Dicembre 1984

44

Servizio Studi
della
Banca d'Italia

TEMI DI DISCUSSIONE

Carlo COTTARELLI

Regressioni lineari con "panel data": una guida alla letteratura

REGRESSIONI LINEARI CON "PANEL DATA":

UNA GUIDA ALLA LETTERATURA

di

Carlo Cottarelli

Questo lavoro presenta una rassegna della letteratura econometrica sull'uso di dati in panel (serie storiche di cross section) per la stima di modelli lineari. Vengono considerati sia i modelli con intercetta variabile, sia quelli in cui tutti i coefficienti differiscono tra individui. Entrambe le ipotesi usualmente adottate sulla natura (stocastica o non stocastica) dei coefficienti vengono discusse. Particolare rilievo viene dato all'esame dei problemi di stima in regressioni dinamiche.

La serie dei "Temi di discussione" intende promuovere la circolazione, in versione provvisoria, di lavori prodotti all'interno della Banca d'Italia o presentati da economisti esterni nel corso di seminari presso l'Istituto, al fine di suscitare commenti critici e suggerimenti. I lavori pubblicati nella serie riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto.

REGRESSIONI LINEARI CON "PANEL DATA": UNA GUIDA ALLA LETTERATURA (*)

> "The origin of the random-effects models like that of the fixed-effects models, lies in astronomical problems; statisticians re-invented random-effects models long after they were introduced by astronomers, and then developed more complicated ones."
>
> (H. Scheffé)

1 - Introduzione

Con il termine "panel data" (o "longitudinal data") ci si riferisce generalmente ad osservazioni su vari individui relative a istanti temporali diversi. L'econometria dei dati in panel, attraverso l'uso combinato (pooling) di tali osservazioni, si occupa della stima di modelli del tipo seguente:

(1) $y_{it} = f(X_{it}, \pi_{it}, E_{it}) \quad i=1, 2,...,N \quad t=1, 2,...,T$

(*) Questo lavoro è stato scritto durante un periodo di studio presso la London School of Economics con la supervisione del Prof. Steve Nickell che l'autore desidera ringraziare per i preziosi suggerimenti forniti. Un ringraziamento anche alla Dott.ssa Bianca de Stavola, al Dott. Giuseppe Marotta ed al Dott. Ignazio Visco per aver letto e commentato una precedente versione dello scritto.

dove l'indice it è relativo all'osservazione di una variabile per un individuo i al tempo t, y_{it} è la variabile dipendente, X_{it} è un vettore di variabili esplicative, π_{it} è un insieme di parametri (che in generale possono differire nel tempo e per individuo) e E_{it} è un insieme di disturbi casuali. Regressioni basate su serie storiche di cross section, rispetto a regressioni su dati aggregati, presentano svariati problemi (relativi soprattutto al modo migliore di modellare differenze individuali e temporali) ed il vantaggio di far pieno uso di tutta l'informazione disponibile, in qualche modo persa nel processo di aggregazione. Stime su dati individuali consentono, ad esempio, di controllare le possibili distorsioni che, su dati aggregati, possono insorgere come conseguenza dell'arbitraria assunzione di omogeneità nel comportamento delle varie unità decisionali.

L'uso di dati in panel può diventare una necessità se solo poche osservazioni temporali sono disponibili, sicché l'uso di una serie storica aggregata produrrebbe troppo pochi gradi di libertà. D'altro canto, ogni qual volta il ricercatore è interessato a modellare differenze tra individui, l'econometria dei panel data offre anche un'alternativa superiore alla stima con una singola cross section, anche perchè consente di tener conto di possibili effetti non osservati, costanti nel tempo, ma diversi tra individui, che produrrebero stime cross section distorte.

Nonostante la disponibilità di serie storiche di cross section sia diventata sempre più comune, è difficile trovare trattamenti unificati su questo argomento nei libri di testo di econometria ed il ricercatore applicato deve basarsi sulla complessa letteratura sul tema sviluppata a partire dagli anni '50[1]. Questo lavoro costituisce una guida a tale letteratura.

L'ampiezza del soggetto trattato ha tuttavia imposto dei vincoli. Innanzitutto, solo la stima di singole equazioni verrà considerata[2]. In secondo luogo, si esamineranno solo modelli lineari [3] del tipo:

$$(2)\quad y_{it}=\Sigma_k \beta^k_i \; x^k_{it} + \alpha_i + \lambda_t + v_{it} \qquad i=1,\dots,N \quad t=1,\dots,T$$

Nei paragrafi 2-7 si considererà il caso più comune di utilizzo dei dati in <u>panel</u>, e cioè i modelli con coefficienti angolari uguali tra individui e nel tempo e di intercette variabili. La maggior parte dei problemi che possono sorgere in questo contesto sarà tuttavia discussa sotto l'assunzione di assenza di effetti temporali ($\lambda_t=0$), anche se si forniranno indicazioni per il caso più complesso. Negli ultimi due paragrafi, si esaminerà invece il caso di coefficienti angolari diversi tra individui, mentre l'ipotesi di coefficienti variabili tra individui e nel tempo non viene considerata dato il trascurabile impatto di tali modelli sulla pratica econometrica.

In entrambe le parti si esamineranno sia l'ipotesi di natura non stocastica che quella di natura stocastica dei coefficienti, dato che nessuna delle due ipotesi ha finora presentato una chiara superiorità. Infine, dato lo scopo di questo lavoro, si seguirà una esposizione cronologica, anche se si verificheranno eccezioni per i casi in cui è sembrato conveniente raggruppare parte dei lavori esaminati in base all'argomento, piuttosto che alla data di pubblicazione.

2. Analisi della covarianza e problema del pooling.

2.1 Analisi della covarianza. L'uso di dati in panel per l'analisi econometrica ha origine negli anni '50 con l'applicazione a problemi economici dell'analisi della covarianza, la tecnica statistica sviluppata per primo da R.A. Fischer (1918) é applicata inizialmente alla ricerca in campo biologico e agronomico.

Scheffé (1959, pp. 4-5) descrive questa tecnica nel modo seguente: data la relazione:

$$y_i = \Sigma_k \beta^k x_i^k + \varepsilon_i \qquad i=1,\dots,N \tag{3}$$

dove l'indice i si riferisce a osservazioni relative all'individuo i, la seguente distinzione può essere tracciata:

a) in analisi della varianza gli x sono interi, tipicamente variabili binarie 0-1; ad esempio x^k può assumere valore 1 per indicare la presenza di una determinata caratteristica (sesso, razza, ecc.) e valore 0 per indicarne l'assenza;

b) in analisi con regressioni gli x sono variabili continue;

c) in analisi della covarianza alcuni x sono variabili binarie, altre sono variabili continue.

Inoltre se il ricercatore dispone di più di una osservazione per ogni individuo e se si ritiene che ogni individuo possieda proprie caratteristiche, diverse da individuo a individuo, è possibile aggiungere un indice t ad ogni osservazione e introdurre sia un effetto individuale (α_i) costante nel tempo ma diverso per individuo, sia un effetto temporale (λ_t) che influisce su tutti gli individui in misura uguale al tempo t scrivendo il modello come:

(4) $y_{it} = \Sigma_k \beta^k x^k_{it} + \alpha_i + \lambda_t + \varepsilon_{it} \qquad i=1,\dots,N \quad t=1,\dots,T$

Occorre sottolineare che, mentre nelle usuali analisi di regressione il ricercatore è principalmente interessato nella stima dei coefficienti della (4), in analisi della varianza e della covarianza l'interesse è volto a verificare l'ipotesi che i coefficienti, in particolare gli effetti α_i e λ_t (nella letteratura sull'argomento chiamati abitualmente layouts) siano uguali per gruppi diversi di osservazioni (individui, anni o, in generale, "celle"). Per questo scopo, l'analisi della covarianza scompone la varianza delle variabili in parti attribuibili a diverse origini e forma rapporti che sotto l'ipotesi di eguaglianza dei coefficienti seguono una distribuzione F, semprechè i residui della (4) siano normalmente e indipendentemente distribuiti. I modelli usati in analisi della covarianza (e la forma dei test proposti) risultano diversi a seconda della natura stocastica degli effetti, che possono essere considerati fissi o casuali, cioè provenienti da una distribuzione con media abitualmente assunta uguale a zero ed una varianza sconosciuta. Fino alla seconda metà degli anni sessanta, tuttavia, i modelli con effetti casuali (random effects models) non vennero utilizzati in applicazioni economiche su dati in panel. E' perciò conveniente considerare prima di tutto i modelli a effetti fissi.

2.2 Rilievo dell'analisi della covarianza per le regressioni con dati in panel - L'importanza dell'analisi della covarianza, e in generale, di metodi statistici che consentano di sottoporre a verifica l'ipotesi di omogeneità nel comportamento degli individui è essenziale. Da questo punto di vista la disponibilità di informazioni disaggregate consente innanzitutto di accertare in che misura stime basate su dati aggregati (che comunque possono risultare più convenienti per il ricercatore se non altro per i minori problemi di calcolo che esse generano) possano risultare distorte da bias di aggregazione. E' noto infatti che se il comportamento individuale è descritto, per es., dalla relazione

(5) $$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it} \qquad i=1,\dots,N \quad t=1,\dots,T$$

stime su dati aggregati del tipo $Y_t = \alpha + \beta X_t + \varepsilon_t$ con ($Y_t = \sum_i Y_{it}$) produrranno in generale stime distorte a causa del vincolo di uguaglianza dei parametri imposto erroneamente (cfr. Theil (1954)) [4].

In secondo luogo, un test sulla eguaglianza dei coefficienti individuali è una condizione per la scelta del metodo di stima più adeguato quando si decida di operare con dati individuali. Infatti, se l'ipotesi di omogeneità risulta confermata, una regressione con OLS su dati disaggregati (con NT informazioni per ogni variabile), o persino una semplice regressione su una cross section, produrrà stime non distorte. Al contrario, il venir meno della condizione di omogeneità, anche soltanto della costante individuale, produrrà in generale stime distorte, come evidenziato dalla fig. 1, in cui il coefficiente β stimato su micro-osservazioni degli individui A e B eccede il coefficiente angolare corretto.

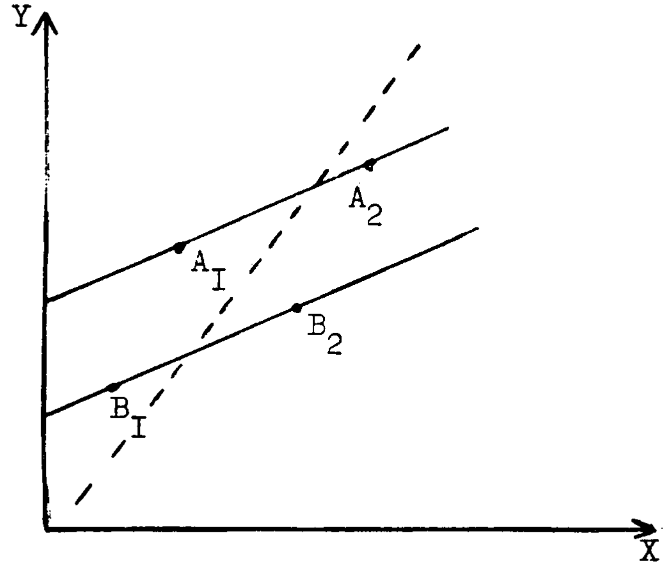
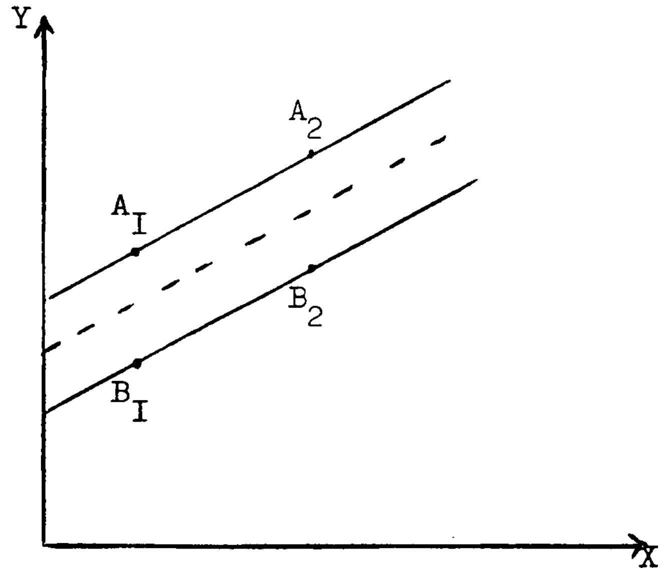


Figura 1

Figura 2



Si noti che è questo un tipico caso di distorsione da variabili omesse (gli effetti individuali). Si può dimostrare che per il caso di regressione semplice illustrato nella fig. 1 la distorsione risulterebbe pari a:

$$E(\hat{\beta}) - \beta = \frac{\Sigma_i \Sigma_t \tilde{x}_{it} (\alpha_i - \bar{\alpha})}{\Sigma_i \Sigma_t x_{it}^2} = \frac{\Sigma_i (\alpha_i - \bar{\alpha}) \Sigma_t \tilde{x}_{it}}{\Sigma_i \Sigma_t x_{it}^2} \qquad (6)$$

dove $\tilde{x}_{it} = x_{it} - \bar{x}$ e $\bar{x} = \frac{1}{NT} \Sigma_i \Sigma_t x_{it}$ $\bar{\alpha} = \frac{1}{N} \Sigma_i \alpha_i$

Il numeratore della (6), diviso per T, è anche il numeratore di una regressione con minimi quadrati degli α_i sulla media degli x presa rispetto al tempo, cioè:

$$\alpha_i = \delta + \gamma x_{i.} + v_i \qquad \text{dove} \quad x_{i.} = \frac{1}{T} \Sigma_t x_{it}$$

La distorsione è perciò nulla in assenza di correlazione tra gli $x_{i.}$ e gli effetti α_i, come evidenziato dalla fig. 2. E' questo un problema ben noto quando si usano dati longitudinali (vedi, per es.,Mundlak (1961) e Kuh (1963)) ed esiste indipendentemente dalle assunzioni sulla natura stocastica degli effetti. Nonostante ciò, come si vedrà in seguito, per parecchi anni praticamente tutta la letteratura econometrica sull'uso di dati in panel basata sulla assunzione di effetti casuali ha trascurato completamente la possibilità di correlazione tra gli effetti non osservati ed i regressori osservati[(5)].

2.3 Tests sulla omogeneità dei coefficienti - I test di analisi della covarianza descritti precedentemente trovano un esatto equivalente nei più usuali test F sui vincoli imposti sui coefficienti di una regressione ed è in questa forma che verranno qui considerati (cfr. Maddala (1977 pp.320-326) e Harvey (1981 pp. 63-67)).

Si consideri il caso di regressione semplice:

$$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it} \qquad i=1,\dots,N \qquad t=1,\dots,T$$

ovvero:

$$(7) \qquad \underline{y}_i = \underline{\alpha}_i + \underline{x}_i \beta_i + \underline{\varepsilon}_i \qquad i=1,\dots,N$$

dove $\underline{y}_i' = \left[y_{i1} \cdots\cdots y_{iT}\right]$ $\quad \underline{\alpha}_i' = \left[\alpha_i \cdots\cdots \alpha_i\right]$

$\underline{x}_i' = \left[x_{i1} \cdots\cdots x_{iT}\right]$ $\quad \underline{\varepsilon}_i' = \left[\varepsilon_{i1} \cdots\cdots \varepsilon_{iT}\right]$

Il modello non vincolato, quello cioè in cui tutti i coefficienti sono diversi, assume la forma seguente:

$$(8) \qquad \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 1 & \cdots & 0 \\ 0 & \cdots & \vdots \\ \vdots & & 0 \\ \vdots & & 1 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} x_{11} & \cdots & 0 \\ \vdots & & \vdots \\ x_{1T} & \cdots & \vdots \\ \vdots & & 0 \\ \vdots & & x_{N1} \\ \vdots & & \vdots \\ 0 & \cdots & x_{NT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1T} \\ \vdots \\ \varepsilon_{N1} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix}$$

o: $\underline{Y} = (I_N \otimes e_T)\underline{\alpha} + \tilde{X}\underline{\beta} + \underline{\varepsilon}$

dove $\underline{Y}' = (\underline{y}_1' \cdots\cdots \underline{y}_N')$ con $\underline{y}_i'$ definito come nella (7)

$\underline{\alpha}' = (\alpha_1 \cdots\cdots \alpha_N)$

$\underline{\beta}' = (\beta_1 \cdots\cdots \beta_N)$

$\underline{\varepsilon}' = (\underline{\varepsilon}_1' \cdots\cdots \underline{\varepsilon}_N')$ con ε_i' definito come nella (7)

e_T è un vettore unitario di T elementi, I_N è una matrice unitaria

NxN, mentre il simbolo $\otimes$ indica il prodotto di Kronecker e $\overset{*}{X}$ è la matrice NTxN delle osservazioni sulla variabile esogena. Si noti che nel modello (8) le osservazioni sono solo apparentemente combinate: le stime dei coefficienti sono le stesse che si otterrebbero stimando N regressioni su dati individuali. Per lo stesso motivo, la somma dei quadrati dei residui non vincolati (<u>unrestricted residual sum of squares</u> = URSS) è la somma delle N somme dei quadrati dei residui delle regressioni individuali.

