course for that purpose.

4. Data organization

4.1. The databases

The NDUs transmit the micro-data to Eurostat in four files:²⁸

1. the household register or D-file, containing administrative variables,²⁹ sample related variables,³⁰ fieldwork related variables,³¹ and variables related to household composition and dynamics;³²
2. the personal register or R-file, containing administrative variables,³³ personal characteristics variables,³⁴ residence related variables,³⁵ interview related variables,³⁶ and identification link variables;³⁷
3. the household questionnaire or H-file;
4. the personal questionnaire or P-file.

The variables in these files are used by Eurostat to construct a number of additional variables, such as household type, activity status, equivalence scales, total net household income, hourly wage rate, etc.

After deriving the constructed variables, imputing missing values and computing weights, data are organized by Eurostat in three data sets. The first has one record per person in the sample, included the ones that refused to answer or are under age 16, and contains the data from the R-file, the weights and imputed variables. The second data set has one record per household with a completed household interview and contains the data from the H-file, some variable from the D-file, the weights, and household constructed and imputed variables. The third data set has one record per person with a completed personal interview and contains the data from the P-file, some variable from the R-file, the weights and some individual constructed and imputed variables.

The data set are made available in compressed ASCII format. The complete database for 1994 (12 countries) is now ready, the one for 1995 is expected to be ready by June 1997.

²⁸ In the case of Germany, micro-data were not transmitted for 240 households (out of 5,000) who denied permission to such transmission.

²⁹ Wave, country, etc

³⁰ Household identification and split number, region, primary strata, primary sampling unit, sample weights, type of sample unit, etc.

³¹ Sequence number of the household, interviewer, date of interview or final contact or last call-back, number of contacts attempted, final result of the household interview, mode of interview, etc.

³² Total number of household members, persons moving out, deaths, persons returned, persons moving in, new borns, number of persons eligible for individual interview, number of individual interviews completed, etc.

³³ Wave, country, identification number, sequence number of the household, person line number in the household roster, whether the person is a sample member, etc.

³⁴ Month and year of birth, sex, relation to the household head, etc.

³⁵ Residential status, reason why no more a member of the household, reason for temporary absence, etc.

³⁶ Respondent status, status if aged less than 16, final result of the personal interview, mode of interview, line number of the person providing the information (in the case of proxy interview), whether the proxy information was verified by the subject concerned, etc.

³⁷ Household identification number of the respondent's current household, etc.

4.2. Data access

Data are stored on Eurostat secure system. Access to the databases is restricted to a degree that depends on the category of users. Eurostat distinguishes four categories: Eurostat itself, the NDUs, consultants contracted by Eurostat, and other users (the European Commission, consultants contracted by the European Commission, and all other categories of researchers).

The other users may receive tabulations upon request, or go to Eurostat and access the micro-data directly in the secure area. Only Ireland, Portugal and the UK allow unrestricted distribution of their national micro-data. All other countries place more or less severe restrictions on the other users.

5. Data editing

This section reviews briefly the procedures adopted in the ECHP to detect inconsistencies and implausibilities, both at the cross-section and longitudinal level.

As pointed out by Olsen (1992), "while cross-sectional, univariate, descriptive work, such as the tabulation of means and rates as often published in government reports, can tolerate [small] error rates, a complex longitudinal analysis requiring only three variables in each of ten years could lose about 30% of the sample just due to interview error. Missing data can be imputed, but the result can induce spurious transitions and the effect of measurement error on nonlinear methods often resists analytical evaluation".

5.1. Cross-sectional editing

Data are checked and edited by the NDUs using checking programs provided by Eurostat as well as national standard checking rules.³⁸ Regoli (1996) reports some data for the Italian household questionnaires in wave 2. The mean number of corrections per variable is 100.2. No corrections were made in 43 percent of the questionnaires, and no more than five were made in 90 percent of them. The mean number of corrections per questionnaire is 1.7, the maximum is 87. After receiving the datasets, Eurostat's checking is also applied. The NDUs are informed about the results of these checking procedures. In some cases, they are requested to review the data and edit items where possible.

To my knowledge, the ECHP has not been supplemented by reinterview programs to measure the reliability of the responses. An alternative way of estimating data quality is to collect information for an overlapping period in successive interviews [O'Muircheartaigh (1989)]. This could easily be done in the ECHP.

5.2. Longitudinal editing

Eurostat's minimalist approach to forward feeding of information from one wave to another implies that, in addition to cross-sectional editing for each wave, longitudinal editing is also necessary to correct longitudinally inconsistent or implausible responses.

³⁸ For example, the Italian Statistical Office uses both manual and computer-assisted interactive editing. Manual editing is used to control the identification numbers, variables such as birth date and sex, main inconsistencies in the answers, and coding of some information. Interactive editing is used to control fields and skip patterns.

The criteria suggested by Eurostat for longitudinal editing are three. The first is plausibility of the distribution in terms of mean, variance, and outliers. The second is plausibility of covariance relationships. In either case, plausibility has to be based on information derived from external sources. The third is consistency, that is, "longitudinally edited variables should be consistent with other variables".

Although I have little disagreement with these general criteria, their practical implementation is far from clear. In the case of income, Eurostat recommends selecting the interviewees with respect to a suspect change if the change in income with respect to the previous wave exceeds 50 percent and the income amount in both years is above a certain threshold. For other variables, such as age, education level, etc., obvious checks based on the fact that these variables are not decreasing may easily be incorporated into CAPI.

6. Nonresponse

Nonresponse is defined as failure to measure some of the units in the selected sample. It is useful to distinguish between five types of nonresponse.

1. Noncoverage: it refers to failure to include certain units of the population in the initial sample.
2. Unit (or complete) nonresponse: it occurs when no information is available from a sample unit. It may be caused by refusal or inability of the unit to participate or by a missing questionnaire.
3. Item nonresponse: it occurs when a unit refuses to answer, or fails to provide a useful response, to a particular question included in the survey.
4. Attrition: is a special kind of unit nonresponse arising in panel data and is defined as failure to obtain data from a sample unit at any wave after it has been selected into the survey.
5. New entry: it gives rise to a special kind of unit nonresponse due to failure to obtain retrospective data from a unit that has been added to the initial sample.

Attrition and new entry, both of households and persons, are the sources of the ECHP sample dynamics. Notice that, due to the ECHP follow-up rules, attrition is much higher for nonsample persons than for sample persons.

The ECHP tries to deal with unit nonresponse, sample attrition and new entry by weighting, and with item nonresponse by imputation. Because noncoverage has already been dealt with in Section 2.1, here we focus on unit nonresponse, item nonresponse, sample attrition and new entry. Some of the issues raised here are not new, and have already been discussed with reference to the PSID [see e.g. Beckett et al. (1988), Duncan and Hill (1989), Lillard (1989), Fitzgerald, Gottschalk and Moffitt (1994)].

6.1. Unit nonresponse

We distinguish between unit nonresponse at the household and the individual level.

Table 2 gives, for the first two waves, household response rates defined as the ratio of household interviews completed to the number of eligible households. For wave 1, the latter is just the number of households selected into the sample, excluding the cases which turned out to be nonexistent or otherwise ineligible.³⁹ For wave 2, it is the number of households

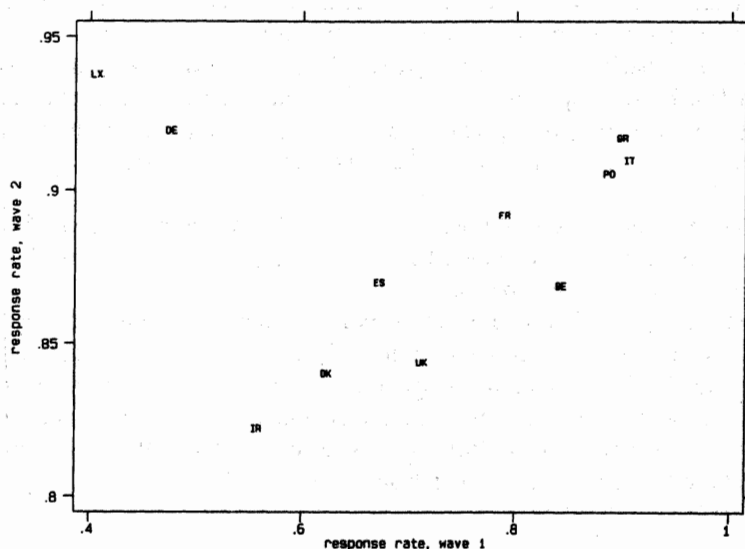
³⁹ In the case of Belgium, Portugal and Spain, calculation of household response rates in wave 1 is complicated by the fact that nonresponding households have been substituted by others. In Spain, for example, households noninterviewed because of final refusal to respond, non-contact or address not located have been replaced by households with similar characteristics drawn from a preselected substitute sample [see INE (1996)].

forwarded from wave 1, minus those which became ineligible between the two waves, plus the newly formed households.

Table 2 – Household response rates in the first two waves, by country

Country	Wave 1	Wave 2
Belgium	.844	.868
Denmark	.624	.839
France	.790	.891
Germany	.477	.918
Greece	.901	.916
Ireland	.558	.821
Italy	.907	.909
Luxembourg	.407	.937
Netherlands	.875	
Portugal	.889	.904
Spain	.673	.869
UK	.713	.843
Total	.730	

Figure 1 – Household response rates in the first two waves of the ECHP



The household response rate in wave 1 is 73 percent for the EU as a whole, but varies considerably among countries. It is above average in Italy, Greece, Portugal, Netherlands, Belgium, and France, and below average in Spain, Denmark, Ireland, Germany and Luxembourg. The UK is about average. Response rates are highest in Italy and Greece (91 and 90 percent respectively), which are the only countries where survey participation is compulsory, and lowest in Germany and Luxembourg (48 and 41 percent respectively).

Response rates in wave 2 are higher, often notably, as it is typical in panel surveys. Figure 1 shows the existence of a positive relationship between the response rates in the two surveys, with Germany and Luxembourg as notable outliers.

Within each country, response rates vary considerably by demographic group and by region. As an example, Table 3 shows wave 1 household response rates by county in Denmark. Response rates vary between a minimum of 49 percent for Kobenhavn and a maximum of 80 percent for Bornholms County.

Table 3 – Household response rates by county in Denmark. Wave 1

COUNTY	
København	.486
Frederiksberg	.500
København County	.542
Frederiksberg County	.573
Roskilde County	.606
Vestsjællen County	.642
Storstøms County	.598
Bornholms County	.800
Fyns County	.622
Sønderjyllands County	.714
Ribe County	.672
Vejle County	.716
Ringkøbing County	.705
Arhus County	.666
Viborg County	.581
Nordjyllands County	.716

Source: Eurostat (1995b)

The high nonresponse rates and their wide variation, both between and within countries, are worrisome. As recognized by Eurostat (1995d), despite the efforts, the start in terms of response rates in a few countries has been less than best.

Table 4, taken from Regoli (1996), provides some information about the reasons of household nonresponse in the Italian case. It is interesting to notice that non-contact decreased from from 6 percent in wave 1 to 4.6 percent in wave 2, while address not located, invalid or nonexistent dropped from 2.5 to .8 percent. At the same time, however, refusal to cooperate (initial or final) increased from 1.3 to 3.5 percent.

Table 4 – Results of the household interview in Italy

	Wave 1		Wave 2	
	No.	%	No.	%
Interview completed	7115	89.1	7128	90.9
Unable to respond (illness, etc.)	8	.1	20	.3
Initial refusal to cooperate	74	.9	86	1.1
Final refusal to cooperate	35	.4	185	2.4
Non contact	482	6.0	363	4.6
Address invalid or nonexistent	48	.6	20	.3
Interview not completed, reason missing	79	1.0	-	-
Address not located	148	1.9	42	.5
Total	7989	100.0	7844	100.0

Source: Regoli (1996)

Regoli (1996) also shows that the probability of successfully interviewing a household in wave 2 is about 94.5 percent if the household has been successfully interviewed in wave 1. It is 10 percentage points lower if the household is new, or it changed address, or it has

been interviewed by different interviewers in the two waves. The success probability drops to 40.5 percent if the household was not contacted in wave 1.