Modelli ristretti sono ottenuti imponendo l'uguaglianza tra individui di alcuni coefficienti. Diversi casi possono essere d'interesse per il ricercatore. Una ipotesi sottoposta spesso a verifica è che tutti i coefficienti angolari siano uguali, mentre le intercette, riflettendo caratteristiche individuali costanti nel tempo, siano diverse tra individui. Sotto questa ipotesi il modello ristretto può scriversi:

$$(9) \qquad \underline{Y} = (I_N \otimes e_T)\underline{\alpha} + \beta \underline{X} + \underline{\varepsilon}$$

dove $\underline{X}' = (\underline{x}'_1 \ldots\ldots x'_N)$, un vettore 1xNT.

Il modello (9) può essere stimato con minimi quadrati sotto le usuali assunzioni sull'errore ε_{it}, più quella di normalità, necessaria perchè il test sotto riportato sia valido anche per campioni piccoli. La somma dei quadrati dei residui della (9) (<u>restricted residual sum of squares</u> = RRSS), avendo imposto (K'-1)(N-1)=1(N-1) restrizioni (K'=2 nell'esempio [6]) può essere utilizzata per costruirne la statistica F(N-1, N(T-2)):

$$F = \frac{(RRSS - URSS)/(N-1)}{URSS/N(T-2)}$$

dove $N(T-K') = N(T-2)$ sono i gradi di libertà del modello non ristretto.

Varie ipotesi potrebbero essere sottoposte a verifica, per esempio che tutti i coefficienti inclusa l'intercetta sono uguali contro una ipotesi nulla che l'intercetta è diversa. Quando l'ipotesi di uguaglianza di tutti i coefficienti non può essere rigettata si possono applicare i minimi quadrati su osservazioni combinate. In questo caso il modello si riduce a:

$\underline{Y} = \begin{bmatrix} e_{NT} & \vdots & \underline{X} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, dove e_{NT} è un vettore unitario di NT elementi. Alla stima con OLS di questo modello si da il nome di minimi quadrati su osservazioni combinate (<u>pooled</u> OLS o POLS). Ipotesi addizionali sulla correlazione degli errori tra diversi individui (simili a quelle del modello SURE considerato al par. 9.1) potrebbero però essere considerate, come descritto da Kmenta (1971 pp. 508-514) [7].

Dato che il modello (9) utilizza variabili <u>dummy</u> additive per modellare differenze tra individui, lo si denomina spesso <u>dummy variable model</u>; viene anche denominato modello a effetti fissi data l'assunzione sulla natura (fissa) degli effetti ed il corrispondente stimatore, a causa delle sue origini, viene chiamato stimatore di covarianza [8]. L'uso di variabili <u>dummy</u> per modellare differenze tra gruppi di dati, raccomandato in questo contesto per la prima volta da Hildreth (1950), fu praticamente l'unico metodo per trattare dati in <u>panel</u> fino alla metà degli anni '60. Il modello (9) può essere facilmente esteso per tener conto di effetti temporali (uguali tra individui ma diversi nei diversi periodi; per una rassegna si veda in proposito Judge et al. (1980, pp. 339-341)) e per altri effetti che interessano solo un sottoinsieme di dati (come per esempio i "village effects" di Mundlak (1961)). Le caratteristiche di questo approccio possono tuttavia essere discusse con riferimento al più semplice modello (9).

3. Il modello a effetti fissi

3.1 Stime con variabili dummy. Il modello (9), nel caso di K regressori, può essere scritto come:

$$(10) \qquad \underline{Y} = Z\,\underline{\alpha} + X\,\underline{\beta} + \underline{\varepsilon} = H\gamma + \underline{\varepsilon}$$

dove Z è la matrice $I_N \otimes e_T$ di dimensione NTxN, X è la matrice NTxK delle variabili esogene, $\underline{\beta}$ è un vettore Kx1 di coefficienti angolari, $H = [Z \,\vdots\, X]$, $\gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ e si assume:

$$E(\varepsilon_{it}) = 0 \;,\; E(\varepsilon_{it}^2) = \sigma_\varepsilon^2 \quad e \quad E(\varepsilon_i\, \varepsilon_j) = 0 \;(\text{per } j \neq i)$$

In molti casi di dati in panel, il numero degli individui N è elevato sicchè l'ordine di H'H è piuttosto alto; di conseguenza, invece di applicare i minimi quadrati direttamente dalla (10) risulta più conveniente adottare la seguente trasformazione che, fra l'altro, rende manifesta la forma dello stimatore con variabili dummy (dummy variabile estimator = DVE). Usando la formula della regressione a blocchi, lo stimatore OLS dei coefficienti angolari della (10) può infatti scriversi:

(11) $$\hat{\beta}_W = (X'QX)^{-1}X'Q\underline{Y} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$$

dove $Q = I_{NT} - Z(Z'Z)^{-1}Z'$ è idempotente e $\tilde{X} = QX$, $\tilde{Y} = Q\underline{Y}$.
Dato che $Z = I_N \otimes e_T$ si ha che $Q = I_{NT} - I_N \otimes \frac{1}{T} e_T e'_T$
e perciò la (11) può essere scritta come:

(12) $$\hat{\beta}_W = (\Sigma_i X_i' D_T X_i)^{-1} \Sigma_i X_i' D_T Y_i' \quad \text{dove } D_T = I_T - \frac{e_T e_T'}{T} .$$

D_T è una matrice che trasforma ogni osservazione in deviazione dalla media individuale presa nel tempo (per es. y_{it} è sostituito da $y_{it} - \frac{1}{T} \Sigma_t y_{it}$). Dalla (11) e dalla (12) si desume che la procedura con variabili dummy equivale a premoltiplicare la (10) per una matrice che trasforma ogni osservazione in deviazione dalla media (in tal modo eliminando tutti i regressori invarianti rispetto al tempo) e a stimare con OLS l'equazione trasformata:

(13) $$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

E' questa una conseguenza di imporre diverse intercette; il DVE minimizza congiuntamente la somma dei quadrati delle deviazioni dalle rette di regressione individuali,il che è più efficiente di minimizzare distintamente la deviazione di N rette di regressione (come sarebbe fatto senza combinare i dati) , anche se, si sottolinea, esattamente lo stesso tipo di informazione (deviazione dalle medie individuali) verrebbe utilizzato. Dato che il DVE utilizza la variazione di ogni variabile all'interno di ogni serie individuale lo stimatore è spesso chiamato "within estimator".

Sotto le usuali ipotesi sul comportamento stocastico dell'errore, sull'esogenità dei regressori e sulla natura non stocastica dei coefficienti, il DVE è BLUE (best linear unbiased estimator). Una stima delle N intercette è ricavata da:

(14) $\alpha_i = y_{i.} - \Sigma_k \hat{\beta}^k x^k_{i.}$

dove (qui e in seguito) si definisce $y_{i.} = \frac{1}{T} \Sigma_t y_{it}$ (e similmente per $x_{i.}$). Infine la matrice varianze-covarianze dei coefficienti è, come usuale:

$$\mathrm{var}(\hat{\gamma}) = \sigma_\varepsilon^2 (H'H)^{-1} \text{ o, per } \hat{\beta}, \ \mathrm{var}(\hat{\beta}) = \sigma_\varepsilon^2 (X'X)^{-1}$$ (9).

Anche nel caso in cui il DVE è BLUE, il metodo ha due svantaggi. Innanzitutto, dato che l'introduzione di variabili dummy comporta la descritta trasformazione, se sono presenti regressori fissi nel tempo la stima dei coefficienti su tali regressori non risulta possibile dato che essi non appaiono nella (13)(10). In secondo luogo, la trasformazione aggrava gli effetti di possibili errori nella misurazione delle variabili (Griliches (1979 pp. 42-43)).(11)

3.2 Regressioni dinamiche. Qualche problema sorge quando uno dei regressori è una variabile dipendente ritardata. Il DVE dei coefficienti angolari può infatti scriversi come:

$$\hat{\beta}_W = \beta + (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{\varepsilon}$$

Il vettore $\tilde{X}'\tilde{\varepsilon}$ comporta termini quali:

$$\Sigma_i \Sigma_t (x^k_{it} - x^K_{i.}) (\varepsilon_{it} - \varepsilon_{i.})$$

o, in regressioni dinamiche, $\Sigma_i \Sigma_t (y_{i,t-1} - y_{i.})(\varepsilon_{it} - \varepsilon_{i.})$

Dato che $y_{i,t-1}$ dipende da $\varepsilon_{i,t-1}$ e $\varepsilon_{i.} = \frac{1}{T}(\varepsilon_{i1} + \ldots + \varepsilon_{iT})$ si ha che:

$$E\, \Sigma_i \Sigma_t (y_{i,t-1} - y_{i.})(\varepsilon_{it} - \varepsilon_{i.}) \neq 0$$

sicchè il DVE non è corretto (cfr. Nickell (1981)). E' questo un tipico caso di distorsione di Hurwicz (1950) in modelli dinamici, ben documentato in ogni libro di testo di econometria; in tal caso il ricercatore si affida a risultati asintotici per affermare la consistenza dei minimi quadrati. Con dati in panel, tuttavia, il comportamento asintotico degli stimatori dipende dal modo in cui il numero di osservazioni NT tende ad infinito (cioè se è N o T a crescere). Nel caso precedente, all'aumentare di T l'effetto di ε_{it} su $\varepsilon_{i.}$ diventa trascurabile e la distorsione tende a zero. Ma ciò non accade quando il numero di individui N tende a infinito. Si ha così che: $\operatorname{plim}_{T\to\infty} (\hat{\beta}_W - \beta) = 0$, ma $\operatorname{plim}_{N\to\infty} (\hat{\beta}_W - \beta) \neq 0$(12) e nel secondo caso il DVE non è consistente.(13)

Questo problema è piuttosto serio. Nella maggior parte dei casi, la dimensione temporale del panel è abbastanza corta, mentre tipicamente è N ad essere grande (cfr.tav. 1). La direzione e dimensione della distorsione del DVE sono stati studiati con esperimenti Montecarlo da Nerlove (1967, 1971), ma il solo risultato analitico è fornito da Nickell (1981); per il caso di autocorrelazione del primo ordine:

$$y_{it} = \alpha_i + \vartheta y_{i,t-1} + \underline{x}_{it} \underline{\beta} + \varepsilon_{it}$$

Tavola 1 Lavori empirici con panel data

Autore	**Natura degli eff.**	**Metodo di stima**	**N**	**T**	**Tipo di equazione**	**Oggetto**
Johnson(1960)	FE	DV	31	8	statica	funzione di produzione
Mundlak(1961)	RE	DV	66	5	statica	funzione di produzione
Hock(1962)	FE	DV	63	6	statica	funzione di produzione
Schipper(1964)	FE	DV	30	3	statica	domanda di consumo
Johnson(1964)	FE	DV	34	4	statica	funzione di costo
Balestra-Nerlove(66)	RE	BN	36	6	dinamica	domanda di gas
Lee(1966)	FE	DV	2605	4	statica	domanda di att. finanz.
Palda-Blair(1970)	FE	DV	24	5	statica	domanda di consumo
Houthakker et al.(1974)	RE	GLS	49	40	dinamica	domanda di gas
Halvorsen(1975)	//	POLS	48	9	statica	domanda di elettricità
Hester-Pierce(1975)	FE	DV	184	10	dinamica	domanda di att. finanz.
Hester(1978)	//	POLS	320	287	dinamica	domanda di att. finanz.
Lillard-Willis(1978)	RE	GLS	1144	7	statica	rend. del capitale umano
Lillard-Weiss(1979)	RE	GLS	4330	7	statica	rend. del capitale umano
Keifer(1979)	FE	DV	608	6	statica	rend. del capitale umano
Taylor(1980)	RE	DV/BN	10	20	statica	domanda di investimento
Hausman-Taylor(1981)	RE	IGLS	750	2	statica	rend. del capitale umano
Marotta(1982)	FE	DV	17	14	dinamica	domanda di investimento
Griffith-Anderson(82)	RE	GLS	38	10	statica	funzione di produzione
Bhargava et al.(1982)	FE	DV	962	10	statica	rend. del capitale umano
Bhargava-Sargan(1983)	RE	ML	962	9	dinamica	rend. del capitale umano
Marcus(1983)	FE/RE	IDV/IGLS	44	13	dinamica	domanda di capitale banc.
Mairesse(1984)	FE	*	307 145 422	10	statica	domanda di investimento

FE=effetti fissi RE=effetti stocastici
DV=OLS con variabili dummy BN=Stimatore Balestra-Nerlove
GLS=Minimi quadrati generalizzati diversi da BN
POLS=OLS su osservazioni combinate (nessun effetto individuale o temporale)
ML=Stimatori di massima verosimiglianza
IGLS=GLS con variabili strumentali
IDV=DV con variabili strumentali
*=OLS su variazioni annuali e multiannuali

(dove $\underline{x}_{it}$ è un vettore di regressori e $\underline{\beta}$ contiene i relativi coefficienti), Nickell trova una espressione esatta per la distorsione asintotica ($N \rightarrow \infty$):

$$(15) \qquad \underset{N \to \infty}{\text{plim}} (\hat{\vartheta} - \vartheta) = \left\{ \underset{N \to \infty}{\text{plim}} \frac{1}{NT} \tilde{Y}'_{-1} M \tilde{Y}_{-1} \right\} \underset{N \to \infty}{\text{plim}} \frac{1}{NT} \tilde{Y}'_{-1} \tilde{\varepsilon}$$

dove $M = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$, X è la matrice delle variabili esogene, Y_{-1} il vettore della dipendente sfasata, ε quello dell'errore e ~ indica che le variabili sono state sostituite da deviazioni dalla media individuale presa rispetto al tempo. Come illustrato da Nickell, la (15) comporta che quando $\vartheta > 0$ la distorsione su $\hat{\vartheta}$ è negativa (il segno è altrimenti incerto); il <u>bias</u> su $\hat{\beta}$ è invece pari a

$$\underset{N \to \infty}{\text{plim}} (\hat{\beta} - \underline{\beta}) = -\underset{N \to \infty}{\text{plim}} \left[(\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{Y}_{-1} \right] \underset{N \to \infty}{\text{plim}} (\hat{\vartheta} - \vartheta)$$

il cui segno dipende dalla relazione tra $y_{i,t-1}$ e la variabile esogena in questione:

"If the exogenus variable is positevely related (in the regression sense) to $y_{i,t-1}$, then.....its coefficient will be upward biased and viceversa"(p. 1424).

Simili conclusioni, anche se non precise, erano state raggiunte da Maddala (1971 - a, pagg. 348-349), anche per campioni piccoli.

Come illustrato nel par. 4.2, la potenziale inconsistenza del DVE, la cui soluzione è rinviata al par. 7.2, stimolò la ricerca di Nerlove (1965) e Balestra e Nerlove (1966) per un nuovo stimatore.

4. <u>Il modello a effetti casuali</u>

4.1 <u>Il between estimator</u> - Prima di introdurre lo stimatore di

Balestra e Nerlove è utile presentare il così detto "between estimator". E' perciò conveniente scrivere la (10) senza dummies come:

(16) $\underline{Y} = \underline{\alpha}° + X\underline{\beta} + \underline{\varepsilon}$

(dove $\underline{\alpha}° = Z\underline{\alpha}$ è un vettore NT x 1 in cui i primi T elementi sono α_1, i secondi T α_2 e così via). Moltiplicando la (16) per una matrice P $= I_{NT} - Q = I_N \otimes \frac{1}{T} e_T e'_T$ che trasforma ogni osservazione nella corrispondente media individuale rispetto al tempo si ottiene:

(17) $P\underline{Y} = \alpha° + PX\underline{\beta} + P\underline{\varepsilon}$ dato che $P\alpha° = \alpha°$

Si noti che, nonostante $P\underline{Y}$ (come pure le colonne di PX ed il vettore $P\underline{\varepsilon}$) contengono NT elementi solo N elementi diversi sono presenti e perciò tutto il contenuto informativo della (17) può essere riassunto in una regressione cross section su dati medi:

(18) $\bar{Y} = \alpha + \bar{X}\beta + \bar{\varepsilon}$

dove $\bar{Y}$, $\bar{X}$ e $\bar{\varepsilon}$ sono vettori N x 1 di medie individuali. Gli elementi di α, come evidente dalla (18), non possono in questo caso essere stimati e devono perciò essere considerati come parte dell'errore della regressione. La stima con OLS della (18) (o della (17)[14]) produce il "between estimator":

(19) $\hat{\beta}_b = (X'PX)^{-1} X'PY$

La matrice dei momenti X'PX di questo stimatore, se una colonna di 1 è inclusa tra i regressori, implicherà termini del tipo $\Sigma_i (x^j_{i.} - x.^j)(x_{i.}^k - x.^k)$ dove x^j e x^k sono due regressori e $x.^k = (1/NT) \Sigma_i \Sigma_t x^k_{it}$, cioè somme di prodotti di deviazioni delle medie individuali dalla media generale. Dato che la (19) utilizza la variazione della media delle variabili tra diversi individui, lo stimatore in questione è detto "between estimator". Come il "within estimator", il "between estimator" è corretto se i regressori sono esogeni ed è distorto in presenza di una endogena ritardata. Tuttavia, in questo caso la distorsione si manifesta non solo quando T è piccolo, dato che essa è anche dovuta all'omissione degli effetti di α_i che sono certamente correlati con il regressore $y_{i,t-1}$ (Balestra e Nerlove (1966, p. 603)). Tuttavia Maddala (1971 - a, pp. 348-349) osservò che la direzione della distorsione tende ad essere opposta a quella del "within estimator" e che perciò: "... we should be able to do better by taking a linear combination of these two regressors" il che è appunto una caratteristica dello stimatore proposto da Balestra e Nerlove [15].

4.2 Il Balestra-Nerlove Estimator (BNE) - Nonostante fosse stato sviluppato nell'ambito dell'analisi della covarianza molto prima degli anni '60 e fosse già stato applicato a problemi econometrici da Kuh (1959)[16], il modello a effetti casuali (random effects model o modello RE) divenne di uso comune per gli studi econometrici solo dopo un celebre articolo di Balestra e Nerlove (1966). Ideato per affrontare la stima di un modello dinamico su cui il DVE sembrava dare risultati implausibili, il nuovo approccio era basato sull'osservazione che, in molti casi, è più appropriato considerare gli effetti individuali o temporali come variabili casuali non osservabili e quindi inclùderli nell'errore della regressione. Escludendo gli effetti dalla parte sistematica della regressione, come già osservato, si può incorrere in un errore di specificazione e in una distorsione, a causa della possibile correlazione tra effetti omessi e regressori inclusi. Tuttavia, praticamente tutta la letteratura econometrica nel decennio seguente l'articolo di Balestra e Nerlove trascurò completamente il problema e fu perciò implicitamente, e talvolta esplicitamente (cfr. Maddala (1971 - a, p. 357), basata sull'assunzione di indipendenza tra effetti e regressori. Nell'interesse dell'argomento, questa assunzione è adottata nei par. 4 e 5.

Il modello descritto può scriversi:

$$(20) \qquad y_{it} = \underline{x}_{it}\beta + v_{it} \qquad i=1,\dots,N \quad t=1,\dots,T$$

$$v_{it} = \alpha_i + \lambda_t + \varepsilon_{it} \qquad \text{e} \quad E(\alpha_i) = 0 \quad E(\alpha_i^2) = \sigma_\alpha^2$$

$$E(\lambda_t) = 0 \quad E(\lambda_t^2) = \sigma_\lambda^2$$

dove la matrice delle variabili predeterminate può includere un vettore unitario per inserire una intercetta comune. Dato il modello (20) ed essendo interessati alla stima di β, il problema è quello di regredire un'equazione con una matrice varianze-covarianze dei residui non scalare, la cui forma specifica dipenderà dalle assunzioni sulla struttura delle correlazioni tra α_i, λ_t e ε_{it}.

Seguendo Balestra e Nerlove si assume che $E(\alpha_i \alpha_j)=0$, che non ci sia alcun effetto temporale e che ε_{it} sia "white noise".
Ne segue che per gli N individui il modello può scriversi come:

(21) $Y = X\beta + V$ [17]

dove $E(V) = 0$ e $E(VV') = \Omega = \sigma^2 \begin{bmatrix} A & & 0 \\ & A & \\ & \ddots & \\ 0 & & A \end{bmatrix} = \sigma^2 \Omega^{\circ}$

e $A = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & & \vdots \\ \vdots & & \ddots & \\ \rho & & & 1 \end{bmatrix}$

è una matrice TxT con $\sigma^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ e $\rho = \sigma_\alpha^2 / \sigma$

Se la variabile dipendente sfasata non appare tra i regressori, la stima OLS produce risultati corretti, ma inefficienti, mentre i minimi quadrati generalizzati (GLS) sono efficienti. Lo stimatore GLS è dato in questo caso da:

(22) $$\hat{\beta}_{BN} = (X' \Omega^{\circ -1} X)^{-1} X' \Omega^{\circ -1} Y$$

la cui matrice varianze-covarianze è stimata da:

$$\mathrm{var}(\hat{\beta}_{BN}) = \frac{\hat{e}' \Omega^{\circ -1} \hat{e}}{NT-K} (X' \Omega^{\circ -1} X)^{-1}$$

Inoltre, se una stima consistente di ρ, il così detto coefficiente di correlazione tra classi, è disponibile, i GLS che utilizzino tale stima condividono asintoticamente le proprietà (efficienza e consistenza) dei GLS che utilizzino il vero ρ.

4.3 Decomposizione del BNE - Alcuni importanti proprietà del BNE vennero messe in luce da Maddala (1971 - a). Le conclusioni di Maddala possono derivarsi espandendo la (22). Dato che $\Omega°$ è una matrice diagonale a blocchi $\Omega°^{-1}$ ha forma:

$$\Omega°^{-1} = \begin{bmatrix} A^{-1} & & \\ & \ddots & \\ & & A^{-1} \end{bmatrix}$$

Perciò, dalla scomposizione $X' = \left[X'_1 \vdots X'_2 \ldots\ldots \vdots X'_N\right]$ la (22) viene scritta nel modo seguente:

(23) $\hat{\beta}_{BN} = (\Sigma_i X'_i A^{-1} X_i)^{-1} \Sigma_i X'_i A^{-1} y_i$

Dato che $A = (1-\varrho)I_T + \varrho e_T e'_T$ la sua inversa è $A^{-1} = \lambda_1 e_T e'_T + \lambda_2 I_T$ (18)

dove: $\lambda_1 = \dfrac{-\varrho}{(1-\varrho)(1-\varrho+T\varrho)}$ e $\lambda_2 = \dfrac{1}{1-\varrho}$

Aggiungendo e sottraendo $\lambda_2 \dfrac{ee'}{T}$, A^{-1} può scriversi come:

(24) $A^{-1} = (T\lambda_1+\lambda_2)\dfrac{ee'}{T} + \lambda_2(I-\dfrac{ee'}{T})$

Sostituendo la (24) nella (23):

$$\hat{\beta}_{BN} = \left\{\Sigma_i X'_i \left[(T\lambda_1+\lambda_2)\frac{ee'}{T} + \lambda_2(I-\frac{ee'}{T})\right] X_i\right\}^{-1} \Sigma_i X'_i \left[(T\lambda_1+\lambda_2)\frac{ee'}{T} + \lambda_2(I-\frac{ee'}{T})\right] y_i$$

$$= \left[\Sigma_i X'_i (T\lambda_1+\lambda_2)\frac{ee'}{T} X_i + \Sigma_i X'_i \lambda_2(I-\frac{ee'}{T})X_i\right]^{-1} \left[\Sigma_i X'_i (T\lambda_1+\lambda_2)\frac{ee'}{T} Y_i + \Sigma_i X'_i \lambda_2(I-\frac{ee'}{T})y_i\right]$$

e notando che $\Sigma_i X'_i \dfrac{ee'}{T} X_i = X'PX$ e che $\Sigma_i X'_i (I-\dfrac{ee'}{T}) X_i = X'QX$

si ottiene:

(25) $$\hat{\beta}_{BN} = \left[(T\lambda_1+\lambda_2)X'PX + \lambda_2 X'QX\right]^{-1} \left[(T\lambda_1+\lambda_2)X'PY + \lambda_2 X'QY\right]$$

$$= (\vartheta X'PX + X'QX)^{-1}(\vartheta X'PY + X'QY)$$

dove: $\vartheta = \frac{T\lambda_1 + \lambda_2}{\lambda_2} = \frac{1-\varrho}{1-\varrho+\varrho T} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$

Infine, facendo uso della (11) e della (19), il BNE può essere espresso come una media ponderata degli stimatori "within" e "between" :

(26) $$\hat{\beta}_{BN} = \left[X'PX + \vartheta X'QX\right]^{-1} (X'PX \; \hat{\beta}_b + \vartheta X'QX \; \hat{\beta}_w)$$ (19)

Tre osservazioni sono necessarie:

a) sotto le assunzioni del modello RE, in regressioni statiche, lo stimatore GLS descritto, per ϱ noto (o per ϱ stimato consistentemente e N o T tendenti a infinito), è più efficiente sia del within che del between estimator (cfr. Wallace e Hussain (1969) e Taylor (1980 pp. 210-211)) e raggiunge il limite inferiore di Cramer e Rao. Una ragione per questa maggiore efficienza risiede nell'uso di maggiore informazione sulla variazione delle variabili, dato che lo stimatore considera sia la variazione individuale (within) che quella tra individui (between) (Maddala (1971 - a, p. 344));

b) dalla (26) si vede che al crescere di T ϑ tende a zero e $\hat{\beta}_{BN}$ diventa simile al DVE, che è perciò asintoticamente $(T \rightarrow \infty)$ efficiente[20]; lo stesso accade quando ϱ tende a uno e quando la variazione within è alta relativamente a quella between, mentre quando ϱ tende a zero (e ϑ tende quindi a 1) $\hat{\beta}_{BN}$ tende ai POLS (vedi nota 15).

c) le proprietà asintotiche del BNE dipendono dal comportamento asintotico degli stimatori between e within. Se una dipendente sfasata è inclusa, entrambi questi stimatori sono inconsistenti per N tendente a infinito, con T fisso, sicchè BNE non è consistente in regressioni dinamiche a meno che T tenda a infinito. Tuttavia, dato che la direzione della distorsione degli stimatori within e between sotto certe condizioni (Maddala (1971 - a, pp. 348-349)) è di segno opposto, ne segue che il BNE può dare risultati migliori quando T è piccolo[21].

4.4 Problemi di calcolo - Come usuale, invece di applicare direttamente la (22) è più conveniente trovare una matrice H tale che $HH' = \Omega$ e moltiplicare i dati per H^{-1} per ottenere una equazione con una matrice varianze covarianze dei residui scalare. Fuller e Battesee (1973 pp. 627-628) suggeriscono $H^{-1} = I_{NT} - (1 - \vartheta^{1/2})P$ dove ϑ e P sono state definite in precedenza nei par. 4.1 e 4.3. Di conseguenza gli OLS possono applicarsi all'equazione trasformata:

$$Y - (1 - \vartheta^{\frac{1}{2}})y_{i.} = \{X - (1 - \vartheta^{\frac{1}{2}})x_{i.}\}\beta + u$$

dove $u = \vartheta a_i + \left[\varepsilon + (1 - \vartheta)\varepsilon_{.}\right]$ è l'errore con matrice varianze e covarianze scalare(22).

4.5 Sviluppi - L'approccio a effetti casuali, che promette una maggiore efficienza in virtù di una semplice assunzione di casualità, divenne nella pratica econometrica e soprattutto nell'analisi teorica di gran lunga il favorito, almeno fino alla seconda metà degli anni '70(23). Nella sezione seguente, alcuni degli sviluppi più interessanti saranno esaminati ancora con riferimento alle più semplici assunzioni di Balestra e Nerlove. Due principali linee di ricerca verranno considerate: la ricerca di stimatori consistenti dei componenti della varianza (e la valutazione della bontà relativa del modello a effetti casuali rispetto al modello a effetti fissi quando $\Omega°$ debba essere stimata); e la ricerca di uno stimatore di massima verosimiglianza (maximum likelihood estimator = MLE) del modello a effetti casuali (par. 5.2). Infine nel par. 5.3 verrà esaminato un test sull'eguaglianza, tra individui, dei coefficienti angolari, simile a quello del par. 2.3, ma valido nel caso di effetti casuali [24].

5. Sviluppi del modello a effetti casuali

5.1 Stime dei componenti della varianza e maggiore efficienza - Per avere una utilità pratica il BNE, come ogni stimatore GLS, richiede stime consistenti dei componenti della varianza σ_ε^2 e σ_α^2 (in particolare del loro rapporto ϱ). In questo contesto, tuttavia, è ancora importante specificare il significato del termine consistente. Infatti una stima di ϱ che risultasse consistente per T tendente a infinito risulterebbe di scarsa utilità, se non per tutti i modelli a effetti casuali, almeno per il BNE, dato che questo stimatore converge al DVE al crescere di T. Inoltre Amemiya (1967) ha dimostrato che se N non è grande è impossibile ottenere stime consistenti e distinte dei componenti della varianza (cfr. anche Anderson e Hsiao (1982)). Un secondo punto riguarda l'efficienza: come usuale, l'utilizzo di stime di componenti della varianza in luogo dei parametri effettivi non tocca le proprietà asintotiche dei GLS. Infatti, al tendere di N o T a infinito, $\sqrt{NT}\,(\beta_{GLS} - \beta)$ converge in distribuzione ad una distribuzione normale con media zero e matrice varianze covarianze minima (nel senso che raggunge il limite inferiore di Cramer e Rao).

Quel che più importa da un punto di vista pratico è che, come dimostrato da Taylor (1980), il GLS che utilizzi stime di ϱ è certamente più efficiente del DVE a meno che il campione sia molto piccolo sia nella direzione temporale che in quella individuale.

Molti stimatori dei componenti della varianza sono descritti nella letteratura sull'argomento. Qui si riportano solo i due comuni:

a) quando tutti i regressori sono esogeni tanto lo stimatore <u>within</u> che quello <u>between</u> sono non distorti e quindi i rispettivi residui possono usarsi per stimare ρ . Dai residui <u>within</u> l'espressione:

(27) $$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}_w' \hat{\varepsilon}_w}{N(T-1)-K}$$

produce una stima non distorta di σ^2; dai residui <u>between</u> l'espressione:

(28) $$\hat{\sigma}_1^2 = \frac{\hat{\varepsilon}_b' \hat{\varepsilon}_b}{N-K}$$

fornisce una stima non distorta di $\sigma_\varepsilon^2 + T\,\sigma_\alpha^2$ e dalla (27) e (28) σ_α^2 può essere facilmente calcolato. Questa procedura, introdotta in questo contesto da Maddala (1971 - a, pp. 347-348) (cfr. anche Maddala e Mount (1973), Judge et al. (1981 p. 335)e Hausman e Taylor (1981)), è una applicazione dell'analisi della covarianza ed è noto che può derivarne una stima negativa di σ_α^2. Maddala (1971 - a,) suggerisce in questo caso di prendere $\sigma_\alpha^2 = 0$ (il che implica l'uso di POLS) o di attribuire il problema a variabili omesse, in particolare a effetti temporali non inclusi.

b) Balestra e Nerlove (1966) e Nerlove (1971 - a) suggeriscono vari stimatori dei componenti della varianza. Tra questi, quello che ha fornito i migliori risultati negli esperimenti Montecarlo tentati da Nerlove (1967 e 1971 - a), anche in presenza di una endogena ritardata, è basato su una stima preliminare col DVE (Nerlove (1971 - a, pp. 373-374)). Come nella (27) i residui del <u>within</u> estimator sono usati per calcolare $\hat{\sigma}_\varepsilon^2$; una stima di $\hat{\sigma}_\alpha^2$ è quindi fornita dalla varianza campionaria delle stime delle N intercette, cioè:

(29) $$\hat{\sigma}_\alpha^2 = \frac{\Sigma_i \left(\hat{\alpha}_i - \Sigma_i \hat{\alpha}_i/N\right)^2}{N}$$

Come mostrato da Maddala (1971 - a, p. 349) lo stimatore (29) è leggermente distorto verso l'alto, ma ha il vantaggio di essere sempre positivo. Naturalmente se una dipendente sfasata è inclusa tra i regressori queste stime sono distorte e inconsistenti (se T è fisso). In questo caso una alternativa è fornita da una stima dei componenti della varianza ricavata da regressioni consistenti con variabili strumentali (Nerlove (1971 - a, p. 373)), come indicato nel par. 7.2.