6.2 Item nonresponse

Again, we distinguish between item nonresponse at the household and the individual level. In evaluating item nonresponse, a number of issues have to be taken into account.

First, calculation of nonresponse rates is complicated by the presence of branching and skip patterns in the questionnaire that require identifying precisely the "universe" for each question.

Second, in the case of income sources, there are various degrees of item nonresponse. For example, although both gross and net earnings amounts are requested, some respondents may report only the gross or the net. In some cases, when the actual income amount is not specified, the questionnaire allows the response to take the form of a range. For most types of income, the questionnaire asks for the monthly average along with the number of months received. If these are not specified, the annual total may be recorded.

Third, in some countries certain variables are not available in the standard form. This may be either because the questions concerned were genuinely inapplicable, or because the questions were applicable but, for some reason, they were excluded or not asked in a standard form. An example is Denmark, where no information on widow's pension income has been collected.

Fourth, like for unit nonresponse, it is important to distinguish between two types of item non response. The first is the "missing at random" case. The second is the "nonignorable nonresponse" case, where the probability of nonresponse for the variable of interest depends upon the value of that variable [see Greenlees, Reece and Zieschang (1982) and Little (1982)]. This distinction is not made by the ECHP, although it has important implications for the imputation methods discussed in Section 8.1.

Verma (1995a) gives frequency distributions of item nonresponse⁴⁰ in wave 1 for all household questionnaire variables and two countries, Denmark and the UK. Table 5 presents, for each of the two countries, the 10 questions with the highest household nonresponse rate, defined as the ratio of nonstated cases to the number of applicable cases. It turns out that, except for 5 questions that, remarkably, are the same in both countries (property income range, net monthly income range, minimum income needed "to make ends meet", amount of inheritance received, and feeling about the current situation), item nonresponse is low and does not exceed 1.5 percent.

Table 5 – The 10 largest nonresponse rates (ratio of nonstated cases to applicable cases). Denmark and the UK, household questionnaire, wave 1

Denmark		%	UK		%
H01113	property income range	21.5	H01113	property income range	68.1
H01078	net monthly income range	10.4	H01078	net monthly income range	42.7
H01109	minimum income needed	6.6	H01109	minimum income needed	6.1
H01115	inheritance received amount	3.9	H01115	inheritance received amount	5.0
H01119	feeling about present situation	1.6	H01119	feeling about present situation	1.8
H01004	reasons for move	1.5	H01025	amount paid gross/net	1.1
H01003	moved from	1.3	H01057	is debit a burden?	.8
H01043	can't afford/don't want car	.8	H01035	rent includes sewage removal	.5
H01035	rent includes sewage removal	.8	H01110	property income during 1993	.3
H01034	rent includes water charges	.7	H01079	whether can save	.3

Source: Verma (1995a)

⁴⁰ Defined as the percentage of eligible households not responding.

Table 6 shows the distributions of nonresponse to the question on total net monthly income. In both countries, roughly 10 percent of the households do not provide actual amounts for total monthly income. In Denmark, 10.7 percent specify a range, while this proportion is only 5.6 percent in the UK. By contrast, complete nonresponse (not even the range is given) is very low (1.2 percent) in Denmark, but higher (4.2 percent) in the UK.

Table 6 – Percentage distribution of nonresponse on total net household monthly income. Denmark and the UK, wave 1

	Denmark	UK
Amount specified	88.1	90.2
Only range given	10.7	5.6
Not even range given	1.2	4.2
Total cases	100.0	100.0

Source: Verma (1995a)

In all cases, the information provided by Eurostat refers only to the unconditional distribution by type of response. To my knowledge, no analysis of item nonresponse as a function of observable household characteristics is available.

Turning to the individual questionnaire, Eurostat (1995a) provides some information on item nonresponse⁴¹. In general, item nonresponse for men is higher than for women. This may be due to women answering for their husbands, therefore missing some information.

In relation to earnings (current monthly earnings and earnings in 1993), one may distinguish between countries that give both gross and net earnings (Denmark, France, Ireland, UK), and the other countries that mainly or only provide net earnings. According to Eurostat, gross monthly earnings appear to be underestimated by 10 percent in general. Item nonresponse is high for current earnings in Belgium, Germany and Portugal, and for last year's earnings in Belgium, Germany and the UK. Item nonresponse on extra payments for overtime work is less than 10 percent in all countries except Belgium (and, according to Table 7, the UK as well).

According to Eurostat, item nonresponse on self-employment income is less than 10 percent in 6 countries (Greece, Ireland, Italy, Portugal, Spain, UK), is between 10 and 25 percent in 4 other countries (Denmark, France, Luxembourg, Netherlands), and "is a problem" in Belgium and Germany.

In the case of income from other work (agricultural, secondary or casual jobs), Eurostat admits that "for all countries except Portugal, item nonresponse is so high that the reliability of this variable is doubtful". Finally, in the case of social protection benefits, data are considered to be of good quality for unemployment insurance benefits, old-age and survivor pensions (first pillar), maternity allowances and sickness benefits.

Verma (1995b) gives frequency distributions of item nonresponse in wave 1 for all individual questionnaire variables and two countries, Denmark and the UK. Table 7 presents, for each of the two countries, the 20 questions with the highest individual nonresponse rate, defined as the ratio of nonstated cases⁴² to the number of applicable cases. The table seems to indicate that item nonresponse is much more serious for the individual than for the household questionnaire.

⁴¹ Defined as the percentage of eligible persons not responding, for income variables in wave 1.

⁴² For income variables, the number of non stated cases is the sum of the cases when it is not stated whether a particular income is received and the cases when the income amount received is not stated.

Table 7 – The 20 largest nonresponse rates (ratio of nonstated cases to applicable cases). Denmark and the UK, individual questionnaire, wave 1

Denmark	%	UK	%
P01166	77.5	P01166	66.7
P01354	38.0	P01011	58.3
G343	36.6	G336	41.9
P01344	36.3	F177	39.3
F072	28.5	P01174	39.2
P01120	15.6	P01313	39.0
P01173	15.5	P01173	15.9
P01353	14.3	P01167	14.4
P01115	14.2	P01169	14.4
P01114	13.6	P01140	14.2
P01119	13.6	F138	14.2
P01113	13.5	P01143	13.8
F068	10.2	P01170	13.3
P01170	10.0	P01168	13.3
P01169	0.0	F170	13.3
P01167	9.8	F310	12.0
P01168	9.8	F154	11.4
F170	9.8	F158	11.3
P01297	9.7	F156	10.8
F172	8.0	P01153	10.8

Source: Verma (1995b)

As for the household questionnaire, the information provided by Eurostat refers only to the marginal distribution by type of response. To my knowledge, no analysis of item non-response as a function of observable personal characteristics is available.

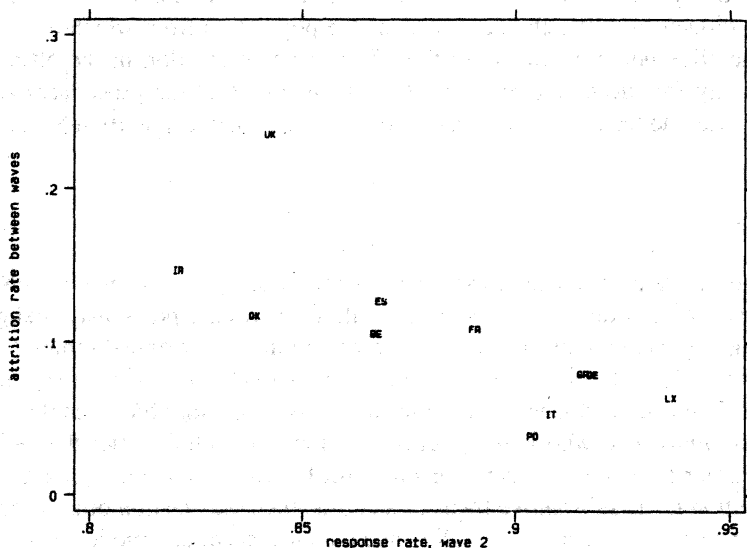
6.3. Attrition

Table 8 presents some data on household attrition between the first two waves of the ECHP. The attrition rate is defined here as one minus the ratio between the number of households interviewed in both waves and the number of households interviewed in wave 1 who are still existing in wave 2.

Table 8 – Household attrition rates between wave 1 and wave 2, by country

COUNTRY	Interviewed in wave 1	No longer existing	Interviewed both waves	Attrition rate
Belgium	4189	37	3723	.103
Denmark	3482	77	3014	.115
France	7344	51	6518	.106
Germany	5054	0	4667	.077
Greece	5523	41	5060	.077
Ireland	4048	36	3433	.144
Italy	7115	59	6697	.051
Luxembourg	1011	4	945	.062
Netherlands	5187			
Portugal	4881	51	4654	.036
Spain	7206	16	6295	.124
UK	5779	34	4405	.233
Total	60819			

Figure 2 – Household attrition rates between waves 1 and 2 and household response rates in wave 2



Attrition rates are highest in the UK (23 percent) and lowest (5 percent or less) in Portugal and Italy. In the remaining eight countries for which data are available, they range between 6 and 14 percent. Figure 2 shows the existence of a clear negative relationship between household attrition rates and household response rates in wave 2.

Turning to individuals, it is useful to distinguish between two types of sample attrition. The first is monotone attrition: it occurs when survey participants leave the sample and never return. The second is nonmonotone attrition: it occurs when survey participants who at some point leave, return to the sample at a later date. Since both types are possible in the ECHP, there are four kinds of sample participants: full-time participants (no attrition), base-year participants with a monotone attrition pattern, base-year participants with a nonmonotone attrition pattern, and non-base year participants that have been added to the sample. An important issue is in which dimensions these four types of sample participants differ, both cross-sectionally and longitudinally.

The lesson from the PSID is that even low attrition rates may accumulate and become substantial over many years. Thus, for example, monotone attrition of 10 percent per year implies that 27 of the initial sample is lost after 3 years and 47 percent after 6 years. When attrition is exogenous, that is, unrelated to the response variable of interest, we only have an efficiency problem but no bias. This is no longer true when attrition is endogenous, that is, systematically related to the response variable.

In a household panel, attrition is typically related to important transitions in a person's life: going to college, finding a new job, marriage or divorce, retirement, death. These events are, at the same time, the main object of study using panel data. Attrition cannot therefore be ignored, for it may lead to invalid inference, even when attrition rates are modest.

Most evaluation studies [see e.g. Beckett *et al.* (1988) and Peracchi and Welch (1995)], focus on the effect of attrition on the cross-sectional representativeness of a panel, that is, its representativeness of the relevant population at a given point in time. However, because panel data are specifically collected to study dynamic behavior, it is longitudinal representativeness that ought to be the primary concern of an evaluation.

Another basic assumption in the literature seems to be that attrition stems mostly from individual behavior. Yet, an interesting feature of the NLS is the fact that its Survey Groups are characterized by very different attrition rates [see Center for Human Resource Research (1994)]. This reflects both the characteristics of the population from which each sample is drawn and the NLS policy towards attrition. The fact that attrition in the NLS has been declining steadily for successive Surveys Groups suggests that the procedures adopted to control for attrition do have a major impact and should be taken explicitly into account.

6.4. *New entry*

In addition to sample attrition, the other source of sample dynamics is new entry of households and persons due to household formation, marriage, previously ineligible persons becoming eligible, replenishment or supplementation of the original sample, etc.

As with attrition, if the process that determines new entry is not exogenous with respect to the phenomenon under study, biases may arise. Consider for example, as in Fienberg (1989), an individual who joins or leaves a sample household, after the beginning of the panel, in order to qualify for cash- or non-cash benefits from some program. "If program participation is the substantive focus of analysis, then the creation of new family units is clearly a form of outcome-based sampling and the analysis of any model using these data may need to incorporate the selection mechanism into the model under study." Because no replenishment or supplementation to the original ECHP sample has yet been made, new entry of sample persons is largely confined to individuals who were not eligible for interview in 1994 because aged less than 16 and may therefore be considered as exogenous. On the other hand, nonsample persons are individuals who joined the household of an original sample member after the first wave. Because their entry is often related to events, such as marriage, living arrangements, etc., that are at the same time the main object of study using panel data, ignoring new entry may in this case lead to bias.