Numerosi altri stimatori sono descritti da Henderson (1953), Amemiya (1971), Wallace e Hussain (1969), Fuller e Battesee (1973) e Rao (1970 e 1972); per un commento su questi metodi si veda Maddala e Mount (1973)[(25)]. Questa abbondanza di stimatori può stupire. Tuttavia, Swamy (1971) e Fuller e Battesee (1973) hanno mostrato che, sotto certe condizioni, tutti questi stimatori producono stime asintoticamente efficienti[(26)]. Inoltre Taylor (1980) ha provato analiticamente che stime più efficienti dei componenti della varianza non implicano di necessità stime più efficienti dei coefficienti della regressione.

L'equivalenza ai fini pratici dei vari GLS basati su diverse stime di Ω è stata confermata dagli esperimenti Montecarlo di Maddala e Mount (1973) che, per un modello statico con N = 25 e T = 10, indicarono una chiara superiorità di tutti i GLS rispetto agli OLS con intercetta comune e, più limitatamente, rispetto al DVE (cfr. anche Arora (1973) in cui uno stimatore alternativo è derivato sulle linee tracciate da Swamy (1971, ch. 3) e Swamy e Arora (1972)). Nerlove (1967 e 1971) ha confermato la superiorità dell'approccio a effetti casuali anche in regressioni dinamiche (ancora con N = 25 e T = 10). Tra i diversi stimatori utilizzati (GLS con il vero Ω, OLS con intercetta comune, DVE, MLE e due GLS con Ω stimato) i migliori risultati sono forniti dai GLS con Ω noto; tra gli altri, le distorsioni più basse sono ottenute dai GLS che utilizzano (27) e (29) per stimare i componenti della varianza.

5.2 Stime di massima verosimiglianza - Con l'assunzione addizionale di normalità dei residui, varie applicazioni del principio della massima verosimiglianza ai dati in *panel* sono rinvenibili nella letteratura[27]. Nella valutazione di queste applicazioni è necessario distinguere tra regressioni statiche e dinamiche. Per regressioni statiche il MLE e il GLS sono asintoticamente equivalenti, risultando efficienti e consistenti (per N o T tendenti a infinito)[28].

Per il caso di regressioni dinamiche Balestra e Nerlove (1966) derivarono il MLE di σ^2, ϱ e i coefficienti angolari, assumendo però che $y_{i,t}$ fosse non stocastico per $t \leq 0$ e per tutti gli i. Tuttavia la funzione di massima verosimiglianza non raggiunse nella applicazione empirica, un massimo per il possibile intervallo di ϱ ($0 \leq \varrho < 1$)[29]. Inoltre, quando Nerlove (1971 - a) impose la convergenza nell'intervallo 0-1, il massimo fu raggiunto, per un numero di casi sorprendentemente alto, al margine inferiore $\varrho = 0$. Maddala (1971-a, pp. 348-349) giustificò analiticamente questo risultato provando che il MLE che usa osservazioni iniziali fisse produce frequentemente soluzioni al margine in modelli dinamici a causa della quasi singolarità della matrice di covarianza tra individui. Tuttavia, se il modello è correttamente specificato la probabilità che ciò accada tende a zero al crescere di T o N (Anderson e Hsiao (1981, p. 600))[30].

A causa di questi problemi e delle difficoltà di calcolo connesse al metodo della massima verosimiglianza, questo approccio, dopo il tentativo di Balestra e Nerlove, non trovò molto successo nelle applicazioni empiriche. Inoltre l'assunzione di valori iniziali y_{i0} fissi deve essere sembrata troppo irrealistica. Questa assunzione, tuttavia, è estremamente rilevante, come provato di recente da Anderson e Hsiao (1981 e 1982).

Anderson e Hsiao hanno dimostrato che la forma e le proprietà asintotiche del MLE dipendono crucialmente dalle assunzioni sulla natura delle osservazioni iniziali. Più precisamente, dato il modello:

$$y_{it} = \beta y_{i,t-1} + \alpha_i + \varepsilon_{it} \quad (31)$$

quattro diverse assunzioni vengono considerate rispetto a y_{i0}:

a) i valori y_{i0} sono fissi e indipendenti dagli α_i (come in Balestra e Nerlove (1966));
b) gli y_{i0} sono distribuiti casualmente intorno a medie fisse;
c) y_{i0} dipende da α_i, secondo una relazione non stocastica;
d) y_{i0} dipende da α_i, secondo una relazione stocastica.

Sotto queste diverse assunzioni gli autori hanno derivato la forma e la distribuzione asintotica di MLE per T o N tendenti ad infinito ed hanno mostrato che, mentre il MLE è sempre consistente al crescere di T (con N crescente o fisso), lo stimatore può risultare inconsistente quando N tende a infinito con T fisso [32].

Quel che più importa:

"Different assumptions about the initial conditions call for different methods to obtain the MLE. Mistaking one case for the other will generally not lead to asymptotically equivalent formulas. Consequently, the misused estimator may be inconsistent. Unfortunately we usually have little information to rely on in making a correct choice of the initial conditions" (p. 505).

Queste conclusioni gettano forti dubbi sulla possibilità di utilizzare il MLE in modelli dinamici con dati in <u>panel</u> con una ridotta dimensione temporale, a meno di non avere forti "a priori" sulla natura delle osservazioni iniziali.

In un recente articolo, Bhargava e Sargan (1983), hanno tentato un nuovo approccio in proposito. In breve, essi considerano i dati come provenienti da un sistema di T equazioni simultanee, ognuna delle quali genera una cross section al tempo t con N osservazioni per ogni variabile e suggeriscono di usare il metodo della massima verosimiglianza a informazione limitata per stimare simultaneamente il sistema di equazioni sotto il vincolo di eguaglianza tra i coefficienti delle diverse equazioni [33]. L'appropriata funzione di verosimiglianza viene derivata sotto due diverse ipotesi sui valori y_{i0} (una, meno realistica, di esogeneità e una, più realistica e generale, in cui essi sono funzione di tutte le osservazioni passate delle variabili esogene e dell'errore che include gli effetti individuali) e di varie assunzioni sulla forma della matrice varianze e covarianze degli errori (incluso il caso di Balestra e Nerlove)[34]. La consistenza del metodo proposto non viene dimostrata, ma i risultati di un esperimento Montecarlo (con N = 100 e T = 9) sono positivi. L'applicazione ad una stima di una funzione di reddito rivelano tuttavia che, come in Nerlove (1971-a), l'assunzione di y_{i0} fissi conduce a boundary solutions, mentre l'ipotesi alternativa produce stime ragionevoli dei coefficienti del modello.

5.3 Test sulla eguaglianza dei coefficienti - I test F descritti nel par. 2.3 sulla possibile uguaglianza degli effetti α_i possono ancora essere usati (data la consistenza del DVE), nel contesto di effetti casuali, come test dell'ipotesi che $\sigma_\alpha^2 = 0$. Tuttavia Breusch e Pagan (1980) hanno proposto un test alternativo dimostrando che nell'ipotesi che σ_α^2 sia uguale a zero la statistica:

$$g = \frac{NT}{2(T-1)} \left\{ \Sigma_i (\Sigma_t \hat{\varepsilon}_{it})^2 - 1 \right\}^2$$

dove $\hat{\varepsilon}_{it}$ sono i residui da una regressione OLS con intercetta comune, è distribuita asintoticamente come $\chi^2_{(1)}$. Si noti infine che il test riportato non può essere usato per modelli dinamici con T piccolo; Pudney (1983, pp. 15-17), ha suggerito un test per coprire questo caso.

6. Effetti fissi e casuali: la critica di Mundlak

6.1 Effetti fissi o casuali? - Come detto in precedenza, l'assunzione di effetti casuali risultò predominante almeno fino alla fine degli anni settanta. Tra i vantaggi pratici di questa assunzione i seguenti vengono più spesso ricordati:

a) dato che il ricercatore è abitualmente interessato nei coefficienti angolari, i modelli a effetti fissi sprecano un gran numero di gradi di libertà se N, come spesso accade, è elevato (Balestra e Nerlove (1966), Maddala (1971-a), Swamy e Arora (1972) e Nickell (1981));

b) il modello a effetti fissi non utilizza una parte dell'informazione, e cioè la variazione delle variabili tra individui (Maddala (1971-a), Nickell (1981));

c) l'assunzione di casualità è un compromesso tra due assunzioni irragionevolmente forti, che gli effetti siano tutti diversi e che siano tutti uguali (Maddala (1971-a p. 346), Feige e Swamy (1974 p. 251)).

Tuttavia, prima che in termini di convenienza, la scelta tra i due approcci dovrebbe essere logica; in proposito si osserva che:

a) gli effetti individuali rappresentano variabili non osservabili costanti nel tempo su cui esiste scarsa o nessuna informazione sicchè una inclusione esplicita non è possibile; perciò, come si riassume nel residuo ε_{it} l'effetto della "general ignorance", anche questa "specific ignorance" dovrebbe essere inclusa nell'errore della regressione (Maddala (1971-a) e Nerlove (1971-a));

b) quando N individui sono estratti casualmente da una popolazione più ampia, come spesso accade, gli effetti individuali possono certamente essere visti come dipendenti da una scelta casuale (Scheffè (1959 p. 202), Wallace e Hussain (1969 p. 60), Fuller e Battesee (1973 p. 627), Mundlak (1978 - a))[35].

c) Chamberlain (1980 p. 233) suggerisce l'uso del criterio di scambiabilità di De Finetti che è condizione necessaria e sufficiente per la casualità del campione: il criterio è di considerare il campione $\alpha = (\alpha_1 \ldots \alpha_N)$ e vedere se è possibile scambiare α_i e α_j mantenendo la stessa distribuzione probabilistica (Hausman 1978 p. 262). Questo non costituisce solo un criterio per la casualità, ma anche per la assunzione che gli α_i sono indipendenti e identicamente distribuiti.

In merito alla scelta tra effetti fissi o casuali un importante contributo è stato fornito da Mundlak (1978-a) il quale ha sostenuto che la scelta sulla natura degli effetti non implica di per sè una scelta sul metodo di stima, dato che il DVE potrebbe essere giustificato, ed in realtà persino raccomandato, anche sotto l'ipotesi di effetti casuali[36].

6.2 <u>La critica di Mundlak</u> - I vantaggi del DVE diventano evidenti ogniqualvolta gli effetti possano ritenersi dipendenti dagli altri regressori. Seguendo Mundlak (1978-a) si consideri ancora:

$$(30) \qquad \underline{Y} = X\underline{\beta} + Z\underline{\alpha} + \underline{\varepsilon}$$

aggiungendo questa volta l'equazione ausiliaria:

$$\alpha_i = \underline{x}_{it}\underline{\pi} + w_{it} \quad \text{e la sua media nel tempo}$$

$$(31) \qquad \alpha_i = \underline{x}_{i.}\underline{\pi} + w_{i.} \qquad \text{con } w_i \sim (0, \omega^2)^{(37)}$$

dove $\underline{\pi}$ è un vettore KX1, $x_{i.}$ è il vettore delle medie temporali degli x per l'individuo i, e similmente per $w_{i.}$.

Data la (31), $Z\underline{\alpha}$ è equivalente a:

(32) $\quad Z\underline{\alpha} = P(X\underline{\pi} + \underline{w})$

dove P è la matrice definita nel par. 4.1; sostituendo la (32) nella (30) si ottiene:

$$Y = X\underline{\beta} + PX\underline{\pi} + P\underline{w} + \underline{\varepsilon} \qquad \text{o}$$

(33) $\quad Y = X\underline{\beta} + PX\pi + \underline{v} \quad$ con $\underline{v} = \underline{\varepsilon} + P\underline{w} \sim (0, \sigma^2 I_{NT} + T\,\omega^2 P)$

che ha la stessa distribuzione che nel modello di Balestra e Nerlove.

Sulla base di questo schema Mundlak sostenne che il lungo dibattito sulla natura degli effetti era stato nel contempo arbitrario ed inutile. Si può infatti assumere dall'inizio che gli effetti siano casuali e vedere l'inferenza del modello a effetti fissi come condizionale agli effetti presenti nel campione. In quest'ultimo caso il ricercatore è interessato al valore atteso di y condizionale agli x e agli effetti inclusi nel campione, cioè con la parte sistematica della (30), mentre nel modello a effetti casuali è la parte sistematica della (33) ad essere rilevante.

Naturalmente, se si è interessati al valore atteso non condizionale (il che produrrebbe stime più efficienti dato che viene introdotta l'informazione sulla distribuzione degli effetti) si deve tenere in adeguata considerazione il termine $PX\pi$, la cui omissione produrrebbe una distorsione da variabili omesse. Al contrario, il

DVE essendo uno stimatore condizionale non dipende dalla distribuzione degli α_i ed è perciò non distorto, indipendentemente dalla possibile correlazione tra effetti e regressori[38].

7. Sviluppi successivi

7.1 Test dell'indipendenza tra effetti e regressori e stime con variabili strumentali - La critica di Mundlak indirizzò l'attenzione dei ricercatori verso due obiettivi: trovare test capaci di accertare l'ipotesi di indipendenza tra effetti individuali e regressori e cercare di incorporare nel modello l'informazione relativa alla distribuzione degli α_i. Infatti, se questa distribuzione fosse conosciuta sarebbe possibile stimare il modello completo (le equazioni (30) e (31)) o una sua forma ridotta. Quest'ultimo tentativo è stato compiuto da Chamberlain (1980 e 1982) anche se questa procedura può essere giudicata arbitraria dato che:

"the estimates obtained often depend crucially on the distributional assumptions made about the individual effects, something on which economic theory has little to say" (Nickell 1981 p. 1418 n. 5).

Oltre a ciò, Mundlak (1978-a) ha anche provato che quando gli effetti dipendono linearmente da tutti gli x inclusi (e solo da questi), a meno di un errore casuale, lo stimatore efficiente GLS della (33) è lo stesso DVE[39]. Perciò una alternativa più efficiente è disponibile solo se è possibile assumere che qualche regressore non è correlato con gli effetti.

Un interessante trattamento di questi problemi si trova, con riferimento al problema del rendimento dell'educazione scolastica (cfr. per es. Griliches (1975)) in Hausman e Taylor (1981). Generalmente, la relazione tra reddito e anni di educazione prende la forma seguente:

(34) $Y = X\beta + Z\gamma + \alpha^{o} + \varepsilon$ [40]

dove X è una matrice di variabili osservate che variano nel tempo e tra individui, Z è una matrice di variabili osservate, come gli anni di scuola, variante tra individui, ma costante nel tempo e α^{o} è un vettore di effetti non osservati (come l'abilità) diversi tra individui e costanti nel tempo (vedi par. 4.1). Si ritiene comunemente che esista una correlazione tra educazione e abilità, ad es. perchè gli studenti più dotati sono anche quelli che proseguono gli anni di studio più a lungo. Il BNE produrrebbe perciò stime inconsistenti. D'altro canto l'uso del DVE non consentirebbe la stima dei parametri γ , dato che gli Z sono fissi nel tempo. In un caso simile è utile prima di tutto sottoporre a test l'esistenza di correlazione tra gli α_i ed i regressori. Un simile test fu proposto da Hausman (1978) e applicato al modello (34) da Hausman e Taylor (1981)[41]. Il test è basato sul fatto che sotto l'ipotesi che α_i è indipendente da X e Z, entrambi gli stimatori <u>between</u> e <u>within</u> danno stime consistenti di β , e cioè $\underset{N \to \infty}{\text{plim}}\ \hat{q} = \underset{N \to \infty}{\text{plim}}(\hat{\beta}_w - \hat{\beta}_b) = 0$[42].

Dato che $\hat{\beta}_w$ e $\hat{\beta}_b$ non sono correlati[43], vale la relazione $V(\hat{q}) = V(\hat{\beta}_w) - V(\hat{\beta}_b)$, dove V(.) indica la matrice asintotica varianze-covarianze dello stimatore. Grazie a ciò, in Hausman (1978) si dimostra che la statistica $m = \hat{q}'\left[(\hat{V}(\beta_w) - \hat{V}(\beta_b)\right]^{-1}\hat{q}'$ è distribuita come $\chi^2_{(K)}$ dove K è il numero dei regressori della regressione <u>between</u> e $\hat{V}$ è una stima consistente di V.

Rigettare l'ipotesi di non correlazione non implica necessariamente l'uso del DVE. Una alternativa superiore, che consente anche la stima di γ, è disponibile purchè un sufficiente numero di strumenti per le colonne di Z correlate con gli effetti possa trovarsi tra le colonne di X. Infatti, date le stime consistenti β_w, moltiplicando per P (la matrice definita al par. 4.1) il vettore $\hat{d}$ dei residui d_{it} della regressione <u>within</u>, si ottiene un vettore $\hat{d}_{i.}$ tale che:

(35) $P\hat{d} = \hat{d}_{i.} = y_{i.} - X_{i.}\beta_w = PY - X_{i.}\beta_w$

Sostituendo la (11) e la (34) nella (35), $\hat{d}_{i.}$ può essere scritto come:

(36) $\hat{d}_{i.} = Z\gamma + w$

dove $w = \overset{o}{\alpha}_i + \left[P - X_{i.}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\right]\varepsilon$. Dato che l'errore w è correlato con Z, è impossibile ricavare dalla (36) stime consistenti di γ con OLS. Tuttavia se esiste un numero di colonne in X non correlate con α per fornire un sufficiente numero di strumenti per le colonne di Z correlate con α_i uno stimatore a variabili strumentali (γ_w) può essere usato e sarà consistente [(44)]. Infine dalle stime consistenti β_w e γ_w è anche possibile costruire stime consistenti dei componenti della varianza da usare per ottenere stime asintoticamente efficienti di β e γ [(45)].

7.2 <u>Stime consistenti in regressioni dinamiche</u> - La precedente discussione ha mostrato che, nel rilevante caso di regressioni con <u>panel</u> di dimensione temporale ridotta, i metodi finora considerati non producono stime consistenti ($N \rightarrow \infty$), indipendentemente dalle assunzioni sulla natura degli effetti (46).

La soluzione naturale al problema sembra rinvenirsi nell'uso di strumenti per le variabili dipendenti ritardate. A questo proposito, quattro diverse procedure si ritrovano nella letteratura:

a) Balestra e Nerlove (1966) e Nerlove (1971-a) suggeriscono l'uso della variabile esogena sfasata $x_{i,t-1}$ come strumento per $y_{i,t-1}$; una applicazione di questo metodo si trova in Marcus (1983);

b) Anderson e Hsiao (1981 p. 604) raccomandano la seguente procedura. Data l'equazione $y_{it} = \vartheta y_{i,t-1} + \alpha_i + \varepsilon_{it}$, prendendo le differenze prime si ottiene:

$$y_{it} - y_{i,t-1} = \vartheta(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

è chiaro che sia $y_{i,t-2}$ o $\Delta y_{i,t-2}$ possono essere usati come strumenti per $\Delta y_{i,t-1}$ (47). Si può dimostrare che l'efficienza relativa (in termini di minore varianza asintotica) di questi due strumenti è una funzione dei coefficienti del modello ϑ e ϱ (il coefficiente di correlazione fra classi) che sono sconosciuti. Tuttavia, come regola approssimativa, gli autori suggeriscono l'uso di $y_{i,t-2}$ se si ritiene che ϑ sia positivo e T non è troppo corto.

c) Pudney (1983 pp. 8-10) nota che, se una variabile esogena è presente nel modello , ogni valore ritardato di questa variabile rappresenta un potenziale strumento e suggerisce una procedura per combinare ottimalmente i <u>set</u> di stime ottenute usando come strumenti $x_{i,t-1}$, i primi due ritardi su x_t, i prime tre, ecc. interrompendosi quando i problemi di calcolo si fanno più complessi.

d) Nickell (1980) suggerisce, in luogo dell'uso di variabili strumentali, la correzione delle stime ottenute col DVE $\hat{\delta}$ (il vettore delle stime dei coefficienti delle esogene) e $\hat{\vartheta}$ per il <u>bias</u> asintotico derivato in Nickell (1980 e 1981) (cfr. par. 3.2). In termini di variabili osservate, Nickell esprime le stime corrette come:

$$(37) \qquad \begin{aligned} \vartheta &= \hat{\vartheta} - (\tilde{Y}'_{-1} M \tilde{Y}_{-1})^{-1} A(\delta,\vartheta) \\ \delta &= \hat{\delta} + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}_{-1}(\tilde{Y}'_{-1} M \tilde{Y}_{-1})^{-1} A(\delta,\vartheta) \end{aligned}$$

dove $A(\delta\,\vartheta) = \dfrac{-\sigma_\varepsilon^2(\delta,\vartheta)}{T(1-\vartheta)} \left\{ 1 - \dfrac{1}{T} \; \dfrac{(1-\vartheta^T)}{1-\vartheta} \right\}$

e $\quad \sigma_\varepsilon^2(\delta,\vartheta) = \dfrac{1}{NT} \; (\tilde{Y} - \vartheta \tilde{Y}_{-1} - \tilde{X}\beta)'\,(\tilde{Y} - \vartheta \tilde{Y}_{-1} - \tilde{X}\beta)$

e gli altri simboli sono definiti nel par. 3.2.

Ottenere stime consistenti dei coefficienti angolari rappresenta solo il primo passo nella procedura per ricavare stime efficienti nell'ipotesi di effetti casuali. Una volta raggiunte stime consistenti dei parametri, diventa possibile ottenere stime consistenti dei componenti della varianza (sul tipo di quelle indicate nel par. 5.1) ed usarle per costruire uno stimatore strumentale con minimi quadrati generalizzati che tenga in considerazione la natura non scalare della matrice varianze-covarianze dei residui (cfr. Pudney (1983 pp. 10-12) e Anderson e Hsiao (1981 p. 604))[48].

L'uso di dati in panel nell'ambito di modelli dinamici è stato di recente considerato anche in riferimento alla stima di ritardi distribuiti. E' noto che la stima di tali modelli su dati aggregati richiede l'assunzione di valide restrizioni sul valore dei coefficienti. Pakes e Griliches (1984) considerano in che modo il problema delle restrizioni si ponga quando si disponga di dati longitudinali con una modesta estensione temporale. Oltre che ad estendere al caso di panel data l'imposizione di restrizioni sui coefficienti, gli autori mostrano come l'identificazione dei primi coefficienti di ritardi distribuiti possa conseguire dalla assunzione di restrizioni sul processo stocastico che genera le variabili ritardate e descrivono uno stimatore ML per la stima consistente del modello indicato[49]

7.3 <u>Autocorrelazione dei residui</u> - Un altro effetto della critica di Mundlak è consistito nel rinnovato interesse per il modello a effetti fissi che è stato arricchito per tener conto di una possibile matrice varianze-covarianze dei residui non scalare (cfr. Keifer (1980) o Bhargava et al. (1982)), caso in cui il DVE è consistente, ma non efficiente.

In particolare, Bhargava et al.(1982) hanno sviluppato un test per l'ipotesi che i disturbi seguano il processo:

$$(38) \qquad u_{it} = \vartheta u_{i,t-1} + \varepsilon_{it}$$

con ε_{it} <u>white noise</u>. La statistica suggerita:

$$d_p = \frac{\Sigma_i \Sigma_t (\hat{e}_{it} - \hat{e}_{i,t-1})^2}{\Sigma_i \Sigma_t \hat{e}_{it}}$$

(dove $\hat{e}_{it}$ sono i residui di una regressione consistente col DVE) è una estensione della statistica DW e la sua distribuzione dipende dal numero dei regressori, dal numero dei periodi e da quello degli individui. Valori critici dei limiti inferiori e superiori di d_p sono tabulati nell'articolo per T=6 e 10 e per vari valori di N e K e viene indicato come calcolarli per valori differenti da T, N e K. Nel caso in cui l'ipotesi di autocorrelazione non possa venir rigettata si suggerisce l'uso di GLS basato su uno stimatore di ϑ (50). La procedura per ricavare ϑ valida per N elevato, è basata sul fatto che $\vartheta_d = 1-d_p/2$ fornisce una stima consistente (per $T \to \infty$) di ϑ e che la relazione tra E (ϑ_d) e ϑ per campioni piccoli è data da:

$$E(\vartheta_d) = 1 - \frac{1-\vartheta(T-1)}{T - \frac{1+\vartheta}{1-\vartheta} + \frac{2\vartheta(1-\vartheta^T)}{T(1-\vartheta)^2}}$$

sostituendo a $E(\vartheta_d)$ il valore $1 - d_p/2$, la relazione precedente può essere risolta iterativamente per ϑ.

Infine viene anche descritto un test per l'ipotesi che i residui siano generati da un random walk.

L'introduzione di autocorrelazione temporale dei residui è tipica del BNE ma si è visto come in tal caso la correlazione sia di tipo particolare. Una estensione in questo campo è dovuta a Lillard e Willis (1978) che considerano un errore $\mu_{it} = \alpha_i + v_{it}$ con $v_{it} = v_{i,t-1} + \varepsilon_{it}$ con ε_{it} white noise. Ciò equivale ad assumere che la correlazione degli errori per lo stesso individuo declini asintoticamente a $\sigma^2_\alpha = E(\alpha_i^2)$. I due autori però non derivano un test della loro assunzione. Inoltre i coefficienti angolari del modello sono stimati con OLS con intercetta comune mentre la stima dei componenti della varianza, i parametri in cui Lillard e Willis sono più interessati, è ottenuta con la procedura ML dai residui della regressione OLS. In un lavoro successivo Lillard e Weiss (1979) stimano un modello leggermente più complicato con GLS, ancora basati su una stima ML dei componenti della varianza.

8. Modelli con coefficienti diversi tra individui

Abbastanza stranamente, fino agli anni più recenti la letteratura sui modelli in cui tutti i coefficienti sono diversi tra individui si era sviluppata indipendentemente da quella sui modelli con intercetta differenziata considerati nei paragrafi precedenti. Una ragione di questa separazione risiede nel fatto che i modelli in esame in questo e nel seguente paragrafo, come il modello SURE, non furono ideati per trattare specificatamente dati in panel e trovarono una vasta applicazione anche ad altri problemi. D'altro canto, i ricercatori che utilizzano dati in panel sembrano aver trovato una più soddisfacente approssimazione della realtà in modelli con intercetta variabile, certo anche in conseguenza della dimensione temporale abitualmente modesta dei dati in panel.

In effetti, i modelli con coefficienti diversi tra individui, dato l'elevato numero di parametri da stimare, richiedono un valore elevato sia per T che per N. E' questo il motivo per cui nei paragrafi seguenti scarso rilievo verrà dato al problema di un aumento delle osservazioni nella dimensione temporale piuttosto che in quella individuale: entrambe saranno di massima assunte essere sufficientemente ampie. Tuttavia, bisogna sottolineare che molti dei problemi incontrati nel contesto dei modelli a intercetta variabile sono comuni al caso di coefficienti variabili, in particolare quelli relativi alla scelta tra coefficienti fissi o stocastici, e alla possibile uguaglianza dei coefficienti tra individui.

9. Coefficienti non stocastici

9.1 Il modello SURE di Zellner - Il prototipo di modello con coefficienti non stocastici è dato dal modello SURE (seemingly unrelated regression equations) di Zellner (1962). In questo modello tutti i coefficienti sono non stocastici, ma differiscono tra individui. Il sistema può essere perciò rappresentato come

$$(39) \qquad \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & & & & 0 \\ & \ddots & & & \\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ v_N \end{bmatrix}$$

ovvero: $$(40) \qquad \underline{Y} = X \underline{\beta} + \underline{v}$$

~~dove~~ Y_i e V_i sono vettori Tx1 delle osservazioni per l'i-esimo individuo, X_i è una matrice TxK di variabili esogene, mentre β_i è un vettore Kx1 dei coefficienti, diverso per ogni individuo(51). Dato che tutti i coefficienti differiscono tra individui e sono non stocastici, l'unico modo per aumentare l'efficienza delle stime rispetto a OLS su dati individuali si manifesta se è possibile assumere l'esistenza di una correlazione degli errori relativi a diversi individui. La principale assunzione di Zellner è che gli errori siano correlati tra individui allo stesso tempo, ma non correlati nel tempo, consentendo una diversa correlazione tra due diverse coppie di individui:

$$E(vv') = \Omega = \begin{bmatrix} \begin{matrix}\sigma_{11} & 0\\ 0 & \ddots & \\ & & \sigma_{11}\end{matrix} & \begin{matrix}\sigma_{12} & 0\\ 0 & \ddots & \\ & & \sigma_{12}\end{matrix} & \cdots & \begin{matrix}\sigma_{1N} & 0\\ 0 & \ddots & \\ & & \sigma_{1N}\end{matrix} \\ \begin{matrix}\sigma_{21} & 0\\ 0 & \ddots & \\ & & \sigma_{21}\end{matrix} & \ddots & & \\ \vdots & & \ddots & \\ \begin{matrix}\sigma_{N1} & 0\\ 0 & \ddots & \\ & & \sigma_{N1}\end{matrix} & & & \begin{matrix}\sigma_{NN} & 0\\ 0 & \ddots & \\ & & \sigma_{NN}\end{matrix} \end{bmatrix} = \varphi \otimes I_T \quad \text{dove: } \varphi = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{N1} & \cdots & \cdots & \sigma_{NN} \end{bmatrix}$$

Data la natura non scalare di Ω , lo stimatore corretto ed efficiente è costituito dai GLS. Tuttavia Zellner (1962, p. 351) dimostra che i risultati in tal modo ottenuti sarebbero equivalenti a OLS individuali, non solo se Ω è diagonale [52], ma anche nel caso in cui $X_1 = \ldots\ldots = X_n$ (cfr. anche Dwivedi e Srivastava (1978)).

Come al solito, Ω non è noto e deve essere stimato. Zellner suggerisce l'uso del seguente stimatore di σ_{ij} ricavato dai residui $\hat{v}_i$ e $\hat{v}_j$ di OLS individuali:

$$\sigma_{ij} = \frac{\hat{v}'_i \, \hat{v}_j}{T - K}$$

Con le abituali assunzioni, lo stimatore GLS con Ω stimato ha la medesima distribuzione asintotica del vero GLS ed è dato da:

(41) $\beta = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y$

con varianza stimata come $(X' \hat{\Omega}^{-1} X)^{-1}$.

Si noti che, data la forma di Ω , non è necessario invertire direttamente tale matrice che è di ordine NTxNT [53]. Dhrymes (1971) suggerisce una procedura alternativa, equivalente al MLE, che utilizza i residui della (41) per ristimare Ω fino a convergere.

Nonostante lo stimatore SURE sia asintoticamente più efficiente degli OLS (il guadagno in efficienza è tanto maggiore quanto più gli errori sono correlati e quanto meno correlate sono le variabili esplicative (cfr. Zellner (1962, par. 3.2)), il rendimento relativo in campioni piccoli è ancora questione aperta. Risultati analitici sono stati derivati solo sotto assunzioni molto particolari (vedi Judge et al. (1980, p. 248) per una breve rassegna) e concludono per una maggiore efficienza dello stimatore SURE. Esperimenti Montecarlo, che mostrano come l'efficienza dello stimatore di Zellner aumenti al crescere di N, sono riportati in Kmenta e Gilbert (1968), i quali derivano il MLE sotto ipotesi di normalità.

9.2 <u>Tests sull'uguaglianza dei coefficienti</u> - Zellner sottolinea l'importanza di sottoporre a test l'ipotesi di uguaglianza dei coefficienti per diversi individui; il test suggerito è equivalente al test F esposto nel par. 2.3 con la differenza che i residui, sia vincolati che non vincolati, devono essere relativi a equazioni che tengano conto della struttura di correlazione dei disturbi. Perciò, mentre i residui non vincolati sono ricavati dalla (41), i residui vincolati sono quelli della stima GLS dell'equazione:

$$(42) \quad Y = Z\,\beta + V \quad \text{dove } Z = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_N \end{bmatrix} \quad \text{(una matrice NTxK)}$$

e V è distribuito come nella (40). Come in precedenza una stima per σ è necessaria e in questo caso sembra più appropriata una stima dai residui della (42). Nell'ipotesi che i coefficienti siano eguali tra individui, la statistica rilevante:

(43) $g = \dfrac{(RRSS - URSS)/K(N-1)}{URSS/(NT-K)}$ (54)

a causa della stima delle varianze, è distribuita come $F(K(N-1), NT-K)$ solo asintoticamente (per $T \to \infty$) (Theil (1971, p. 402)).