7. **Weighting**

In order to compensate for the differential sampling probabilities and response rates, both within and between waves, the ECHP assigns weights to the response cases (households and persons). Eurostat seems to have devoted a considerable amount of its resources to this effort.

There are several problems with weighting. First, because of the substantial departure from simple random sampling, due to both the initial survey design and subsequent attrition, unweighted sample statistics are likely to be biased. Second, due to the complexity of the survey, it is not at all clear which is an appropriate set of weights, especially for longitudinal analysis. Third, as pointed out by DuMouchel and Duncan (1983), while weighting may be appropriate for estimating population means and totals, relying too much on weighted estimates and ignoring unweighted results may be dangerous in regression problems. Fourth, as stressed by Hoem (1989), weighting "can be superfluous in situations where the investigator is really involved in modeling human behavior rather than in calculating descriptive statistics for the finite population and ... may divert attention away from more important concerns of modeling and analysis".⁴³

⁴³ One exception being, of course, the case of outcome-based or choice-based sampling, where the sampling plan is informative about the relationship of interest.

7.1. Initial weights

Initially, in wave 1, household and personal weights coincide. These weights, defined for all interviewed households, are computed according to the following steps:⁴⁴

1. Compute design weights $\omega^{(1)}$ inversely proportional to household selection probabilities. If households are selected with uniform probabilities, as in the case of Denmark, then $\omega^{(1)} = 1$.
2. Compute nonresponse weights $\omega^{(2)}$ inversely proportional to the response rates of each weighting class in which the sample has been partitioned. For example, in the case of Denmark, the weighting classes correspond to the counties and the nonresponse weights vary from a minimum of .775 (= 61.96/80.00) for Bornholms County to a maximum of 1.275 (= 61.96/48.58) for København [see Table 3].
3. Compute weights $\omega^{(3)}$ correcting the distribution of households. These weights make the distribution of the households by various socio-economic characteristics agree with the corresponding distribution derived from the census or the Labor Force Survey. For example, in the case of Denmark, the external controls include the household type, the number of economically active persons in the household, the household tenure, and the county of residence.
4. Compute weights $\omega^{(4)}$ correcting the distribution of persons. These weights make the distribution of the persons by sex and age agree with the corresponding distribution derived from the census or the Labor Force Surveys.
5. Compute the final weights ω as

$$\omega = \omega^{(1)} \times \omega^{(2)} \times \omega^{(3)} \times \omega^{(4)}$$

Steps 1 and 2 were carried out by the NDU's, Steps 3 and 4 by Eurostat (except for Belgium, France, Germany, Ireland and Italy), Step 5 by Eurostat. The weights computed at each step are trimmed in order to avoid extreme values, and normalized in such a way that they average to one.

Table 9 – Weighting of wave 1 for Denmark

	Population %	ECHP %
Household type:		
- Single male aged 60+	4.7	5.7
- Single female aged 60+	12.6	13.5
- Single male under 60	15.8	15.3
- Single female under 60	10.2	9.5
- Couple with no children	25.5	26.0
- Couple with children all less than 16	13.9	13.5
- Couple with some children aged 16+	10.9	10.3
- Other	6.4	6.2
Total	100.0	100.0
Number of economically active persons:		
- No person	26.3	28.4
- One person	35.7	35.2
- Two or more persons	38.0	36.4
Total	100.0	100.0

Source: Eurostat (1995b)

⁴⁴ For a theoretical justification of this procedure, see Sa'ndal, Swensson and Wretman (1992), Section 15.6.

Weighting is far from perfect, however. Even after weighting, some discrepancies remain between the population distribution of households and their distribution in the ECHP. For example, in the case of Denmark (the only country for which results have been provided), the ECHP weights computed after Step 5 seem to overrepresent households consisting of a single person aged 60 or more, and households with no economically active persons [Table 9]. Given the role of the ECHP in social policy design, the consequences of these discrepancies are potentially worrisome.

7.2. Subsequent waves

Starting from wave 2, a distinction is made between household and personal weights, and also between cross-sectional and longitudinal weights. The latter should help making the panel longitudinally representative. Personal weights for any wave $t > 1$ are obtained recursively from those of the previous one. While the methodology for initial weighting is essentially uncontroversial, problems arise with the updating scheme.⁴⁵

The "basic weight" $\omega_{ij}(t)$ of sample adult j in household i interviewed at wave $t > 1$ is computed as

$$\omega_{ij}(t) = f_{ij}(t-1, t) \omega_{ij}(t-1) \quad [1]$$

where $\omega_{ij}(t-1)$ is the basic weight of sample adult j in household i at wave $t-1$ and $f_{ij}(t-1, t)$ is an adjustment factor. For nonsample persons, instead, the basic weight is set equal to zero.

In the absence of changes in the distribution of the population by demographic characteristics, Eurostat computes the adjustment factor as

$$f_{ij}(t-1, t) = \frac{1}{r_{ij}(t-1, t) c_{ij}(t-1, t)} \quad [2]$$

where $r_{ij}(t-1, t)$ is an estimate of the probability that a sample person from wave $t-1$ is successfully contacted in wave t and $c_{ij}(t-1, t)$ is an estimate of the probability that a contacted person is successfully interviewed. These contact and interview rates are computed either partitioning the sample in weighting classes, or by probit or logit. At least in principle, the adjustment factor $f_{ij}(t-1, t)$ is further modified to take into account changes in the distribution of the population by demographic characteristics.

There are two problems with the recursive formula [1] - [2]. First, if retention probabilities are the same for all households and all persons, then no updating is necessary since the weights remain unchanged between the two periods. Second, if there are changes in the distribution of the population by demographic characteristics, then it is far from obvious how the appropriate adjustment factors should be computed. This is likely to be a serious problem if weighting classes are based on the region of residence and the population is geographically mobile. The problem is further complicated by the fact that some of the changes may involve unmeasured demographic characteristics, that is, characteristics that do not enter in the definition of weighting class.

⁴⁵ A more fundamental issue, not discussed here, is whether updating the weights makes sense. As pointed out by Lillard (1989), if one believes that weighting is appropriate for behavioral models, then "the original sample weights would seem most appropriate. There is no need of update the weights over time."

7.3. Derived weights

The weights $\omega_{ij}(t)$ assigned at each wave $t > 1$ to sample adults are the basis for deriving a number of other weights. Given the basic weights, the weight $\omega_i(t)$ of household i in wave t is set equal to the average of the basic weights of its adult members. Thus

$$\omega_i(t) = \frac{1}{h_i} \sum_{j=1}^{h_i} \omega_{ij}(t) = \frac{1}{s_i} \sum_{j=1}^{s_i} \omega_{ij}(t)$$

where $h_i = s_i + n_i$ is the total number of adults, s_i is the number of sample adults (receiving a positive basic weight) and n_i is the number of nonsample adults (receiving zero basic weight). In order to allow for cross-sectional representation of the population at t , Eurostat suggests assigning the household weight $\omega_i(t)$ to each member of household i , whether sample or nonsample person and whether adult or child.

Eurostat proposes a similar approach for longitudinal weighting. Denoting by $s_i(T)$ the number of sample adults in household i at time T and by $n_i(t, T)$ the number of nonsample adults who entered the household at time t , or after, and are still present at time T , Eurostat suggests computing the weights for a longitudinal analysis covering the time interval between t and T as

$$\omega_i(t, T) = \frac{1}{s_i(T) + n_i(t, T)} \sum_{j=1}^{s_i(T)} \omega_{ij}(T)$$

The same longitudinal weight $\omega_i(t, T)$ is then assigned to each member of household i , whether sample or nonsample person.

8. Imputation

Imputation is inserting values that are missing because of item nonresponse. Why imputing missing values? The conventional answers are: to make the data easier to analyze, to reduce nonresponse bias, and to increase precision.

Imputation has several drawbacks, however. First, it can lead to biased estimates of bivariate and multivariate parameters, such as correlations. Second, in the "nonignorable response" case, it can even lead to biased estimates of the population mean. Third, it can lead to overestimate precision. Fourth, inserting imputed values as predictors in a regression relationship leads to the problems typical of generated regressors. Fifth, imputation is time-consuming and, if carried out at the level of the data collecting agency, tends to delay the release of the data.

According to the guidelines set in Eurostat (1995c), "the tight time schedule in the ECHP production process and limitation on manpower dictate concentration of the imputation on the most important variables. Eurostat's major concern is with income and its components, together with the preservation of covariates". More precisely, imputation is carried out for household's total net monthly income and a number of income components, namely gross and net monthly earnings, gross and net annual earnings, holiday pay or allowance, annual pre-tax profit, unemployment benefit per month, unemployment assistance per month, income from old-age pensions per month, benefits from early retirement schemes, other old-age related schemes or benefits, and annual income from capital and investent.

For other income components, missing values have simply been treated as zeros in the construction of total income.

8.1. Imputation methods

Most common imputation methods may be viewed as special cases of the formula

$$\hat{X}_i = BW_i + V_i, \quad [3]$$

where \hat{X}_i is an m -vector of imputed values for the i -th sample unit, B is an $m \times k$ matrix of coefficients, W_i is a k -vector of auxiliary variables, and V_i is an m -vector of residuals. The various methods differ in the choice of B , W_i and V_i . Methods that set $V_i = 0$ are called deterministic, and are called stochastic otherwise.

Examples of deterministic imputation methods are the overall mean imputation, the class mean imputation, and the deterministic regression imputation. In overall mean imputation, \hat{X}_i is set equal to the overall respondent mean \bar{X} , that is, $k = 1$, $W_i = 1$ and $B = \bar{X}$ is the m -vector of means computed from the subsample of the respondents. In class mean imputation, the sample is divided into k classes and \hat{X}_i is set equal to the respondent mean \bar{X}_h for the class that contains the i -th sample unit, that is, W_i is the k -vector of dummy variables that represent the classes and B is a matrix whose h -th column is equal to \bar{X}_h . In deterministic regression imputation, \hat{X}_i is the predicted value of X_i given W_i from a multivariate regression of X on W using the subsample of the respondents.

The stochastic versions of these methods are the random hotdeck overall imputation, the random hotdeck imputation within classes, and the stochastic regression imputation. In random hotdeck overall imputation, a "donor" j is randomly selected from the subsample of the respondents and \hat{X}_i is set equal to X_j . Thus, $V_i = \hat{X}_i - \bar{X}$ is the donor's deviation from the overall mean. In random hotdeck imputation within classes, a "donor" j is randomly selected from the subsample of the respondents in the class h that contains the i -th sample unit and \hat{X}_i is set equal to X_j . Thus, $V_i = \hat{X}_i - \bar{X}_h$ is the donor's deviation from the h -th class mean. Stochastic regression imputation adds to the predicted value of X_i a residual V_i which is randomly drawn from a distribution with mean zero and variance equal to the residual variance of the regression.

8.2. Cross-sectional imputation

This is the only kind of imputation carried out by Eurostat. The method chosen is random hotdeck imputation within classes. In order to preserve key covariances, the same donor is used for imputation of related missing variables.

8.3. Longitudinal imputation

In principle, all methods described in Section 8.1 may be generalized to longitudinal imputation. The need for the latter has sometimes been advocated on the ground that cross-sectional imputation of a variable may produce an implausible measure of change in that variable.

The issue is whether longitudinal imputation makes sense. As pointed out by Lillard (1989), "while imputation procedures may be useful for generating missing values for ran-

dom item nonresponse, the idea seems futile in the case of wave nonresponse generated by attrition, or the entry of nonsample individuals”.

9. Conclusions

So far, the analysis of a large set of issues (labor force participation decisions, human capital investment decisions, retirement decisions, earnings structure and dynamics, effects of training programs, effects of social protection programs, etc.) has been made substantially more difficult in Europe than in the USA by the lack of adequate data on the two main labor market indicators: prices (earnings) and quantities (employment). While the information on European employment and unemployment is comparable to the one available for the USA in terms of quantity and quality, the information on earnings is difficult to obtain and its quality is unsatisfactory.

From this viewpoint, Eurostat's decision to launch the ECHP is to be praised, as it represents an important step forward towards filling the informational gap between Europe and the USA, although the fundamental question raised by Heckman and Robb (1989) remains of whether a longitudinal survey is really necessary for most of the analyses for which the ECHP has been designed. In particular, couldn't a suitably designed repeated cross-sectional survey do just as well?

Like any other survey, the ECHP raises important methodological and statistical problems that require careful examination. Unfortunately, the information released so far by Eurostat is fairly limited. To make things worse, the laws about privacy that have been passed in most EU member countries are very strict and severely restrict access to the micro-data. Unlike the USA, the approach towards data access prevailing in Europe is “no access unless authorization has been granted”. This situation has obvious negative consequences, not just for the potential users, but for Eurostat itself who is deprived of important feedback.

Appreciation, by the academic community and the public at large, of the effort carried out by Eurostat with the ECHP survey can only come if the data are made widely available to researchers, and the associated methodological and statistical issues (attrition, selection, measurement error, etc.) are openly discussed and well understood.

References

- BECKETTI, S., W. GOULD, L. LILLARD, AND F. WELCH (1988), “The Panel Study of Income Dynamics after fourteen years: An evaluation”, *Journal of Labor Economics*, 6, 472-492.
- CENTER FOR HUMAN RESOURCE RESEARCH (1994), *NLS. User's Guide 1994*, Columbus, OH, The Ohio State University.
- DUMOUCHEL, W.H., AND G.J. DUNCAN (1983), “Using sample survey weights in multiple regression analysis of stratified samples”, *Journal of the American Statistical Association*, 78, 535-543.
- DUNCAN, G.J., AND D.H. HILL (1989), “Assessing the quality of household panel data: The case of the Panel Study of Income Dynamics”, *Journal of Business and Economic Statistics*, 7, 441-452.
- EUROSTAT (1995a), “Preparation of micro-data files within Eurostat”, PAN 45/95, Luxembourg, Eurostat.
- EUROSTAT (1995b), “Weighting results of wave 1: empirical results”, PAN 46/95, Luxembourg, Eurostat.
- EUROSTAT (1995c), “Cross-sectional imputation rules, and application to the micro-data files”, PAN 47/95, Luxembourg, Eurostat.
- EUROSTAT (1995d), “ECHP replenishments/supplementation”, PAN 56/95, Luxembourg, Eurostat.
- EUROSTAT (1996), *The Household Panel Newsletter*, 1, Theme 3 (Population and social conditions), Series B (Short-term statistics), Luxembourg, Office for Official Publications of the European Communities.

- FIENBERG, S.E. (1989), "Modelling considerations: Discussion from a modelling perspective", in D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.
- FITZGERALD, J., P. GOTTSCHALK, AND R. MOFFITT (1994), "A study of sample attrition in the Michigan Panel Study of Income Dynamics", paper presented at the Conference on Attrition in Longitudinal Surveys, February 24-25, 1994, Washington, D.C.
- GREENLEES, J.S., W.S. REECE, AND K.D. ZIESCHANG (1982), "Imputation of missing values when the probability of response depends on the variable being imputed", *Journal of the American Statistical Association*, 77, 251-261.
- HECKMAN, J.J., AND R. ROBB (1989), "The value of longitudinal data for solving the problem of selection bias in evaluating the impact of treatments on outcomes", in D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.
- HILL, M.S. (1992), *The Panel Study of Income Dynamics. A User's Guide*, Newbury Park, CA, Sage Publications.
- HOEM, J.M. (1989), "The issue of weights in panel surveys of individual behavior", in D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.
- LILLARD, L. (1989), "Sample dynamics: Some behavioral issues", in D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh (eds.), *Panel Surveys*, Wiley, New York.
- LITTLE, R.J.A. (1982), "Models for nonresponse in sample surveys", *Journal of the American Statistical Association*, 77, 237-250.
- O'MUIRCHARTAIGH, C. (1989), "Sources of nonsampling error: Discussion", in D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh (eds.), *Panel Surveys*, New York, Wiley.
- PERACCHI, F., AND F. WELCH (1995), "How representative are matched cross-sections? Evidence from the Current Population Survey", *Journal of Econometrics*, 68, 153-180.
- REGOLI, M. (1996), "Empirical and analytical experiences from the first waves of the Italian ECHP", Rome, Istat, mimeo.
- SÄRNDAL, C.-E., B. SWENSSON, AND J. WRETMAN (1992), *Model Assisted Survey Sampling*, New York, Springer.
- VERMA, V. (1995a), "Item non-response and frequency distributions: Household questionnaire wave 1. Illustration from two countries (DK,UK)", PAN 60/95, Luxembourg, Eurostat.
- VERMA, V. (1995b), "Item non-response and frequency distributions: Individual questionnaire wave 1. Illustration from two countries (DK,UK)", PAN 61/95, Luxembourg, Eurostat.

LA STIMA DEI FLUSSI RISPETTO AL LAVORO IN PRESENZA DI INDICATORI AFFETTI DA ERRORE

Nicola Torelli e Francesca Bassi

Dipartimento di Scienze Statistiche – Università di Padova

1. Introduzione

Le stime correnti, relative alla dimensione degli stock di occupati o disoccupati in determinati istanti temporali, consentono di valutare le variazioni nette. Le informazioni che, invece, risultano di maggior interesse, soprattutto ai fini di analisi e di valutazioni di politiche, sono i flussi lordi. Tali informazioni consentono di misurare, ad esempio, la proporzione di unità che, essendo risultate disoccupate ad un certo istante di tempo, permangono tali o che, complementariamente, risultano avere trovato in un tempo successivo un'occupazione (o sono uscite dal mercato del lavoro); ancora, esse permettono di stabilire in quale misura un aumento della disoccupazione sia dovuto all'aumento del numero di coloro che essendo precedentemente occupati hanno perso il lavoro e sono alla ricerca di nuova occupazione, oppure all'incremento del numero di unità che appartenevano alle non forze di lavoro e che iniziano a cercare un'occupazione.

I dati necessari per la stima dei flussi lordi fra condizioni lavorative derivano, com'è ovvio, dalla rilevazione in una successione di istanti temporali, della condizione rispetto al lavoro per ciascuna unità del campione considerato; in altre parole sono necessarie informazioni longitudinali.

La disponibilità di informazioni longitudinali è attualmente più diffusa rispetto al passato; ciò vale anche per quelle indagini che sono dedicate in particolar modo alla rilevazione delle condizioni rispetto al lavoro della popolazione. Inoltre in disegni di rilevazione che prevedono la reintervista di parte del campione (qual è, ad esempio, il disegno con campione ruotato adottato in Italia da un paio di decenni per la Rilevazione Trimestrale sulle Forze di Lavoro - RTFL), è frequente il ricorso a quesiti retrospettivi che consentono di ricostruire la storia lavorativa recente delle unità oggetto di indagine. Fra i vantaggi che vengono riconosciuti a tali disegni di rilevazione vi è quello di consentire la stima delle matrici di flusso. Tali esperienze sono state tuttavia riconsiderate criticamente alla luce di una maggiore consapevolezza dei problemi metodologici che è necessario risolvere al fine di ridurre eventuali effetti distorsivi che possono condurre a interpretazioni fuorvianti dei comportamenti dinamici dell'offerta di lavoro.

Uno dei principali problemi, in questo ambito, è legato all'impatto degli errori di

Lavoro svolto nell'ambito del progetto strategico di ricerca CNR "La disoccupazione e il basso livello di attività in Italia: cause specifiche e rimedi adeguati", contributo n.95.04495.ST74. Il lavoro è frutto dell'impegno comune dei due autori; per quanto riguarda la stesura, N. Torelli ha curato le sezioni 1, 2.1, 2.2, 4 e 5, F. Bassi le sezioni 2.3 e 3

misura (o errori di classificazione) sulle stime dei flussi lordi. Se è infatti spesso ragionevole assumere che tali errori si bilancino nella stima degli stock, essi possono condurre a sensibili distorsioni quando si voglia studiare il cambiamento; l'errore di classificazione può infatti tradursi in transizioni spurie, ed è necessario disporre di metodi di stima che consentano di separare la vera mobilità da quella eventualmente imputabile alla presenza di errori.

Tale tema è stato oggetto di notevole attenzione negli ultimi anni (per una rassegna, si vedano Singh e Rao, 1995 e Skinner e Torelli, 1993). Soluzioni al problema e possibili linee di intervento passano attraverso due strategie complementari: (i) l'integrazione delle informazioni raccolte mediante il disegno longitudinale ricorrendo, ad esempio, alla reintervista di parte del campione per ottenere in qualche modo "dati di validazione", oppure introducendo ulteriori quesiti nel modello di rilevazione; (ii) la specificazione di modelli statistici che sfruttino appieno la sequenza di dati longitudinali e muovano da assunzioni riguardo al meccanismo dinamico che si ritiene all'origine dei flussi sul mercato del lavoro.

Come detto, le due strategie non sono in stretta alternativa e in entrambi i casi si richiede la disponibilità di informazioni aggiuntive. Inoltre le due strategie possono essere ricondotte ad una formulazione comune che comprende la specificazione di un modello di misura e di un modello per la dinamica del processo di transizione.

Nel seguito viene riconsiderato il problema della stima dei flussi lordi nel mercato del lavoro quando si dispone di indicatori della condizione riguardo al lavoro affetta da errori di classificazione. Nella sez. 2 si richiamano i problemi di stima dei flussi in presenza di errori di misura e in particolare si fornisce una formulazione generale del problema, che consente di pervenire ad aggiustamenti delle stime dei flussi in presenza di errori di classificazione. Nella sez. 3 viene presentata un'applicazione alla stima dei flussi nel mercato del lavoro considerando i dati italiani raccolti nella RTFL e in un'indagine suppletiva alla stessa, mentre nella sez. 4 si discute la possibilità di pervenire a stime dei flussi impiegando i soli dati raccolti attraverso l'indagine corrente sulle forze di lavoro. La sez. 5 contiene alcune considerazioni conclusive.

2. La stima dei flussi lordi in presenza di errore di misura

2.1. I dati per la stima dei flussi lordi

Dati di flusso possono essere ottenuti attraverso diverse procedure di rilevazione: ad esempio, si possono derivare informazioni longitudinali (i) sfruttando le opportunità offerte da disegni di campionamento ruotato adottati per indagini il cui scopo principale è la stima degli stock, (ii) ricorrendo a indagini panel, (iii) mediante quesiti retrospettivi in indagini sezionali convenzionali oppure (iv) combinando opportunamente dati provenienti da fonte amministrativa.

I rispettivi meriti e le debolezze di ciascuna procedura con riferimento allo specifico caso della stima di flussi lordi fra condizioni lavorative sono nella rassegna proposta da Mehran (1990). In questa sede, l'angolazione con cui guardare alle varie procedure riguarda le possibili implicazioni che diverse strategie di collezione di informazioni longitudinali hanno sulle opportunità per la stima di flussi in presenza di errori di classificazione, ponendo eventualmente particolare attenzione al caso italiano.

È volutamente trascurata nell'analisi che segue la fonte amministrativa per la difficoltà, almeno nel caso italiano, di adattare concetti e definizioni proprie della gestione amministrativa alle esigenze statistiche di misurazione della condizione rispetto al lavoro.

Si ricorda tuttavia che tale fonte, usata eventualmente per integrare informazioni che provengono da indagini campionarie, è con successo usata in Paesi che hanno una lunga tradizione nell'uso di dati provenienti da registri amministrativi (ad esempio, in Danimarca).

Le informazioni di maggior valore per la stima dei flussi lordi sono usualmente quelle raccolte in indagini sulle forze di lavoro per le quali è previsto un disegno di campionamento ruotato, e quindi è possibile disporre, almeno per una quota del campione, dell'informazione riguardo alla condizione lavorativa in almeno due istanti successivi. È questo lo schema di rilevazione adottato in numerosi Paesi (fra gli altri, Italia, Stati Uniti, Canada, Australia) per le indagini correnti sulle forze di lavoro; la misurazione della condizione rispetto al lavoro è in tali indagini estremamente accurata e, anche al fine di garantire la comparabilità internazionale, vengono in genere adottati scrupolosamente i criteri raccomandati dell'International Labour Office per la definizione dei disoccupati, degli occupati e delle non forze di lavoro.