9.3 Autocorrelazione dei residui - Una rilevante estensione del modello SURE fu suggerita da Parks (1967) con l'introduzione di autocorrelazione del primo ordine degli errori individuali:

$$v_{it} = \varrho_i v_{i,t-1} + \varepsilon_{it} \quad (55)$$

La matrice varianze covarianze dei residui è in questo caso:

$$\Omega = E(vv') = \begin{bmatrix} \Omega_{11} & \cdots & \Omega_{1N} \\ \vdots & \ddots & \vdots \\ \Omega_{N1} & \cdots & \Omega_{NN} \end{bmatrix}$$

dove: $$\Omega_{ij} = E(v_i v_j') = \frac{\sigma_{ij}}{1 - \varrho_i \varrho_j} \begin{bmatrix} 1 & \varrho_j & \cdots & \varrho_j^{T-1} \\ \varrho_i & 1 & & \varrho_j^{T-1} \\ \vdots & & \ddots & \vdots \\ \varrho_i^{T-1} & \varrho_i^{T-2} & \cdots & 1 \end{bmatrix}$$

Ancora, se Ω fosse noto, i GLS sarebbero corretti ed efficienti sotto le usuali assunzioni sulla esogenità dei regressori.

L'inversione di Ω (di ordine NTxNT) può essere difficile da calcolare e viene suggerito (Parks (1967), Judge et al. (1980, p. 262)) di trasformare le osservazioni moltiplicandole per una matrice H tale che $H \Omega H' = \Psi \otimes I_T$, la matrice varianze-covarianze del modello di Zellner. Questa procedura porta all'uso delle seguenti variabili trasformate:

(44) $y^*_{i1} = \sum_j \alpha_{ij} y_{j1} \qquad i=1,\dots,N$

(45) $y^*_{it} = y_{it} - \rho_i \, y_{i,t-1}$ $i=1,\dots,N$ $t=2,\dots,T$

il che significa che mentre le prime osservazioni sono sottoposte ad una trattamento speciale, tutte le altre osservazioni sono modificate dalla trasformazione usuale per residui AR(1) (56). Dato che la procedura per ottenere gli α_{ij} della (44) è piuttosto onerosa, e purchè T non sia troppo piccolo, può risultare conveniente ignorare le prime osservazioni e trasformare i dati come nella (45). Questo GLS approssimato è naturalmente meno efficiente, ma nulla può essere detto quando invece delle vere varianze si usano delle stime. Esperimenti Montecarlo (Kmenta e Gilbert (1970)) suggeriscono che ci sia poca differenza dal punto di vista pratico tra la procedura esatta e quella approssimata e che entrambe diano risultati migliori dello stimatore di Zellner, quando il modello sottostante abbia in realtà errori temporalmente autocorrelati (57).

Scegliendo la procedura semplificata, i parametri ρ_i possono essere stimati dai residui di N regressioni OLS individuali (58). Le variabili, trasformate come nella (45), sono poi usate per stimare le covarianze degli errori tra individui σ_{ij} come:

$$\hat{\sigma}_{ij} = \frac{\hat{v}'_i \, \hat{v}_j}{T-K-1}$$

Infine la stima dei coefficienti è ottenuta da:

$$\hat{\gamma} = (X^{*\prime} \hat{\Omega}^{-1} X^*)^{-1} X^{*\prime} \hat{\Omega}^{-1} Y^*$$

con matrice varianze covarianze $E(\hat{\gamma}\,\hat{\gamma}') = (X^{*\prime} \hat{\Omega}^{-1} X^*)^{-1}$

dove l'asterisco indica la trasformazione secondo la (45) delle variabili originali. Parks (1967) dimostra che, sotto semplici condizioni, lo stimatore ottenuto in questo modello è asintoticamente ($T \rightarrow \infty$) equivalente al GLS che utilizzi la vera matrice varianze-covarianze.

Normalmente è consigliabile un test sulla eguaglianza dei parametri tra individui, tenendo conto della struttura di autocorrelazione dei residui. Il test da usare è quello della (43); il modello vincolato, da cui ricavare RRSS, è descritto, per es., da Kmenta (1971, pp. 512-513).(59)

10 - Modelli con coefficienti stocastici

10.1 Il modello di Swamy

- I modelli a coefficienti casuali non hanno trovato una applicazione tanto ampia quanto il modello SURE a coefficienti non stocastici, anche se si restringe l'analisi a modelli con coefficienti diversi tra individui e non nel tempo(60). In quest'area il modello più rilevante (e praticamente l'unico) è stato suggerito da Swamy (1970), cui si devono anche numerosi sviluppi ulteriori e applicazioni empiriche.

Swamy (1970) ha considerato un modello in cui le osservazioni sono generate da:

(46) $y_{it} = \Sigma_k \beta^k_i x^k_{it} + u_{it} \qquad i=1....N,\ t=1....T$

dove u_{it} è un errore distribuito indipendentemente con media 0 e varianza σ^2_i (ammettendo quindi eteroschedasticità dei residui(61)), x^k_{it} sono regressori non stocastici e β^k_i rappresenta coefficienti casuali con media β^k e matrice varianze-covarianze $E(\beta_i \beta'_i) = \Delta$ dove β_i è il vettore dei coefficienti dell'individuo i.

Questo significa che per ogni individuo i coefficienti includono una componente casuale e che queste componenti sono correlate tra di loro per lo stesso individuo, ma non correlate tra individui.

Il modello completo perciò può scriversi come:

$$(47) \qquad \underline{Y} = Z\underline{\beta} + \underline{u} = \begin{bmatrix} X_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & X_N \end{bmatrix} \begin{bmatrix} \underline{\bar{\beta}} + \underline{\delta}_1 \\ \vdots \\ \underline{\bar{\beta}} + \underline{\delta}_N \end{bmatrix} + \begin{bmatrix} \underline{u} \\ \vdots \\ \underline{u}_N \end{bmatrix}$$

dove $\underline{\bar{\beta}}$ è il vettore Kx1 delle medie $\left[\bar{\beta}^1 \cdots\cdots \bar{\beta}^K\right]$, $\underline{\delta}_i$ è il vettore Kx1 dei componenti casuali dei coefficienti per l'individuo i e $\underline{u}_i$ è il vettore Tx1 $\underline{u}'_i = \left[u_{i1} \cdots\cdots u_{iT}\right]$

Il modello (47) può essere scritto come:

$$(48) \qquad Y = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \underline{\bar{\beta}} + \begin{bmatrix} X_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & X_N \end{bmatrix} \begin{bmatrix} \underline{\delta}_1 \\ \vdots \\ \underline{\delta}_N \end{bmatrix} + u = X\underline{\bar{\beta}} + Z\delta + u$$

dove $X' = (X'_1 \cdots\cdots X'_N)$ e $\delta' = (\underline{\delta}'_1 \cdots\cdots \underline{\delta}'_N)$.

Sotto l'ipotesi che gli x siano non stocastici (o equivalentemente che si faccia inferenza condizionale alla loro distribuzione), la stima di $\underline{\bar{\beta}}$ può essere compiuta con OLS e risulterebbe consistente anche se non efficiente. Lo stimatore efficiente sarebbe però ottenuto usando GLS. La matrice varianze-covarianze dell'errore della (48) è:

$$\Omega = \begin{bmatrix} \Omega_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \Omega_N \end{bmatrix} \quad \text{dove} \quad \Omega_i = X_i \Delta X_i' + \sigma_{ii} I$$

e i GLS sono dati da:

$$(49) \qquad \hat{\underline{\beta}} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \underline{Y} = (\Sigma_j X_j' \Omega_j^{-1} X_j)^{-1} \Sigma_i X_i' \Omega_i^{-1} \underline{Y}_i$$

Questa espressione ha due interessanti interpretazioni. Swamy (1970, p. 314) ha mostrato che la (49) è equivalente a:

$$(50) \qquad \hat{\underline{\beta}} = \Sigma_i W_i \underline{b}_i$$

$$\text{dove: } W_i = \left[\Sigma_j (\Delta + \sigma_j (X_j' X_j)^{-1})^{-1} \right]^{-1} \left[\Delta + \sigma_i (X_i' X_i)^{-1} \right]^{-1}$$

e $\underline{b}_i$ sono le stime dei coefficienti β_i da N regressioni OLS su dati individuali. La (50) mostra che i GLS sono in questo caso equivalenti a una media ponderata delle stime OLS individuali con pesi proporzionali alle relative matrici varianze covarianze $\sigma_i (X'_i X_i)^{-1}$ (62). Si noti infine che la (50) richiede solo una inversione di una matrice di ordine K ed è perciò preferibile dal punto di vista del calcolo alla (49). D'altro canto, Mundlak (1978-b) ha dimostrato che la (49) è equivalente a una media ponderata di stimatori <u>between</u> e <u>within</u>.

La stima con GLS richiede una stima della matrice Ω. Swamy ha proposto il seguente stimatore non distorto:

$$\hat{\sigma}_i = \frac{\hat{e}_i' \hat{e}_i}{N-K}$$

(dove $\hat{e}_i$ sono i residui della regressione OLS con dati dell'i-esimo individuo) e $\hat{\Delta} = S_b/(N-1) - (1/N)\ \Sigma_i\ \hat{\sigma}_i\ (X'_i X_i)^{-1}$ dove: $S_b = \Sigma_i\ b_i\ b'_i - (1/N)\ \Sigma_i\ b_i\ \Sigma_i\ b'_i$

E' naturalmente possibile che $\hat{\Delta}$ non risulti essere positiva semidefinita. Tra le varie possibilità Swamy (1971) ha raccomandato l'uso dello stimatore consistente $\hat{\Delta} = S_b/(N-1)$. Le proprietà di consistenza di questo GLS (discusse in Swamy (1970, pp. 316-318) per N e T tendenti entrambi a infinito) sarebbero mantenute, anche se nulla si può dire sul rendimento comparato dei due stimatori GLS in campioni piccoli. Come nel modello a coefficienti fissi, anche per il modello di Swamy è conveniente verificare l'ipotesi di uguaglianza dei coefficienti. Swamy (1970) ha dimostrato che nell'ipotesi che $\beta_1 = \ldots\ldots = \beta_N$ la statistica:

$$H_\beta = \Sigma_i \frac{(\underline{b}_i - \hat{\underline{\beta}})'\ X'_i\ X_i\ (\underline{b}_i - \hat{\underline{\beta}})}{\hat{\sigma}_i}$$

dove $\hat{\beta} = \left[\Sigma_i \dfrac{X'_i\ X_i}{\hat{\sigma}_i}\right]^{-1} \Sigma_i \dfrac{X'_i\ X_i}{\hat{\sigma}_i}\ b_i$ è distribuita come $\chi^2_{[K(N-1)]}$

per $T \rightarrow \infty$. Come indicato da Judge et al. (1980, p. 351) questo test può essere usato anche per l'ipotesi di coefficienti fissi; un test dell'eguaglianza dei coefficienti che è basato sull'assunzione di casualità (cioè un test di $\Delta = 0$) è derivato in Swamy (1971, pp. 122-124).

10.2 Estensioni del modello - Dopo l'articolo del 1970, Swamy ha prodotto svariate estensioni del suo modello:

a) Swamy (1971 e 1973) ha sviluppato il MLE del modello (47)

b) Swamy (1973 e 1974) ha introdotto autocorrelazione del primo ordine del residuo u_{it} (nessuna semplice trasformazione dei dati è tuttavia nota e questo stimatore richiede l'inversione di una matrice di ordine NTxNT);

c) Swamy (1974) ha considerato l'uso di variabili strumentali per il caso in cui la dipendente ritardata sia presente tra i regressori; tuttavia Pudney (1983, p. 32) ha mostrato che in questo caso nessun stimatore (GLS, IV o ML) produce stime consistenti.

10.3 Esperimenti Montecarlo - Esiste un interessante esperimento Montecarlo sulla validità dello stimatore di Swamy. Johnson e Lyon (1973) produssero dati da un modello dinamico con coefficienti variabili tra individui ed un errore omoschedastico [63]. Tra i vari stimatori utilizzati (OLS, DVE, BNE, Swamy), sorprendentemente, il miglior risultato venne fornito dal DVE che utilizzava dummies individuali e temporali. In una delle due specificazioni usate lo stimatore di Swamy produsse risultati molto peggiori del DVE e solo leggermente migliori nella seconda specificazione. Inoltre, il test per l'uguaglianza dei coefficienti suggerito da Swamy (1970), illustrato nel paragrafo 10.1, diede risultati insoddisfacenti, non essendo in grado di discriminare il precedente modello da un modello alternativo con coefficienti fissi e uguali per tutti gli individui (quasi sempre l'ipotesi nulla di uguaglianza dei coefficienti venne rigettata).

10.4 <u>Correlazione tra coefficienti e variabili esogene</u> - La critica di Mundlak, relativa ai problemi che sorgono con inferenza statistica non condizionale ai coefficienti presenti nel campione, originariamente diretta al caso di intercetta diversa tra individui, è stata estesa da Mundlak (1978-b) al modello di Swamy, visto in contrapposizione all'approccio a effetti fissi del modello SURE. Mundlak ha sostenuto di nuovo che lo stimatore a coefficienti fissi può essere visto come una stimatore condizionale ai coefficenti presenti nel modello; questa può essere una necessità nel caso di correlazione tra coefficienti e variabili esplicative. Una alternativa superiore esiste, di nuovo, se è possibile introdurre esplicitamente nel modello K regressioni ausiliarie che rappresentino la citata correlazione. Nel caso in cui i coefficienti dipendano da tutti i regressori inclusi (e solo da questi), Mundlak ha derivato l'appropriato stimatore GLS che è consistente per $N \rightarrow \infty$. Un test per verificare la dipendenza tra distribuzione dei coefficienti e regressori è stato ideato da Pudney (1983, p. 32)

NOTE

(1) Ampie esposizioni dell'econometria dei panel data si trovano in Maddala (1977) e, soprattutto, in Judge et al. (1980). Scarso o nessun rilievo viene tuttavia dato all'uso di dati in panel per regressioni dinamiche, un caso molto rilevante, nonchè all'uso di stimatori di massima verosimiglianza. Numerosi temi inerenti all'utilizzo di panel data sono trattati in AA. VV. (1978).

(2) Lo studio di modelli a più equazioni basati su dati longitudinali è stato finora abbastanza limitato (cfr. Maddala (1977, p. 332)), Joreskog (1978) e Pudney (1983, pp. 34-36)), anche se di recente numerose applicazioni si sono avute relativamente alla stima simultanea di equazioni di domanda di beni di consumo (cfr. Jorgenson, Lau e Stocker (1982)).

(3) Il problema del "selectivity bias", particolarmente importante nel caso di modelli non lineari, non verrà di conseguenza trattato; una dettagliata trattazione del problema può trovarsi in Maddala (1978).

(4) L'uguaglianza, tra individui, dei coefficienti è però una condizione sufficiente, ma non necessaria per l'assenza di bias di aggregazione nel senso definito da Theil (1954). Infatti se i coefficienti sono diversi, ma possono essere considerati come variabili casuali indipendenti derivanti dalla medesima distribuzione (come nel modello di Swamy descritto nel par. 10.1), le stime aggregate risulteranno non distorte (cfr. Zellner (1969)). Per una trattazione in italiano del problema del bias di aggregazione con riferimento ai vantaggi derivanti dall'uso di dati in panel si veda Marotta (1983).

(5) Svariate sono le ipotesi in cui la costante può risultare correlata con i regressori. Ad es., nella stima di una funzione di produzione, α_i può rappresentare un input non osservato, costante nel tempo, ma diverso fra individui (come la capacità manageriale) correlato con gli altri input del processo produttivo (cfr. Mundlak (1961)).

(6) E' stata adottata la convenzione di usare il simbolo K' per indicare il numero delle variabili esplicative, inclusa la costante, e K per indicare K'-1.

(7) Il problema di combinare o meno le osservazioni relative a diversi individui è considerato anche da Maddala (1971-b) che suggerisce di mantenere abbastanza alto il grado di significatività dei test F ("we should be using something like twenty five to thirty per cent level of significance in our preliminary test of significance" p. 949)

(8) E' da notare tuttavia che l'analisi della covarianza è stata estesa anche all'ipotesi di effetti casuali sicché il termine, seppur frequentemente usato, appare fuorviante.