Le indagini panel hanno quale esclusivo o principale obiettivo quello di studiare il comportamento di una specifica popolazione nel tempo, a tal fine uno stesso campione di individui (o famiglie) viene intervistato in successivi istanti. Si tratta spesso di indagini su larga scala progettate esplicitamente per ottenere dati longitudinali che evidentemente consentono di ottenere stime di flusso. In tali indagini comunque la rilevazione della condizione rispetto al lavoro assume un rilievo marginale e questa viene di solito rilevata senza tentare di ricondursi a convenzioni definitorie opportune. Piuttosto comune è il ricorso all'autodichiarazione della condizione da parte del rispondente, informazione questa che viene considerata, e con buone ragioni, scarsamente attendibile.

In indagini sezionali periodiche, il ricorso a quesiti retrospettivi è piuttosto comune. I quesiti retrospettivi hanno il vantaggio di fornire con costi contenuti informazioni sulla sequenza di eventi (relativi alla storia lavorativa) che hanno riguardato la popolazione di riferimento. Nel caso della misurazione della condizione lavorativa (o non lavorativa), la possibilità di adattare ad uno schema di interrogazione retrospettiva gli stringenti criteri definitori adottati nel caso delle indagini sulle forze di lavoro è tuttavia limitata drasticamente. Inoltre è ben noto che i dati retrospettivi sono particolarmente esposti alla presenza di errori tipicamente legati alla fallacia della memoria del rispondente.

Nessuna delle procedure fornisce, com'è ovvio, dati che siano esenti da errori di classificazione, ma è necessario avere in mente che nei diversi casi l'incidenza e le caratteristiche degli errori di misura possono notevolmente differire; tale aspetto va ovviamente tenuto presente quando si procede alla stima di flussi lordi. È inoltre piuttosto comune la situazione in cui le informazioni longitudinali possano derivare da un miscuglio delle strategie di rilevazione segnalate. Il ricorso a quesiti retrospettivi è frequente in indagini panel, allo scopo di coprire il periodo che separa due rilevazioni successive. È il caso dell'indagine panel statunitense *Survey of Income and Program Participation* (David, 1985), ove l'intervista viene riproposta con cadenza di quattro mesi, ma si ricorre anche alla ricostruzione retrospettiva con riferimento alla storia lavorativa nel periodo che separa le due interviste; un'analoga strategia di interrogazione è presente anche per l'indagine sulle forze di lavoro francese (Magnac e Visser, 1996). Ciò consente, fra l'altro, di stimare flussi lordi relativamente a tempi più brevi di quelli che separano le interviste. Va poi ricordato che per le indagini sulle forze lavoro con campione ruotato, almeno in occasione dell'indagine comunitaria per i Paesi dell'Unione Europea, si richiede retrospettivamente quale era la condizione lavorativa nel corrispondente mese dell'anno precedente. In tali casi si può disporre all'interno della medesima indagine di più informazioni per la stima di flussi lordi, cioè il flusso lordo può essere ricavato o sfruttando il dato fornito dallo stesso rispondente nelle due occasioni di indagine o utilizzando l'informazione retrospettiva.

Si è già osservato come la possibilità di formulare modelli che consentano di pervenire a stime dei flussi lordi in presenza di errori di classificazione dipenda dalla disponibilità di informazioni aggiuntive, in tal senso l'uso combinato delle diverse strategie per ottenere dati di flusso fornisce interessanti opportunità.

L'informazione di maggior rilievo per valutare gli effetti della presenza di errori di classificazione e per operare conseguenti correzioni degli effetti distorsivi è comunque quella che risulta dalla possibilità di avere, almeno per una parte del campione, un dato di "validazione", di conoscere cioè la condizione vera al momento dell'intervista. L'ottenimento di tali informazioni è in genere legato a dati raccolti mediante programmi di reintervista: si dispone cioè una seconda visita presso un selezionato insieme di rispondenti al fine di verificare il dato riportato. Tali strategie sono state adottate, ad esempio, negli Stati Uniti per la *Current Population Survey* (il programma di reinterviste prevede inoltre la riconciliazione del dato precedentemente riportato per almeno una parte del campione reintervistato; una succinta descrizione del programma di reinterviste per l'indagine statunitense è in Abowd e Zellner, 1985 e Sinclair e Gastwirth, 1996) ed in Olanda, dove la reintervista è reiterata, per lo stesso sottocampione di unità, anche nella successiva occasione di indagine. Ovviamente in tal caso si presume di poter conoscere mediante la reintervista il "vero" stato e di poter determinare quindi l'entità dell'errore di classificazione. Tuttavia è ragionevole credere che attraverso programmi di reintervista si abbia solo un indicatore più accurato del vero stato e di conseguenza accettare che il vero stato sia comunque non osservabile. Inoltre i programmi di reintervista hanno una diffusione limitata a causa dei costi ad essi connessi; d'altra parte anche in presenza di tali informazioni è indispensabile il ricorso a ulteriori assunzioni per utilizzare procedure di aggiustamento (nel caso dei dati americani, ad esempio, è comunque indispensabile l'assunto che gli errori di classificazione siano serialmente incorrelati). I primi tentativi di produrre stime aggiustate per tener conto degli errori di misura poggiavano proprio sulla disponibilità di dati di reintervista (Poterba e Summers, 1986, Chua e Fuller, 1987 e Abowd e Zellner, 1985).

Un'altra strada che consente di ottenere una molteplicità di indicatori sulla condizione rispetto al lavoro è quella di ricorrere a rilevazioni suppletive alle indagini sulle forze di lavoro. In tal caso l'indagine suppletiva, che interessa le stesse unità che partecipano all'indagine corrente, consente di ottenere retrospettivamente informazioni sulla storia lavorativa. A tale schema si può ricondurre il *March Supplement* della *Current Population Survey* degli Stati Uniti e l'esperienza dell'indagine suppletiva alla rilevazione trimestrale sulle forze lavoro condotta in Italia nel 1989 (vedi oltre alla sez. 3).

La stima di flussi contaminati da errori di classificazione può inoltre avvalersi della rilevazione della condizione rispetto al lavoro per una sequenza di più istanti (almeno 3). In tal caso infatti si possono considerare ragionevoli assunzioni riguardo il processo di transizione, prescindendo quindi dalla presenza di dati di validazione (a tale situazione possono ricondursi ad esempio, i lavori di Van De Pol e Langeheine, 1990 e 1992).

In tutti i casi, le sole informazioni longitudinali rilevate non sono sufficienti. È infatti comunque decisivo il ruolo delle assunzioni (riguardo sia alla struttura degli errori di misura che alla dinamica del processo di transizione) adottate all'interno dei modelli statistici che si possono proporre.

È quindi di qualche utilità ricondurre le diverse situazioni informative all'interno di una formulazione generale per modelli adatti a stimare i flussi lordi in presenza di errori di misura. Fra l'altro essa consente di evidenziare il merito relativo delle eventuali informazioni supplementari e delle particolari assunzioni riguardo al modello di misura ed al modello di comportamento.

2.2. Una formulazione generale dei modelli per la stima dei flussi lordi in presenza di errori di classificazione

Ad ogni istante temporale $t = 1, 2, \dots, T$ la variabile Y_t denoti lo stato osservato per una generica unità di un campione di n elementi. Y_t può assumere i valori $1, 2, \dots, r$ che rappresentano l'insieme di stati alternativi secondo cui viene classificata la popolazione. Si assume che la variabile sia misurata con errore, ovvero che per ogni generica unità sia definita la variabile y_t che designa lo stato vero (non osservato) al tempo t .

I metodi adottati per la stima dei flussi lordi poggiano:

- (a) sulla specificazione della funzione di probabilità congiunta del processo osservato e di quello vero

$$P(Y_1, Y_2, \dots, Y_T, y_1, y_2, \dots, y_T), \tag{1}$$

- (b) sulla marginalizzazione rispetto a quanto non è osservato, cioè

$$P(Y_1, Y_2, \dots, Y_T) = \sum_{y_1=1}^r \sum_{y_2=1}^r \dots \sum_{y_T=1}^r P(Y_1, Y_2, \dots, Y_T, y_1, y_2, \dots, y_T), \tag{2}$$

È quindi cruciale la specificazione della funzione di probabilità congiunta [1]. All'atto pratico, si tratta in sostanza di fare opportune assunzioni di indipendenza per le componenti del vettore $Y_1, Y_2, \dots, Y_T, y_1, y_2, \dots, y_T$, che consentano di pervenire alla [2] con un ridotto numero di parametri ignoti.

Un esempio notevole che si può ricondurre a tale specificazione è costituito dalla classe di modelli a catene di Markov latenti, già utilizzata nell'ambito della stima di flussi lordi contaminati da errori di misura da Van de Pol e Langeheine (1990). In tal caso, per una generica unità si specifica

$$q_t^{j_t l_t} = P(Y_t = j_t | y_t = l_t) \quad t = 1, 2, \dots, T, \tag{3}$$

$$\pi_t^{l_t l_{t-1}} = P(y_t = l_t | y_{t-1} = l_{t-1}) \quad t = 2, 3, \dots, T, \tag{4}$$

$$\pi_1^{l_1} = P(y_1 = l_1). \tag{5}$$

La [3] è relativa al legame tra stato osservato e stato vero, e in tal senso assolve alla funzione di modello di misura; la [4] specifica il modello di comportamento dinamico, in questo caso assunto markoviano; la [5] è la condizione iniziale del processo.¹ Di conseguenza, in accordo con la [2], si ottiene

$$P(Y_1 = j_1, Y_2 = j_2, \dots, Y_T = j_T) = \sum_{l_1=1}^r \sum_{l_2=1}^r \dots \sum_{l_T=1}^r \pi_1^{l_1} q_1^{j_1 l_1} \prod_{t=2}^T q_t^{j_t l_t} \pi_t^{l_t l_{t-1}}. \tag{6}$$

¹ Nel seguito, laddove non ci sia rischio di confusione, per semplicità, si ometteranno gli indici relativi ai valori assunti dalle variabili.

La struttura è quella di un modello a classi latenti, ove il ruolo di classe latente è svolto, per ogni istante temporale t , dalla vera condizione y e il ruolo di indicatore dalla condizione osservata Y . Vale l'assunzione, tipica dei modelli a classi latenti, di indipendenza locale fra gli indicatori: gli indicatori sono parzialmente indipendenti condizionatamente alle variabili latenti. Nel presente contesto, ciò è equivalente all'assunzione di errori di classificazione indipendenti (ECI), spesso introdotta nei modelli per la stima di flussi lordi in presenza di errori di misura. Tale assunzione è stata oggetto di fondate osservazioni critiche motivate dalla particolare natura dell'errore di misura in indagini longitudinali (in particolare, in indagini di tipo retrospettivo; si vedano Skinner e Torelli, 1993 e Singh e Rao, 1995).

Nel caso si ponga $T=2$, si possono ricondurre alla logica presentata in [1] e [2] anche i metodi di correzione dei flussi proposti da Abowd e Zellner (1985) e Poterba e Summers (1986), con riguardo alla stima dei flussi dell'indagine corrente sulle forze di lavoro statunitense, sfruttando la disponibilità di dati da reintervista. La struttura del modello rimane quella formulata dalle [3]-[5]; si ipotizza di disporre di stime esogene dei q_t , appunto dai dati della reintervista. Sostituendo le stime di q_t nella [2] e considerando le stime di $P(Y_1, Y_2)$ ottenute dai dati osservati, è possibile ottenere direttamente la stima dei flussi lordi corretta.

Alla formulazione generale [1]-[2] possono essere ricondotti altri modelli proposti in letteratura. Ad esempio, al posto della [3]-[4], Pfefferman, Skinner e Humphreys (1994) specificano

$$q_t^{jm} = P(Y_t = j | y_t = l, Y_{t-1} = m) \quad t=2,3,\dots,T, \quad [7]$$

$$\pi_t^{lj} = P(y_t = l | y_{t-1} = j) \quad t=2,3,\dots,T. \quad [8]$$

Tale formulazione, com'è evidente, supera l'assunzione ECI. Con l'opportuna specificazione della condizione iniziale, è poi possibile scrivere la probabilità congiunta [1] in funzione dei nuovi parametri, e ricorrere alla [2].

In modo analogo, Van de Pol e Langeheine (1992) e Regoli (1995) cercano di superare l'incorrelazione degli errori di classificazione e di estendere la dinamica che si suppone descritta da una catena di Markov di secondo ordine, specificando

$$q_t^{jm} = P(Y_t = j | y_t = l, y_{t-1} = m) \quad t=2,3,\dots,T, \quad [9]$$

$$\pi_t^{ljm} = P(y_t = l | y_{t-1} = j, y_{t-2} = m) \quad t=3,4,\dots,T. \quad [10]$$

Infine, è possibile considerare specificazioni ancor più generali, che prevedano, ad esempio, che lo stato osservato al tempo t dipenda dallo stato osservato al tempo $t-1$ e dallo stato vero in t e $t-1$.

Il modello può poi essere esteso in più direzioni. In particolare, è conveniente introdurre variabili ausiliarie X che si suppone influenzino sia i parametri q che i parametri π . L'uso di informazioni ausiliarie è stato considerato da Pfefferman, Skinner e Humphreys (1994), che specificano la relazione fra q e π e le variabili X mediante modelli di regressione logistica. Van de Pol e Langeheine (1990) hanno considerato variabili X qualitative, con il ruolo di variabili di stratificazione. L'uso di variabili ausiliarie pare quanto mai opportuno: spesso occorre considerare popolazioni non omogenee, sicché è ragionevole supporre che, per soggetti con caratteristiche diverse, possano permanere differenze nella dinamica temporale e/o che possano esservi differenti propensioni all'errore di classificazione.

Un'ulteriore estensione del modello si ottiene se si ipotizza la presenza di eterogeneità non osservata. È il caso dei modelli di Markov latenti misti, in cui si ipotizza l'esistenza di due o più sottopopolazioni che obbediscono a processi di Markov differenti, ma la variabile che permette di identificare le sottopopolazioni non è osservata (un esempio estremo è il cosiddetto modello *mover-stayer* ove la popolazione è ripartita in due gruppi, e tutti i movimenti avvengono all'interno di uno solo di essi).

Una volta specificato il modello si pongono tre problemi: l'identificazione, la stima e la verifica dell'adattamento. Per i modelli descritti sopra, ad esempio, non è automaticamente assicurata l'identificabilità. In genere, è necessario ricorrere: (i) a restrizioni di uguaglianza fra insiemi di parametri (ad esempio, omogeneità della catena markoviana o costanza nel tempo delle relazioni di misura); (ii) all'impiego di ulteriori indicatori per la variabile latente y . Allo stato, non sono disponibili regole generali per l'identificazione. Esistono risultati per alcune particolari classi di modelli; spesso, conviene verificare almeno l'identificazione locale.

La stima dei parametri incogniti, se si dispone di un insieme di n osservazioni indipendenti su Y , può in generale essere ottenuta tramite massimizzazione della verosimiglianza, desunta in modo diretto dalla [2]. Questa, difatti, rappresenta il contributo alla verosimiglianza per la generica unità. Conviene, tuttavia, ricondurre il problema alla stima di modelli con dati mancanti. Il ruolo di dato mancante è assunto dallo stato vero. In tal caso si può ricorrere all'algoritmo EM che in genere risulta vantaggioso dal punto di vista computazionale.

La verifica dell'adattamento costituisce un problema aperto. A causa della particolare struttura delle matrici di flussi lordi, che presentano spesso frequenze elevate lungo la diagonale e frequenze basse al di fuori di essa, è precluso l'impiego di strumenti classici tipo L^2 e X^2 poiché non è possibile fare riferimento agli usuali risultati asintotici. È necessario usare criteri diversi quali, ad esempio, l'AIC o il BIC, oppure fare ricorso a tecniche di simulazione o di ricampionamento (McLachlan, 1987).

2.3. L'approccio LISREL modificato

I modelli a classi latenti possono essere equivalentemente rappresentati da modelli log-lineari in cui una o più variabili non siano direttamente osservabili. La formulazione in termini di modello log-lineare della specificazione [3]-[5], ad esempio per tre occasioni temporali, è:

$$\begin{aligned} \log P(y_1, y_2, y_3, Y_1, Y_2, Y_3) = \\ = \lambda + \lambda^{y_1} + \lambda^{y_2} + \lambda^{y_3} + \lambda^{Y_1} + \lambda^{Y_2} + \lambda^{Y_3} + \lambda^{y_1 Y_1} + \lambda^{y_2 Y_2} + \lambda^{y_3 Y_3} + \lambda^{y_1 y_2} + \lambda^{y_2 y_3} \end{aligned} \quad [11]$$

o in modo sintetico, nella notazione per modelli log-lineari gerarchici che riporta solo i termini di interazione di ordine superiore: $(y_1 Y_1, y_2 Y_2, y_3 Y_3, y_1 y_2, y_2 y_3)$

I modelli log-lineari consentono però di ipotizzare solo relazioni di tipo simmetrico tra le variabili; non è quindi possibile tenere conto di un eventuale ordinamento di causalità tra di esse. Nelle specificazioni presentate sopra, è invece implicito un ordinamento causale, dato ad esempio dall'assunzione di un processo Markoviano (ma analoga considerazione vale anche per particolari assunzioni sul modello di misura).

L'approccio LISREL modificato proposto da Hagenaars (1990), estendendo i *modified path models* di Goodman (1973) a situazioni in cui siano presenti anche variabili

latenti, consente di stimare l'intensità delle associazioni tra le variabili tenendo in considerazione l'ordinamento di causalità esistente tra di esse. Al modello log-lineare [11] viene sostituito un sistema di equazioni di tipo logit multinomiale, la cui stima e verifica dell'adattamento avviene per passi successivi. Per ogni variabile dipendente, si costruisce una sottotabella marginale contenente la variabile in questione e tutte le variabili indipendenti che la precedono nell'ordinamento causale (in altre parole, la tabella completa viene collassata rispetto a tutte le altre variabili del modello). Su ogni sottotabella si stima un opportuno modello logit; le stime ad ogni passo vengono opportunamente ricombinate per ottenere una stima per il modello completo (Goodman, 1973). L'adattamento si può valutare sommando i valori del rapporto di log-verosimiglianza ottenuti ad ogni passo, oppure con l'usuale confronto tra frequenze stimate ed osservate per la tabella completa. Nei modelli visti in precedenza i vari passaggi vengono effettuati coinvolgendo anche variabili non osservabili. Risulta quindi opportuno procedere alla stima mediante un'opportuna implementazione dell'algoritmo iterativo E-M.

A scopo esemplificativo, si riportano i passi corrispondenti al caso di un modello markoviano latente implicato dalla [6] per tre occasioni temporali:

(y_1, y_2) per la tabella $y_1 \times y_2$;

(y_1, y_2, y_2, y_3) per la tabella $y_1 \times y_2 \times y_3$;

$(y_1, y_2, y_3, y_1, y_2, y_2, y_3, y_3)$ per la tabella completa.

Secondo questa logica, è possibile fornire, ad esempio, una espressione equivalente in termini di modelli LISREL modificati, per ciascuna delle specificazioni viste nella sez. 2.2. In particolare, si possono considerare modelli che superano l'assunzione di indipendenza locale, sulla quale si basano i modelli a classi latenti, con l'introduzione di effetti diretti tra gli indicatori. Gli effetti diretti consentono di rappresentare quella parte di associazione esistente tra le variabili osservate che non è completamente spiegata dalle variabili latenti ed è dovuta, ad esempio, ad errori di classificazione serialmente correlati.

È ovvio che l'introduzione di effetti diretti tra gli indicatori, a causa di una più ricca parametrizzazione pone, in genere, problemi di identificabilità: in particolare, se $T \geq 2$, è sufficiente disporre di almeno due indicatori per ciascuna variabile latente affinché, sia assicurata l'identificazione (Bassi, 1995).

3. La stima dei flussi con la RTFL e un'indagine suppletiva alla stessa

La principale fonte di informazioni sul mercato del lavoro in Italia è la Rilevazione Trimestrale sulle Forze di Lavoro (RTFL) condotta dall'Istat. Si tratta di un'indagine campionaria condotta nei mesi di gennaio, aprile, luglio e ottobre. Essa prevede un piano di campionamento ruotato del tipo 2-2-2, ovvero ogni unità che entra a far parte del campione viene reintervistata nel trimestre successivo, esce dal campione per due trimestri e viene ancora intervistata per altri due. L'indagine, fra l'altro, attraverso un' articolata serie di quesiti, consente di rilevare la posizione dell'intervistato riguardo al lavoro avendo come periodo di riferimento la settimana precedente l'intervista.

Nel maggio 1989, alla RTFL corrente è stata occasionalmente affiancata un'indagine suppletiva (Trivellato *et al.*, 1991) su un sottocampione di famiglie lombarde che partecipavano all'indagine corrente per l'ultima volta nel II trimestre del 1989 (per tali famiglie si dispone, almeno teoricamente, di informazioni rilevate nei primi due trimestri del 1988 e del 1989). Con l'indagine suppletiva sono state raccolte, tra l'altro, informazioni con

cadenza mensile sullo stato lavorativo dei rispondenti nei 17 mesi precedenti (da gennaio 1988 a maggio 1989 compresi).

Vale la pena descrivere brevemente la struttura della parte del questionario dell'indagine suppletiva nella quale si pongono i quesiti retrospettivi sulla partecipazione al mercato del lavoro. Il questionario è costituito da una serie di sette domande e da un griglia/calendario riassuntiva della posizione del rispondente nei singoli mesi a seconda che in ciascuno egli abbia lavorato almeno alcuni giorni, non abbia lavorato ma abbia cercato attivamente un lavoro, o abbia svolto entrambe queste attività. Le sette domande costituiscono dei filtri, ovvero non vengono poste a tutti i rispondenti: brevemente, chi dichiara situazioni più stabili si ferma prima e compila in modo uniforme la griglia riassuntiva. Ad esempio, chi dichiara di avere lavorato con continuità negli ultimi 17 mesi, passa direttamente a compilare la griglia riassuntiva e alle domande della sezione successiva; chi dichiara di aver lavorato solo per alcuni periodi deve sottostare ad alcune ulteriori domande. Con le informazioni raccolte mediante le due indagini è possibile classificare ciascun rispondente secondo l'abituale ripartizione: occupato (O), disoccupato (D) e inattivo (I).

L'abbinamento dei dati dell'indagine RTFL corrente con quelli dell'indagine suppletiva consente quindi di disporre, per il sottocampione in questione e per i mesi di riferimento della RTFL, di due informazioni indipendenti sulla condizione lavorativa dei rispondenti e sulla dinamica del mercato del lavoro in Italia.

3.1. Il problema degli errori di misura nella RTFL corrente e nell'indagine suppletiva

La Tabella 1 contiene le transizioni trimestrali osservate, per il periodo compreso tra luglio 1988 e aprile 1989, tra le condizioni di occupato, disoccupato e inattivo. La posizione nel mercato del lavoro nei mesi di luglio ed ottobre è stata osservata con la sola indagine suppletiva (SUPPL), mentre per gennaio ed aprile si dispone di informazioni raccolte anche con la RTFL. Il campione comprende 6768 unità. Le transizioni verso i diversi stati di destinazione sono espresse come percentuale sul totale delle unità con lo stesso stato di origine.

Tabella 1 – Proporzioni osservate (%) secondo il tipo di indagine (RTFL e SUPPL)

tipo ind.		transizione								
		OO	OD	OI	DO	DD	DI	IO	ID	II
LUG-OTT	SUPPL	98.71	0.47	0.82	16.81	81.42	1.77	1.36	0.86	97.78
OTT-GEN	SUPPL	97.84	0.69	1.47	19.30	75.44	5.26	0.99	0.48	98.53
GEN-APR	SUPPL	98.38	0.72	0.90	16.19	78.57	5.24	1.04	0.29	98.67
GEN-APR	RTFL	96.52	0.85	2.63	20.99	62.98	16.03	2.39	0.97	96.64

Confrontando le ultime due righe della tabella, si vede chiaramente che le transizioni osservate con la RTFL descrivono un mercato del lavoro più dinamico che non le transizioni osservate con l'indagine suppletiva, per la quale la proporzione di unità che non cambiano stato, per tutti e tre gli stati, è sistematicamente più elevata.