(9) Si noti che la matrice varianze-covarianze dell'errore non è scalare:
$E(\tilde{\varepsilon}\,\tilde{\varepsilon}')=E(Q\tilde{\varepsilon}\,\tilde{\varepsilon}'Q)=QE(\tilde{\varepsilon}\,\tilde{\varepsilon}')Q=\sigma^2 QIQ=\sigma^2 Q$
Tuttavia, data l'idempotenza della matrice Q, la stima con minimi quadrati della (13) è efficiente:
$Var(\hat{\beta})=E\,(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = E\left[(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\varepsilon}\,\tilde{\varepsilon}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\right]=$

$=(\tilde{X}'\tilde{X})^{-1}\tilde{X}\sigma_{\varepsilon}^2\,Q\tilde{X}(\tilde{X}'\tilde{X})^{-1}$, ma $Q\tilde{X}=QQX=QX=\tilde{X}$ sicché tutto si semplifica a $\sigma_{\varepsilon}^2(\tilde{X}'\tilde{X})$. Naturalmente dato che Q è singolare, i minimi quadrati generalizzati non possono applicarsi alla (13). Una prova formale dell'efficienza degli OLS può infine basarsi sul fatto che in questo caso corrispondono ad uno stimatore di Aitken non ristretto basato sull'inversa generalizzata di Q (Swamy e Arora (1972)).

(10) Il problema non è risolvibile nell'ambito del modello con variabili _dummies_ perchè non è possibile distinguere tra l'effetto dell'intercetta individuale e quello dei regressori "_time invariant_", sicché la (14) fornirà una stima congiunta della costante e dell'effetto di tutte le variabili _time invariant_. La presenza di regressori costanti nel tempo non crea invece alcun problema sotto l'assunzione di effetti casuali.

(11) Il problema degli errori di misurazione in stime con dati in

panel è riconsiderato in Griliches e Hausman (1984), i quali suggeriscono l'uso combinato dello stimatore *within* e di stimatori basati su differenze tra periodi diversi per valutare gli effetti di errori di misurazione e per ottenere stime consistenti.

(12)Il contrario vale se si introducono effetti temporali (invece che individuali) dato che in questo caso la media è calcolata rispetto agli individui e non al tempo.

(13)Per una prova formale delle proprietà asintotiche del DVE si veda Anderson e Hsiao (1981). Se T è fisso il DVE è anche inconsistente, anche se tutti i regressori sono esogeni, se il modello non è lineare. Si veda Heckman (1979) per prove Montecarlo sulla dimensione della distorsione in alcuni modelli *probit*.

(14)Le stime OLS della (17) e della (18) sono equivalenti perchè $(X'PX)^{-1}=(1/T)\ (\bar{X}'\bar{X})^{-1}$ e $X'PY=T(\bar{X}'\bar{Y})$.

(15)Una rilevante relazione tra stimatori "*within*" e "*between*" è la seguente: la somma della matrice dei momenti dei due stimatori è uguale alla matrice dei momenti dello stimatore OLS che utililizza i dati individuali con un'unica intercetta. Infatti: $(X'QX)+(X'PX)=X'\left[Q+P\right]X=X'X$ dato che $Q+P=I_{NT}$. Questa relazione riassume un ben noto risultato di analisi della varianza, e cioè che la variazione totale delle variabili (considerata dagli OLS su osservazioni combinate) può essere divisa nella variazione delle variabili intorno alle medie individuali (considerata dallo stimatore "*within*") e T volte la variazione delle medie individuali intorno alla media generale (considerata dallo stimatore "*between*").

(16)Anche Mundlak (1961 e 1963) aveva considerato gli effetti individuali come parte dell'errore, ma era poi ricorso all'uso del DVE in ragione della possibile correlazione tra i regressori osservati e gli effetti. Inoltre, la maggior parte dei risultati teorici esposti in Balestra e Nerlove (1966) erano stati sviluppati in Nerlove (1965, appendice al cap. 7).

(17)E' stato notato che il modello (21) equivale ad un modello in cui si ipotizzano errori correlati per ogni individuo e non correlati tra individui (Nerlove (1971-a)). Nella (21) la struttura dell'autocorrelazione è però molto semplice, essendo la stessa per tutti gli individui e per ogni coppia di errori indipendentemente dalla loro distanza temporale.

(18)Questo passaggio è basato su una proprietà dell'algebra matriciale che può trovarsi, per es., in Maddala (1977, p. 445).

(19)O, come in Hausman e Taylor (1981, p. 1381):

$$\hat{\beta}_{BN}=\left[(V_b + V_w)^{-1} V_w\right]\hat{\beta}_b + \left[I-(V_b + V_w)^{-1} V_w\right]\hat{\beta}_w$$

dove $V_b=(\sigma_\alpha^2 +(1/T)\sigma_\varepsilon^2)_1 (X'PX)^{-1}$ (la varianza dello stimatore "between") e $V_w=\sigma_\varepsilon(X'QX)^{-1}$ (la varianza dello stimatore "within"); cfr. anche Taylor (1980, p. 206).

(20)Al contrario l'efficienza dello stimatore "between" non aumenta al crescere di N o T (Taylor (1981, p. 210)).

(21)Si noti che l'origine della distorsione del BNE in regressioni dinamiche risiede anche nella non corretta specificazione per variabili omesse (vedi par. 3.3). Perciò al crescere di T lo stimatore "between" rimane inconsistente e il BNE diventa consistente solo perchè il peso dato a tale stimatore tende a zero.

(22)In precedenti articoli Wallace e Hussain (1969) e, soprattutto, Nerlove (1971-b) avevano studiato le proprietà algebriche della matrice Ω , quando anche un effetto temporale è incluso, trovando la sua inversa e le sue radici e vettori caratteristici. Per il caso di effetti individuali e temporali la traformazione richiesta è derivata in Fuller e Battesee (1974). Per una descrizione di programmi fortran per la stima di modelli a effetti casuali (e di altri modelli, anche non lineari, con panel data) si veda Hall (1978).

(23)Assunzioni più complesse sulla struttura di correlazione degli errori vennero introdotte per tener conto, ad esempio, della possibile eteroschedasticità di ε_{it}. Per una applicazione si veda Griffiths e Anderson (1982) e Lillard e Willis (1978). Per una analisi teorica di modelli con una più complessa matrice varianze-covarianze (incluso in particolare il caso di eteroschedasticità) si veda Pudney (1983, pp. 22-31). Pudney (1983, pp. 16-19) presenta un test, per modelli statici e dinamici, dell'ipotesi che la matrice varianze-covarianze sia del tipo assunto da Balestra e Nerlove contro l'ipotesi di eteroschedasticità.

(24)Il modello a effetti casuali con l'inclusione aggiuntiva di effetti temporali è stato studiato da Wallace e Hussain (1969), Amemiya (1971), Maddala (1971-a par. 5), Swamy (1971), Swamy e Arora(1972) e Fuller e Battesee (1974). Nel caso di effetti temporanei casuali e effetti individuali fissi si veda Hussain (1969). Per il caso opposto (effetti temporali fissi e individuali casuali) si veda Lillard e Willis (1978). Per un approccio bayesiano si veda Chetty (1968) e Swamy e Mehta (1973).

(25)La letteratura sull'analisi della varianza è pure ricca di stimatori di componenti della varianza; si veda, per es., la rassegna di Searle (1971).

(26)Più in generale Swamy e Arora (1972) derivano una intera classe di stimatori asintoticamente efficienti e sostengono che la scelta tra questi può solo fondarsi sulle proprietà in piccoli campioni, qualcosa su cui poco si conosce. Tuttavia Swamy e Arora derivano questi risultati per N e T tendenti entrambi a infinito (e perciò anche il DVE è incluso nella classe di modelli asintoticamente efficienti). Cfr. anche Swamy (1971, cap. 3).

(27)La stima di massima verosimiglianza può naturalmente essere adottata anche per il caso di effetti fissi e si dimostra (Hildreth (1950), Chamberlain (1980), Anderson e Hsiao (1982, p. 52)) che il MLE è uguale al DVE sotto l'ipotesi di normalità. Tuttavia, a causa del classico problema dei parametri incidentali lo stimatore ML del modello a effetti fissi potrebbe in generale non raggiungere la consistenza per N tendente a infinito (due esempi di questo caso si trovano in Chamberlain

(1980, p. 229)). Sul problema si veda anche Keifer (1980). Infine, per una discussione del MLE nel contesto delle analisi della covarianza cfr. Harville (1977).

(28)La derivazione dello stimatore ML dei coefficienti della regressione e dei componenti della varianza si trova, per es., in Maddala e Mount (1973, p. 325).

(29)Per $\rho = 1$ Ω° è singolare e il MLE non è definito (Balestra e Nerlove (1966), p. 599).

(30)Boundary solutions possono anche essere attribuite ad una cattiva specificazione stocastica del modello (Swamy e Arora (1972)) e diverse specificazioni, che garantiscono la massimizzazione della funzione di verosimiglianza per valori plausibili dei parametri, sono state proposte (Swamy (1971, p. 88), Swamy e Mehta (1973) e Berzeg (1978)).

(31)Anderson e Hsiao (1982) considerano, con simili conclusioni, un modello più complesso con variabili esogene costanti e variabili nel tempo. Come nel precedente articolo, gli autori suggeriscono anche metodi per semplificare il calcolo per il MLE, una volta compiuta la scelta sulla natura delle osservazioni iniziali.

(32)Si può provare (cfr. Chowdury (1984)) che il MLE sotto l'ipotesi a) è ancora equivalente ad una combinazione lineare di stimatori within e between, ma la distorsione del primo è asintoticamente uguale e di segno opposto alla distorsione del secondo, sicché lo stimatore è consistente.

(33)In questo modo una equazione dinamica è trasformata in un sistema di T equazioni cross section che sono ovviamente non dinamiche in quanto nessun elemento del vettore della variabile dipendente di ogni equazione appare nel lato destro della stessa equazione.

(34)Sono presentati anche dei likelihood ratio test sulla validità delle diverse alternative proposte.

(35)Si è quindi indotti a usare gli effetti casuali quando i dati provengono da un campione casuale e i fissi quando si esamina l'intera popolazione. Tuttavia in questo secondo caso si potrebbe ugualmente sostenere l'uso di effetti casuali dato che gli α_i potrebbero vedersi come provenienti della più ampia popolazione degli eventi che non si sono verificati. Per es., Feige e Swamy (1974) adottano effetti casuali pur stimando su dati relativi a tutti gli stati americani.

(36)D'altra parte il modello a effetti casuali non si giustifica se gli effetti sono fissi. E' abbastanza sorprendente che nessun tentativo sia stato fatto di valutare le conseguenze di considerare α_i come casuale quando la sua natura è fissa. Anche negli esperimenti Montecarlo citati nel par. 5.1 i dati erano generati da un modello a effetti casuali.

(37)E' questa senza dubbio una formulazione strana, a meno che i regressori non siano costanti nel tempo, dato che implica che la relazione

$$y_{it} = x_{it}\,\beta + \alpha_i + u_{it} = x_{it}\beta + \Sigma_s\, x_{is}\frac{\pi}{T} + w_{i.} + u_{it}$$

dipende dalla dimensione del campione "which should surely be irrelevant to the economic agent in determining his behavior"(Pudney (1983, p. 45, nota 4)). Tuttavia il punto essenziale è di assumere l'esistenza di una correlazione di qualche tipo tra α_i e la media delle variabili esogene. Per una formulazione alternativa cfr. Chamberlain (1980 e 1982) .

(38)Questo vale naturalmente perchè il DVE si libera degli effetti prendendo differenze dalla media individuale. Lo stesso risultato si otterrebbe con differenze prime (o di altro ordine) anche se, in tal caso, minore efficienza sarebbe ottenuta a causa della risultante autocorrelazione dei residui.

(39)Da qui la tesi che: "There is only one estimator. The whole literature which has been based on an imaginary difference between the two estimators, starting from Balestra and Nerlove, is based on an incorrect specification which ignores the correlation between the effects and the explanatory variables." (p. 70). Che questo sia un caso speciale è stato provato da Lee

(1978).

(40)Si mantiene la precedente notazione che differisce da quella adottata "somewhat unconventionally" da Hausman e Taylor.

(41)Hausman (1978) propone tre diverse versioni del test, tutte basate sulle differenze tra stimatori between, within e GLS. Tuttavia Hausman e Taylor (1981, pp. 1382-1383) dimostrano che, se i componenti della varianza fossero noti, la statistica ottenuta sarebbe la stessa per i tre test. Una applicazione del test riportato nel testo si trova in Nickell (1982).

(42)Essendo basato sulla consistenza asintotica dei due stimatori il test è valido solo per regressioni statiche con N sufficientemente grande. Se l'equazione è dinamica lo stimatore between del coefficiente sulla sfasata è senza dubbio non consistente a causa dell'omissione dell'effetto individuale, ma il test potrebbe ancora essere usato per accertare l'esistenza di correlazione tra effetti esogeni purchè T sia lungo. Un test che può essere usato quando T è piccolo e l'equazione è dinamica, è descritto in Pudney (1983, pp. 22-24), insieme ad un test alternativo per il caso statico.

(43)Cfr. Hausman e Taylor (1981, p. 1382).

(44)L'uso di strumenti era già stato raccomandato da Griliches (1977). La scelta degli strumenti, interni al modello piuttosto che esterni come usualmente accade, è giustificabile in quanto derivante dalla minimizzazione di una somma di quadrati di residui (Chowdhury (1983)).

(45)Per derivare uno stimatore efficiente Hausman e Taylor suggeriscono di considerare il sistema completo formato dalla (34) e dalle equazioni esprimenti la dipendenza tra effetti e regressori. Questo può essere trasformato in un sistema formato dalla (34) e dalle forme ridotte in cui le variabili che dipendono da α_i sono espresse in funzione delle esogene. Se stime consistenti dei componenti della varianza della (34) sono state ottenute con la procedura indicata nel testo, la (34) moltiplicata per H^{-1} (vedi par. 4.4) può essere stimata efficientemente con minimi quadrati a due stadi. Se il modello è perfettamente identificato questa procedura produce le stime within (come in Mundlak (1978)); se il modello è sovraindentificato si ottengono stime più efficienti.

(46)L'eccezione è il MLE, ma la consistenza dipende, come visto, dalla scelta appropriata delle condizioni iniziali. Si ricorda inoltre che l'uso di GLS sembra condurre a stime con un bias inferiore, come indicato dagli esperimenti Montecarlo di Nerlove (1971-a) (cfr. par. 4.3 punto c)).

(47)E' necessario prendere differenze prime nell'ipotesi di effetti casuali anche per eliminare la correlazione tra effetti individuali e lo strumento suggerito. Dato che questa differenziazione elimina tutti i regressori invarianti al tempo, i coefficienti su questi non possono essere stimati in questa prima fase. La procedura richiesta in questo caso è descritta da Anderson e Hsiao (1982, p. 59).

(48)Si noti però che Pudney (1983) suggerisce di stimare inizialmente i componenti della varianza dai residui di regressori between e within che utilizzino variabili strumentali. Queste dovrebbero essere usate per ottenere stime da combinare in modo ottimale come indicato in c).

(49)Il modello di Pakes e Griliches è a coefficienti angolari variabili stocasticamente tra individui, come il modello di Swamy (par. 10); le conclusioni raggiunte, con adeguate semplificazioni, rimangono però valide anche nel caso di coefficienti uguali tra individui e di intercetta variabile. Sul problema dell'uso di dati in panel in modelli con ritardi distribuiti si veda anche Glejser (1978).

(50)Si noti che lo stimatore GLS proposto evita la perdita di N osservazioni che si manifesterebbe applicando l'abituale trasformazione con differenze prime corrette per ϱ.
Una diversa procedura per ottenere ϱ è suggerita da Nickell (1982) e si basa sul fatto che l'errore del modello (38) rispetta la relazione

$$u_{it} - u_{i.} = \varrho (u_{i,t-1} - u_{i.,-1}) + (\varepsilon_{it} - \varepsilon_{i.})$$

e una stima consistente di $u_{it} - u_{i.}$ è costituita dai residui del DVE. Una stima con OLS della precedente relazione dinamica fornisce un valore distorto di ϱ , ma Nickell (1980 e 1981) dimostra che il bias è pari a :

$$\underset{N\to\infty}{\text{plim}}\ \hat{\varrho} - \varrho = \left\{ \frac{2\varrho}{1-\varrho^2} - \left[\frac{1+\varrho}{T-1} \left(1 - (1/T) \frac{1-\varrho^T}{1-\varrho} \right) \right]^{-1} \right\}^{-1}$$

per cui sostituendo a ϱ il valore stimato si può risolvere per iterazione, ottenendo una stima corretta di ϱ . Una procedura leggermente diversa è descritta da Nickell (1980) e commentata criticamente da Bhargava et al. (1982).

(51)L'assunzione che lo stesso set di regressori sia presente in ogni equazione non è necessaria, ma verrà mantenuta dato che sembra comune per le regressioni con i dati in panel.

(52)Un test LM sull'ipotesi che Ω sia diagonale si trova in Breusch e Pagan (1980).

(53)Si veda anche Harvey (1981, p. 70) che fornisce anche una diversa esposizione del modello, basata su un riordino dei dati e che evita l'uso di prodotti di Kroneker.