Queste differenze sono, in buona misura, imputabili alla diversa struttura delle due indagini - corrente l'una (quindi con osservazioni ripetute), e retrospettiva l'altra, - e alla diversa incidenza nei due casi degli errori di classificazione. Nell'indagine suppletiva, in particolare, è lecito attendersi che i rispondenti commettano errori dovuti a cattiva memoria, ovvero la completa omissione o l'errato posizionamento degli eventi nel tempo. Inoltre per la natura retrospettiva della rilevazione e, probabilmente anche per la struttura del que-

stionario descritta sopra, è plausibile che si verifichino fenomeni di telescopio ed, eventualmente di trascinamento, legati alla tendenza da parte dei rispondenti a ripetere meccanicamente le risposte fornite. Questi fenomeni possono dare origine alla correlazione seriale degli errori di misura.

Anche per l'indagine corrente, tuttavia, può risultare scorretto assumere l'indipendenza degli errori (a questa conclusione giunge, ad esempio, Regoli, 1995). Si noti inoltre che il meccanismo di rilevazione della condizione rispetto al lavoro è diverso nelle due indagini: nella suppletiva si tratta di una autodichiarazione con riferimento al mese di calendario.

A supporto delle osservazioni fatte, può essere utile confrontare le classificazioni degli individui nei tre stati del mercato del lavoro ottenute con le due indagini, per i mesi di riferimento della RTFL (gennaio e aprile 1988 e gennaio ed aprile 1989). Una prima evidenza è che le discrepanze fra le due fonti aumentano andando a ritroso nel tempo. Ciò autorizza ad ipotizzare perlomeno la presenza di un effetto memoria nell'indagine suppletiva, ovvero è molto probabile che, andando indietro nel tempo, i rispondenti commettano più errori nel dichiarare la propria posizione nel mercato del lavoro. Un effetto analogo, seppure in relazione ad un diverso disegno di indagine, viene ipotizzato da Magnac e Visser (1996).

Se si analizzano in modo un po' più approfondito le risposte fornite dalle unità classificate in uno stato diverso per lo stesso mese di riferimento con le due indagini si osserva che lo stato per il quale si verificano maggiori discrepanze, in percentuale, è quello di disoccupato. Si veda, a questo proposito la Tabella 2 che contiene la classificazione delle unità rispetto alla posizione nel mercato del lavoro nel mese di gennaio 1989 secondo le due fonti.

Tabella 2 – Classificazione delle unità secondo i tre stati nel mercato del lavoro nel mese di gennaio 1989 con le due indagini (RTFL e SUPPL)

RTFL	SUPPL			TOT
	O	D	I	
O	3097 (94.42)	44 (1.34)	139 (4.24)	3280 (100.00)
D	23 (12.70)	107 (59.12)	51 (28.18)	181 (100.00)
I	87 (2.63)	58 (1.75)	3162 (95.62)	3307 (100.00)

I dati in tabella non consentono ovviamente di valutare quale delle due informazioni sia da considerarsi più attendibile, tuttavia l'esame dei casi che forniscono informazioni discordanti nelle due rilevazioni, per i quali è certamente presente in almeno una delle indagini un errore di classificazione, è utile per trarre idee sulla plausibilità di assunti riguardo al meccanismo di misura.

Concentriamo l'attenzione, a mo' di esempio, sulle discordanze rispetto alla condizione di disoccupato rilevata con la RTFL. Dei 23 rispondenti che hanno dichiarato di essere occupati nel mese di gennaio 1989 durante la rilevazione suppletiva, ben 21 avevano dichiarato la stessa posizione nel mese di febbraio, e solo 2 la posizione di disoccupato. Inoltre 7 di essi hanno riportato di avere lavorato con continuità in tutto il periodo coperto dall'indagine suppletiva, senza dunque aver vissuto alcun episodio di disoccupazione; tra i restanti 16, molti dichiarano di avere sperimentato la disoccupazione nei mesi immediatamente precedenti a gennaio; sono 5, ad esempio, nel mese di novembre e 7 nel mese di ottobre. Un comportamento simile si osserva per i 51 rispondenti che si sono dichiarati

inattivi nel mese di gennaio durante l'indagine suppletiva; 47 di loro, inoltre, si sono dichiarati inattivi anche in febbraio, 2 si sono dichiarati occupati e 2 disoccupati.

Ad analoghe conclusioni, si giunge se si considerano i dati relativi agli altri mesi del periodo coperto dalla rilevazione suppletiva. Non si può escludere dunque che nell'indagine suppletiva siano presenti effetti di telescopio e di trascinamento. Tali effetti hanno origine dalla natura retrospettiva della rilevazione e dalla struttura del questionario che probabilmente induce i rispondenti a dichiarare di avere sperimentato situazioni di stabilità in tutto il periodo di riferimento, trascurando eventuali brevi episodi di cambiamento. Questo vale in particolare per coloro che rimangono occupati per la maggior parte del tempo. Ciò è equivalente ad affermare che l'ipotesi di errori di misura indipendenti nell'indagine suppletiva è poco plausibile.

3.2. Un modello per la stima di flussi corretti da errori di misura

Il problema della stima dei flussi in presenza di errori di classificazione, ed in particolare, di errori di classificazione correlati, può essere risolto mediante la specificazione di opportuni modelli a classi latenti. La flessibilità di questo approccio consente di rappresentare la correlazione tra gli errori di misura mediante effetti diretti tra gli indicatori e di specificare poi il meccanismo generatore di questi errori, ovvero le ipotesi che si fanno sul comportamento dei rispondenti, mediante una serie di vincoli sui suoi parametri.

Siano Y_1, Y_2, Y_3 e Y_4 le variabili che denotano lo stato osservato nelle quattro occasioni temporali dell'indagine suppletiva (luglio e ottobre 1988 e gennaio e aprile 1989) e W_1 e W_2 le variabili che denotano lo stato osservato nelle due occasioni dell'indagine corrente (gennaio ed aprile 1989). La struttura di base del modello utilizzato si ottiene specificando in modo opportuno il legame di dipendenza tra le variabili osservate e le latenti. Nel presente caso si è considerata la seguente specificazione per le transizioni a livello latente: l'ipotesi di markovianità del secondo ordine. A tal fine si definiscono:

$$\pi_1^l = P(y_1 = l) \quad [12]$$

$$\pi_2^{lm} = P(y_2 = l | y_1 = m) \quad [13]$$

$$\pi_t^{lijm} = P(y_t = l | y_{t-1} = j, y_{t-2} = m) \quad t=3, 4. \quad [14]$$

La [12] e la [13] sono le condizioni iniziali delle catena di Markov del secondo ordine, mentre la [14] si riferisce alle probabilità di transizione. Si è inoltre assunto che la catena di Markov sia stazionaria, ovvero che i parametri della [14] non dipendano da t^2 .

Per quanto riguarda invece il modello di misura, si specificano:

$$q_{wt}^{jlm} = P(W_t = j | y_t = l, y_{t-1} = m) \quad t=3, 4 \quad [15]$$

che rappresentano le probabilità di risposta per l'indagine corrente. Il modello assume che queste probabilità dipendano dalla transizione latente tra $t-1$ e t , e che, quindi, anche per l'indagine corrente non vi sia indipendenza degli errori. Inoltre si specificano

$$q_{y_4}^{jl} = P(Y_4 = j | y_4 = l) \text{ e} \quad [16]$$

$$q_{y_t}^{jlmk} = P(Y_t = j | y_t = l, y_{t+1} = m, Y_{t+1} = k) \quad t=1, 2, 3 \quad [17]$$

² Si noti che, essendo quattro i periodi considerati, ciò equivale a dire che la catena del secondo ordine è omogenea per i mesi di aprile e gennaio.

le probabilità di risposta nell'indagine suppletiva che dipendono, non solo dalla transizione vera, ma anche dallo stato dichiarato per l'istante precedente, andando a ritroso nel tempo.³ Questi legami causali tra gli indicatori descrivono l'effetto di trascinamento. Sono stati poi imposti dei vincoli su alcune delle probabilità di risposta [16] che si riferiscono all'indagine suppletiva per tenere conto dell'effetto telescopio. In particolare, se l'errore del rispondente consiste nell'aver anticipato un cambiamento di stato, nell'istante successivo si suppone che lo stesso rispondente non commetta più errori.

Si è, infine, introdotta un'ulteriore variabile latente facendo l'ipotesi che la popolazione possa essere suddivisa in due gruppi, non direttamente osservabili, con comportamenti diversi anche per quel che riguarda gli errori di misura: coloro che non cambiano mai veramente la loro posizione nel mercato del lavoro (*stayers*) e coloro che invece hanno sperimentato della mobilità (*movers*).

Tabella 3 – Stime dei flussi (%)

	transizione								
	OO	OD	OI	DO	DD	DI	IO	ID	II
LUG-OTT	98.36	0.47	1.17	21.43	77.02	1.55	1.13	1.66	97.21
OTT-GEN	97.78	1.0	1.19	19.03	72.02	8.95	0.92	0.37	98.71
GEN-APR	97.90	0.95	1.15	18.26	78.69	3.05	0.93	0.43	98.64

La Tabella 3 contiene le transizioni stimate con il modello. Dal confronto tra le stime della Tabella 3 e le proporzioni osservate della Tabella 1 emerge che le transizioni osservate con le due indagini sono quasi sempre corrette in direzioni opposte, coerentemente con l'assunto che tali differenze siano da imputare alla presenza di errori di classificazione serialmente correlati in modo differente nella RTFL e nella SUPPL. Per valutare l'adattamento del modello si è simulata la distribuzione della statistica L^2 mediante tecniche di ricampionamento Monte Carlo. Il valore della statistica, ottenuto nel campione originario è risultato pari a 532.5478. Esso è stato valutato in relazione alla distribuzione simulata ottenendo un livello di significatività osservato inferiore allo 0.05.

4. La stima dei flussi lordi utilizzando i dati correnti della RTFL

Nelle varie esemplificazioni considerate per illustrare le potenzialità dei modelli a classi latenti, ed in particolare dei modelli tipo LISREL modificato, per la stima di flussi lordi in presenza di errori si è assunta la disponibilità di informazioni aggiuntive (ad esempio più indicatori dello stato osservato o dati di validazione) e/o che fosse disponibile una successione piuttosto lunga di informazioni longitudinali (ad esempio nel caso di modelli markoviani latenti). Ciò permette di giungere a specificazioni che tengano conto di strutture complesse per l'errore di misura (senza ricorrere all'assunzione ECI) e di evitare ipotesi troppo stringenti di stazionarietà o di markovianità del primo ordine.

Le informazioni che sono correntemente disponibili nel caso italiano, quelle provenienti dalla RTFL, sembrano quindi offrire minori opportunità per stimare flussi corretti in presenza di errori di misura. Si ricordi, con riferimento alla stima di flussi trimestrali, che il piano di rotazione prevede che una parte delle unità campionarie sia seguita solo

³ I pedici w e y nelle formule [14], [15] e [16] stanno ad indicare che la prima definisce le probabilità di risposta per lo stato osservato con l'indagine corrente (W_p), mentre le altre due definiscono le probabilità di risposta per lo stato osservato con l'indagine suppletiva (Y_p).

per due trimestri successivi. Tuttavia, nel caso della RTFL, è possibile sfruttare adeguatamente:

1. la disponibilità di una sequenza di matrici di flusso, seppure queste siano osservate su insiemi di unità diverse;
2. l'uso di potenziali variabili di stratificazione, considerando la possibilità di imporre vincoli di uguaglianza fra parametri relativi ai diversi gruppi così identificati;
3. la presenza, nell'indagine di aprile nell'ambito dell'inchiesta comunitaria, di un quesito retrospettivo con il quale si rileva la condizione dell'aprile dell'anno precedente;
4. la presenza di un eventuale secondo indicatore della condizione di disoccupazione costituito dal quesito relativo all'iscrizione all'ufficio di collocamento; tale quesito è posto all'esterno del blocco di quesiti che vengono utilizzati per definire la condizione di disoccupato e non viene utilizzato a tale scopo.