(54)La (43) deve essere usata per controllare l'ipotesi di uguaglianza di tutti i coefficienti; naturalmente l'eguaglianza di qualcuno soltanto dei coefficienti può essere verificata con lo stesso metodo modificando la forma della equazione vincolata.

(55)Una assunzione più complessa sulla struttura di autocorrelazione dei residui si trova in Guilkey e Schmidt (1973) che considerano v_{it} dipendente da $v_{j,t-1}$ per tutti i j. Per una discussione di questo modello si veda Judge et al. (1980, pp. 268-274).

(56)Parks (1967) segue un approccio leggermente diverso, che però presenta lo stesso tipo di trasformazione.

(57)Si deve però notare che gli esperimenti di Kmenta e Gilbert sono basati su due individui soltanto sicché si perdono solo 2 osservazioni per variabile.

(58)A questo stadio si possono applicare i test abituali sull'esistenza di autocorrelazione dei residui (cfr. Parks (1969)). Judge et al. (1980) suggeriscono l'uso di un test sull'ipotesi di autocorrelazione congiunta degli N <u>sets</u> di residui.

(59)Il modello SURE è stato esteso in varie direzioni, talvolta con scarsa rilevanza pratica per l'applicazione a dati in <u>panel</u>. Per una rassegna di questi sviluppi cfr. Judge et al. (1980) e Srivastava e Dwivedi (1979).

(60)Modelli con coefficienti variabili nel tempo e per individui sono stati studiati da Hsiao (1974 e 1975), Swamy e Mehta (1975 e 1977) e Pudney (1983), ma hanno trovato scarso successo empirico. Per una rassegna si veda Judge et al. (1980, pp. 353-358).

(61)Quando T è piccolo può risultare necessario abbandonare l'assunzione di eteroschedasticità, dato che le stime di σ^2_i diventano poco affidabili. Un modello con omoschedasticità è considerato da Pudney (1983).

(62)La (50) implica che quando i regressori sono tutti uguali, cioè $X_1=\ldots\ldots=X_N$ e non c'è eteroschedasticità lo stimatore di Swamy si semplifica in una media semplice dei $\underline{b}_i$ (anche se la varianza di $\hat{\beta}_i$ risulterebbe diversa). Lo stesso avviene approssimativamente quando Δ è molto grande rispetto a $\sigma_i(X'_iX_i)^{-1}$. Si veda in proposito Maddala (1977, p. 402).

(63)Johnson e Lyon non indicano i valori di N e T utilizzati nell'esperimento.

BIBLIOGRAFIA

AA. VV. (1978), "The econometrics of panel data", Annales de l'Insee, April-September.

Amemiya T. (1971), "The estimation of variances in a variance components model", International Economic Review, February.

Anderson T. W. and C. Hsiao (1981), "Estimation of dynamic models with error components", Journal of the American Statistical Association, September.

Anderson T. W. and C. Hsiao (1982), "Formulation and estimation of dynamic models using panel data", Journal of Econometrics, January.

Balestra P. and M. Nerlove (1966), "Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas", Econometrica, July.

Berzeg K. (1979), "The error components model. Conditions for the existence of the maximum likelihood estimates", Journal of Econometrics, April.

Bhargava et al. (1980) Durbin-Watson-Sargan type test for serial correlation in models estimated from panel data, ICERD Working Paper.

Bhargava A. et al. (1982), "Serial correlation and the fixed effect model", Review of economic studies, October.

Bhargava A. and J. D. Sargan (1983), "Estimating dynamic random effects models from panel data covering short time series", Econometrica, 51, November.

Breusch T. S. and A. R. Pagan (1980), "The Lagrange multiplier test and its applications to model specification in econometrics", Review of Economic Studies, January.

Chamberlain G. (1980), "Analysis of covariance with qualitative data", Review of Economic Studies, January.

Chamberlain G. (1982), "Panel data", in M. Intrillingator e Z. Griliches, The handbook of econometrics, North Holland

Chetty V. K. (1968), "Pooling of time series and cross section data", Econometrica, April.

Chowdhury G. (1983), "Estimators for correctly specified panel data models", Center for Labour Economics, LSE, Discussion Paper No.

147.

Chowdury G. (1984) "An analysis of the determination of individual earnings in the U.S. 1967-1977 using panel data", LSE, Unpublished Ph. D. dissertation.

Dhrymes P. J. (1971), "Equivalence of the iterative Aitken and maximum likelihood estimators for a system of regression equations", Australian Economic Papers, June.

Dwivedi T. D. and Srivastava (1978), "Optimality of least squares in the seemingly unrelated regression equation model", Journal of Econometrics, June.

Feige L. (1964), "The demand for liquid assets: a temporal cross section analysis", Englewood Cliffs, N. J., Prentice Hall.

Feige L. and P. A. V. B. Swamy (1974), "A random coefficient model of the demand for liquid assets", Journal of money credit and banking, May.

Fisher R. A. (1918), "The correlation between relatives and the supposition of Mendelian inheritance", Trans. Royal Society, 52, pp. 399-433.

Fuller W. A. and G. E. Battesee (1973), "Transformations for estimation of linear models with nested error structure", Journal of the American Statistical Association, September.

Fuller W. A. and G. E. Battesee (1974), "Estimation of linear models with crossed error structure", Journal of Econometrics, May.

Glejser H. (1978), "Truncated distributed lags in small sample panel models", Annales de l'Inseé, April-September.

Griffiths W. E. and J. R. Anderson (1982), "Using time series and cross section data to estimate a production function with positive and negative marginal risks", Journal of American Statistical Association, September.

Griliches Z. (1977), "Estimating the returns to schooling: some econometric problems", Econometrica, January.

Griliches Z. (1979), "Sibling models and data in economics: beginnings of a survey", Journal of Political Economy, October.

Griliches Z. e Hausman J. (1984), "Errors in variables in panel data", National Bureau of Economic Research, Technical Working Paper No.37.

Guilkey D. K. and P. Schmidt (1973), "Estimation of seemingly unrelated regressions with vector autoregressive errors", Journal of the American Statistical Association, September.

Hall B. H. (1978), "A general framework for the time series-cross section estimation", Annales de l'Inseé, April-September.

Harvey A. C. (1981), "The econometric analysis of time series", Phillip Allan, Oxford.

Harville D. A. (1977), "Maximum likelihood approaches to variance components estimation and to related problems", Journal of the American Statistical Association, June.

Hausman J. A. (1978), "Specification tests in econometrics", Econometrica, November.

Hausman J. A. and W. E. Taylor (1981), "Panel data and unobservable individual effects", Econometrica, November.

Heckman J. J. (1979), "The incidental parameter and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Montecarlo evidence", in McFadden D. and C. Manski (eds.), "Structural analysis of discrete data", MIT Press, Cambridge, Mass.

Henderson C. R. (1953), "Estimation of variance and covariance components", Biometrics, June.

Hester and Pierce (1975), "Bank management and portfolio behaviour", Yale University Press, New Haven.

Hildreth C. (1950), "Combining cross section data and time series", unpublished Cowles Commission Discussion Paper, Statistics N° 347, May.

Hildreth C. and J. P. Houck (1968), "Some estimators for a linear model with random coefficients", Journal of the American Statistical Association , September.

Hock I. (1962), "Estimation of production function parameters combining time series and cross section data", Econometrica, February.

Houthaker H. S. et al. (1974), "Dynamic demand analysis for gasoline and residential electricity", American Journal of Agricoltural Economics, May.

Hsiao C. (1974), "Statistical inference for a model with both random cross sectional and time effects", International Economic Review, February.

Hsiao C. (1975), "Some estimation methods for a random coefficient model", Econometrica, March.

Hurwicz L. (1950), "Least squares bias in time series", in Koopmans T. C. (ed.) "Statistical inference in dynamic economic models", New York, Wiley.

Hussain A. (1969), "A mixed model for regressions", Biometrika,

August.
Johnson P. R. (1960), "Land substitutes and changes in corn yields", Journal of Farm Economics, May.
Johnson P. R. (1964), "Some aspects of estimating statistical cost functions", Journal of Farm Economics, February.
Jöreskog K. G. (1978), "An econometric model for multivariate panel data", Annales de l'Inseé, April-September.
Jorgenson et al. (1982), "The trascendental logarithmic model of aggregate consumer behaviour", in Basmann R. e Rhodes G. (eds.) Advances in econometrics, Vol. 1, JAI Press.
Judge G.G. et al.(1980), "The theory and practice of econometrics", John Wiley and Sons, New York.
Keifer N. M. (1979), "Population heterogeneity and inference from panel data on the effects of vocational education", Journal of Political Economy, October.
Keifer N. M. (1980), "Estimation of fixed effects models for time series of cross sections with arbitrary intertemporal covariance", Journal of Econometrics, October.
Kmenta J. (1971), "Elements of Econometrics", New York, MacMillan.
Kmenta J. and R. F. Gilbert (1968), "Small sample properties of alternative estimators of seemingly unrelated regressions", Journal of the American Statistical Association, December.
Kmenta J. and R. F. Gilbert (1970), "Estimation of seemingly unrelated regressions with autoregressive disturbances", Journal of the American Statistical Association, March.
Kuh E.(1959), "The validity of cross-sectionally estimated behaviour equations in time series applications", Econometrica, April.
Kuh E. (1963), "Capital stock growth: a micro-econometric approach", North-Holland Publishing Company, Amsterdam.
Lee L. F. (1978), "Estimation of error components model with ARMA(p,q) time component. An exact GLS approach", University of Minnesota Center for Economic Research Discussion Paper N° 78, Minneapolis.
Lee T.H.(1966), "Substituibility of non-bank intermediary liabilities for money: the empirical evidence", Journal of Finance, September.
Lillard L. and R. Willis (1978), "Dynamic aspects of earning mobility", Econometrica, September.
Lillard L. A. and Y. Weiss (1979), "Components of variation in panel earnings data: American scientists 1960-1970", Econometrica,

March.

Maddala G. S. (1971-a), "The use of variance components models in pooling cross section and time series data", Econometrica, March.

Maddala G. S. (1971-b), "The likelihood approach to pooling cross-section and time series data", Econometrica, November.

Maddala G. S. (1977), "Econometrics", McGrow-Hill, New York.

Maddala G. S. (1978), "Selectivity problems in longitudinal data", Annales de l'Inseé, April-September.

Maddala G. S. and Mount T. D. (1973), "A comparative study of alternative estimators for variance components models used in econometric applications", Journal of the American Statistical Association, June.

Mairesse J. (1984), "Labor and investment demand at the firm level in French, German and U. S. manufacturing 1970-79", presentato all'International Seminar on Macroeconomics tenuto a Perugia il 24-26 giugno 1984.

Marcus A.L.(1983), "The bank capital decision: a time series cross section analysis", Journal of Finance, September.

Marotta G. (1983), "Microfunzioni di investimento: uno studio su un campione longitudinale di imprese italiane 1960-1973", in N. Rossi-R. Rovelli (eds.), "Ricerche di Economia applicata : il caso italiano", Franco Angeli.

Mundlak Y. (1961), "Empirical production function free of management bias", Journal of Farm Economics, February.

Mundlak Y. (1963), "Estimation of production and behavioural functions from a combination of cross-section and time series data", in C. F. Christ (ed.), "Measurement in economics".

Mundlak Y. (1978-a), "On the pooling of time series and cross section data", Econometrica, January.

Mundlak Y. (1978-b), "Models with variable coefficients-integration and extension", Annales de l'Inseé, April-September.

Nerlove M. (1965), "Estimation and identification of Cobb-Douglas production functions", North-Holland, Amsterdam.

Nerlove M. (1967), "Experimental evidence on the estimation of dynamic economic relations from a time series of cross sections", Economic Studies Quarterly, December.

Nerlove M. (1971-a), "Further evidence on the estimation of dynamic economic relations from a time series of cross sections", Econometrica, March.

Nerlove M.(1971-b),"A note on error components models", Econometrica, March.

Nickell S. (1980), "Correcting the biases in dynamic models with fixed effects", LSE, mimeo.

Nickell S. (1981), "Biases in dynamic models with fixed effects", Econometrica, November.

Nickell S. (1982), "The determinants of occupational success in Britain", Review of Economic Studies, January.

Pakes A. e Griliches Z. (1984), "Estimating distributed lags in short panels with an application to the specification of depreciation patterns and capital stock constructs", Review of Economic Studies, April.

Parks R. W. (1967), "Efficient estimation of a system of regression equations when the disturbances are both serially and contemporaneously correlated", Journal of the American Statistical Association, June.

Palda K. S. and L. M. Blair (1970), "A moving cross section analysis of demand for toothpast", Journal of Marketing Research, 7.

Pudney S. E. (1983), "The estimation and testing of some error components models", mimeo L.S.E.

Scheffé H. (1959), "The analysis of variance", New York, John Wiley and Sons.

Schipper L. (1964), "Consumer discretioning behaviour", North-Holland, Amsterdam.

Searle S. R. (1971), "Linear models", New York, John Wiley and Sons

Srivastava V. K. and T. D. Dwivedi (1979), "Estimation of seemingly unrelated regression equations. A brief survey", Journal of Econometrics, April.

Swamy P. A. V. B. (1971), "Efficient inference in a random coefficient regression model", Econometrica, March.

Swamy P. A. V. B. (1971), "Statistical inference in random coefficient regression models", Springer-Verlag, New York.

Swamy P. A. V. B. (1974), "Linear models with random coefficients", in P. Zarembka (ed.), Frontiers in econometrics, New York, Academic.

Swamy P. A. V. B. and S. S. Arora (1972), "The exact finite sample properties of the estimators of coefficients in the error components regression models", Econometrica, March.

Swamy P. A. V. B. and J. S. Mehta (1973), "Bayesian analysis of error components regression models", Journal of the American

Statistical Association, September.

Taylor W. E. (1980), "Small sample considerations in estimation from panel data", Journal of Econometrics, June.

Theil H. (1954), "Linear aggregation of economic relations", North Holland, Amsterdam.

Theil H. (1971), "Principles of Econometrics", John Wiley, New York.

Wallace T. D. and A. Hussain (1969), "The use of error components models in combining cross section with time series data", Econometrica, January.

Zellner A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias", Journal of the American Statistical Association, June.

Zellner A. (1969), "On the aggregation problem: a new approach to a troublesome problem", in K. A. Fox et al. (ed.), "Economic models, estimation and risk programming: essays in honor of Gerhard Tintner", New York, Springer-Verlag.

TEMI DI DISCUSSIONE RECENTEMENTE PUBBLICATI (*)

n. 29 - Real balances, the exchange rate, and indexation: real variables in disinflation, by S. Fischer (giugno 1984)

n. 30 - Il bilancio pubblico per il quinquennio 1984-88: alcune simulazioni, di G. Morcaldo - G. Salvemini (luglio 1984)

n. 31 - Funzioni aggregate d'investimento, di M. Magnani - R. Valcamonici (agosto 1984)

n. 32 - Un'indagine econometrica sui consumi nazionali (1972-1981), di G. Marotta (agosto 1984)

n. 33 - Short-term interest rate linkages between the United States and Europe, by S. Micossi - T. Padoa-Schioppa (agosto 1984)

n. 34 - La condizione di additività nella stima di sistemi di equazioni simultanee, di C.A. Bollino (agosto 1984)

n. 35 - La relazione tra orari di fatto e ore contrattuali nell'industria italiana, di G. Bodo - C. Giannini (settembre 1984)

n. 36 - Corsi e rendimenti dei titoli a medio e lungo termine, di G. Galli (settembre 1984)

n. 37 - Il commercio di manufatti: una specializzazione incompleta, di G. Majnoni (settembre 1984)

n. 38 - Il dibattito sull'inflazione italiana negli ultimi 15 anni, di L. Guiso (settembre 1984)

n. 39 - Estimation of complete demand systems: the trinomial expenditure system in comparison with alternative demand systems, by C.A. Bollino (ottobre 1984)

n. 40 - Un modello di previsione del bilancio pubblico per il breve-medio termine, di G. Morcaldo - G. Salvemini - P. Zanchi (ottobre 1984)

n. 41 - Il mercato degli impieghi bancari in Italia: un'analisi econometrica (1974-1982), di I. Angeloni (ottobre 1984)

n. 42 - Why floating exchange rates fail, by R. McKinnon (novembre 1984)

n. 43 - Una stima delle funzioni di domanda di attività finanziarie, di F. Cotula - G. Galli - E. Lecaldano - V. Sannucci - E. Zautzik (novembre 1984)

(*) I "Temi" pubblicati possono essere richiesti alla Biblioteca del Servizio Studi della Banca d'Italia.