Nel seguito, limitando per ora l'attenzione alla stima dei flussi trimestrali, si discuterà la specificazione di modelli che consentano di utilizzare, eventualmente in modo combinato, tali opportunità. Tali modelli potrebbero eventualmente essere impiegati correntemente per fornire prime valutazioni sull'effetto dell'errore di classificazione e ottenere, sotto assunzioni la cui ragionevolezza va giudicata con attenzione, stime dei flussi corrette.

Lo schema di rotazione è tale per cui per le unità di un ipotetico gruppo di rotazione A si osservano i flussi relativi ai due successivi istanti t_1 e t_2 , sulle unità di un secondo gruppo di rotazione B si osservano i flussi tra t_2 e t_3 , mentre sulle unità di un altro gruppo C si osservano i flussi tra t_3 e t_4 . Per il particolare disegno di rotazione (2-2-2), su un quarto gruppo D si osservano i flussi tra t_1 e t_4 . Se si guarda alla sequenza delle matrici così costruite, la condizione rispetto al lavoro al tempo t_1 , può essere considerata come una variabile con osservazioni mancanti per unità appartenenti ad altri gruppi (B e C), e un analogo argomento vale per la stessa variabile osservata nei successivi istanti di tempo.

Si noti che l'approccio per modelli a classi latenti può essere facilmente esteso per includere problemi legati alla presenza di dati mancanti; d'altra parte si è visto che la variabile latente è formalmente trattata come un dato mancante. Si tratta cioè di specificare comunque la verosimiglianza mediante marginalizzazione rispetto al dato mancante. In una specificazione di tipo LISREL modificato, si costruiscono ad ogni passo le statistiche sufficienti rilevanti per il modello (ovvero una tabella di contingenza) supponendo di utilizzare tutte le informazioni, provenienti dai rispondenti e dai non rispondenti. La procedura di stima dei parametri sarà ovviamente più complessa per tenere conto dei dati mancanti. Le stime di massima verosimiglianza possono essere quindi calcolate mediante una procedura di tipo EM che comprende un passo di E leggermente più elaborato.

Tuttavia l'uso di tali modelli nel presente caso, essendo fondato su informazioni più povere rispetto alle specificazioni considerate nelle precedenti sezioni, rende indispensabile l'adozione di assunzioni di regolarità sugli errori nel tempo e sulla natura dei flussi veri, allo scopo di ottenere l'identificabilità. È comunque possibile, nel caso dei dati della RTFL, sfruttare una serie piuttosto lunga di rilevazioni ed il fatto che le unità di ciascun gruppo di rotazione vengono intervistate in quattro occasioni successive seppur a distanza di nove mesi tra la seconda e la terza.

Con i dati relativi a quattro occasioni di indagine, pare quindi possibile stimare un modello a classi latenti markoviano classico come quello presentato nelle equazioni [3]-[5].

Si noti poi che, utilizzando una serie più lunga di osservazioni temporali e inserendo nel modello ulteriore eterogeneità osservata, potrebbe essere possibile considerare ipotesi meno restrittive anche per il processo di misura. La concreta praticabilità di questo approccio è stata

verificata ricorrendo ad esercizi di simulazione orientati, in questa fase, solo a verificare se il modello fosse operativamente proponibile. Seppure non sia opportuno trarre da queste evidenze conclusioni con carattere di generalità, l'adozione di tale modello appare promettente.

Se si considera il problema di stimare flussi annuali corretti, il quesito retrospettivo posto nel mese di aprile sulla condizione lavorativa dell'anno precedente è di qualche interesse. Esso infatti può essere utilizzato come un secondo indicatore della condizione lavorativa e permette di specificare modelli più generali soprattutto riguardo alla struttura degli errori di classificazione.

Va infine valutata la possibilità di ricavare dal medesimo questionario, usato per l'indagine corrente, altre informazioni relative alla condizione lavorativa; in particolare, occorre verificare se si possa ricorrere al dato relativo all'iscrizione all'ufficio di collocamento. Come già detto, tale quesito è posto a tutti i rispondenti ed è nel questionario all'esterno dell'insieme di quesiti utilizzati per definire la condizione di disoccupato. Un'attenta analisi esplorativa di insiemi di dati reali della RTFL è quindi indispensabile per trarre qualche prima indicazione sulla concreta possibilità di sfruttare utilmente tale informazione.

5. Conclusioni

Vale la pena di ricordare che i problemi posti dalla presenza di errori di classificazione non sono i soli aspetti metodologici connessi alla stima dei flussi lordi rispetto al lavoro. Altre classi di problemi cui è necessario far fronte riguardano: (i) la definizione della popolazione di riferimento; (ii) la presenza di non risposte, complicata dalla struttura longitudinale dell'indagine; (iii) la procedura di abbinamento, necessaria per ricostruire dati longitudinali a partire da indagini con campione ruotato; (iv) il calcolo dei pesi di campionamento.

In parte tali problemi possono interagire con quelli derivanti dalla presenza di errori di classificazione e ciò dà la misura di quante cautele siano necessarie nella stima dei flussi lordi. Sebbene il problema della stima dei flussi lordi includa questioni non risolte adeguatamente, è comunque unico il contributo che deriva dalle informazioni che essi contengono. Risulta pertanto di estrema utilità proporre convincenti soluzioni ad almeno alcuni di questi problemi. La presenza di errori di misura, anche in considerazione delle difficoltà che insorgono quando si voglia classificare la popolazione con riferimento alla sua condizione rispetto al lavoro, è certamente centrale.

Nel testo si è mostrato come la stima dei flussi lordi rispetto al lavoro in presenza di errori di classificazione può essere affrontata efficacemente nell'ambito dei modelli a classi latenti. In particolare, se si considerano modelli LISREL modificati, è possibile estendere notevolmente le generalità e l'ambito di applicazione di tali modelli, essendo più agevole introdurre assunzioni e vincoli che derivano dalla particolare natura del problema in esame.

Ad oggi, tale approccio soffre peraltro dell'assenza di criteri generali per l'identificazione dei modelli; mancano inoltre criteri statistici che consentano un'agevole valutazione dell'adattamento e, conseguentemente, risulta difficile la formulazione di strategie per la ricerca di specificazione. Si ricorda infatti che la tecnica più convincente per valutare l'adattamento risulta essere quella di utilizzare opportunamente metodi di ricampionamento.

Con riguardo alla possibilità di adottare modelli di tale natura nel caso italiano, si è visto come sia possibile pervenire a specificazioni di qualche interesse anche disponendo solo dell'informazione corrente. Tuttavia non si può dimenticare che i risultati che ne derivano sono fortemente condizionati alla correttezza delle assunzioni che si fanno. La validità di queste può essere almeno occasionalmente esaminata se si dispone di informazioni alternative, in particolare:

- le indagini suppletive emergono come uno strumento estremamente valido; tuttavia, il carattere di occasionalità che necessariamente le caratterizza e i costi che possono derivare dalla loro gestione possono costituire un limite alla adozione. La ripetizione di tali esperienze, magari con formato ridotto e con obiettivi più circoscritti va comunque considerata con interesse;

- è auspicabile che vengano progettate con qualche regolarità esperienze limitate di programmi di reintervista. Nonostante si abbia consapevolezza che i dati che si ottengono non conducono automaticamente alla soluzione del problema, le informazioni che si ricavano consentono quantomeno di controllare la validità delle assunzioni riguardo gli errori di misura. D'altra parte tale tipo di esperienze potrebbe risultare agevolato dal crescente utilizzo di tecniche di rilevazione assistite con il computer;
- è conveniente considerare il ricorso, nel questionario correntemente utilizzato, a quesiti disegnati anche con l'obiettivo di ottenere informazioni utili per valutare l'attendibilità delle classificazioni rispetto al lavoro. Ad esempio, può essere utile estendere il ricorso all'interrogazione retrospettiva.

References

- ABOWD J.M., ZELLNER A. (1985), "Estimating gross labor force flows", *Journal of Business and Economic Statistics*, 3, pp.254-283.
- BASSI F. (1995), *Errori di misura correlati in dati longitudinali e analisi a classi latenti*, Tesi di Dottorato, Università di Padova.
- CHUA T.C., FULLER W.A. (1987), "A model for multinomial response error applied to labor flows", *Journal of the American Statistical Association*, 82, pp.46-51.
- DAVID M.H. (1995), "Introduction: the design and development of SIPP", *Journal of Economic and Social Measurement*, 13, pp.215-224.
- GOODMAN L.A. (1973), "The analysis of a multidimensional contingency table when some variables are posterior to the others", *Biometrika*, 60, pp. 179-192.
- HAGENAARS J.A. (1990), *Categorical Longitudinal Data: Log-Linear Panel, Trend and Cohort Analysis*, Newbury Park, Sage.
- MAGNAC T., VISSER M. (1995), "Transition models with measurement errors", Institut National de la Recherche Agronomique (INRA), Inry-sur-Seine, mimeo.
- MCLACHLAN G.J. (1987), "On bootstrapping the likelihood ratio test statistics for the number of components in a normal mixture", *Applied Statistics*, 36, pp.318-324.
- MEHRAN F. (1990), "Labour force flow statistics: an examination of objectives, collecting methods and measurement issues", in Turvey R. (ed.), *Developments in International Labour Statistics*, pp. 189-202, London, Printer Publishers.
- PFEFFERMAN D., SKINNER C.J., HUMPHREY S K. (1998), "The estimation of gross flows in the presence of measurement error using auxiliary variables", *Journal of the Royal Statistical Society, Series A*, 161, pp. 13-32.
- POTERBA J.M., SUMMERS L.H. (1986), "Reporting errors and the labor market dynamics", *Econometrica*, 54, pp.1319-1338.
- REGOLI A. (1995), "L'analisi delle transizioni nel mercato del lavoro in presenza di errori di classificazione", *Economia e Lavoro*, XXVII, 3-4, pp.35-52.
- SINCLAIR M., GASTWIRTH L. (1996), "On procedures for evaluating the effectiveness of reinterview survey methods: application to labor force data", *Journal of the American Statistical Association*, 91, pp.961-969.
- SINGH A.C., RAO J.N.K (1995), "On the adjustment of gross flows estimates for classification errors with application to data from the Canadian Labor Force Survey", *Journal of the American Statistical Association*, 90, pp.1-11.
- SKINNER C., TORELLI N. (1993), "Measurement error and the estimation of gross flows from longitudinal economic data", *Statistica*, 3, pp.391-405.
- VAN DE POL F., LANGEHEINE R. (1990), "Mixed Markov latent class models", in: C.Clogg (ed.) *Sociological Methodology*, New York, Blackwell, pp.213-247.
- VAN DE POL F., LANGEHEINE R. (1992), "Analysing measurement error in quasi-experimental data: an application of latent class models to labour market data", *Working Papers of the European Scientific Network on Household Panel Studies*, 57, Colchester, University of Essex.
- TRIVELLATO U., DE NICOLA I., DI PIETRO E., GHELLINI G., RETTORE E., TORELLI N. (1991), "Un'indagine suppletiva alla rilevazione sulle forze di lavoro incentrata sulla storia lavorativa", in: U.Trivellato (ed.), *Forze di Lavoro: Disegno dell'Indagine e Analisi Strutturali*, Annali di Statistica, Serie IX, Vol. 11, Roma, ISTAT.

Serie Annali di Statistica - Volumi pubblicati

Anno 1996

L'Attività del Consiglio Superiore di Statistica dal 1949 al 1989 - serie X - Vol. 3 (Ristampa)

L'Istat e il Consiglio Superiore di Statistica dal 1926 al 1945 - serie X - Vol. 8

Proceedings of the Second International Forum on Tourism Statistics - Venice, May 30 - June 2 / 1995 - serie X - Vol. 9

L'informazione statistica per il governo dell'ambiente - Atti del Convegno - Roma 29-30 gennaio 1996 - serie X - Vol. 10

Verso il nuovo sistema di contabilità nazionale - serie X - Vol. 11

Indici degli Annali di statistica - anni 1871-1996 - serie X - Vol. 12

Contabilità ambientale - serie X - Vol. 13

Anno 1997

Statistica, storia e nazione: la statistica ufficiale tra passato e futuro - una prospettiva comparata - serie X - Vol. 14

Anno 1998

La misurazione delle variabili economiche e i suoi riflessi sulla modellistica econometrica - serie X - Vol. 15

