PANEL DATA ECONOMETRICS: THEORY AND APPLICATIONS IN STATA

Maria Elena Bontempi e.bontempi@economia.unife.it

Roberto Golinelli golinell@spbo.unibo.it

This version: 30 agosto 2006[§]

1. Overview

The aim of this note is to provide some basic theoretical background on panel econometrics. We also add references, indicated by ☺A, to the corresponding applications in STATA.

1.1. The econometric methodology: a review

L'attività di modellazione econometrica (cioè di costruzione di modelli econometrici) si articola nelle tre fasi della specificazione, stima e test.

La **specificazione** muove dalla teoria economica, che suggerisce l'elenco delle variabili di interesse del problema che si intende affrontare e la direzione di causalità (ad esempio, lo studio della

[§] Very preliminary. Comments welcome. Thanks to Luigi Bidoia (Prometeia, Bologna), Lucio Picci (Department of Economics, University of Bologna), Jacques Mairesse (CREST-ENSAE, Paris) for providing stimulus and suggestions. Manuel Arellano, Badi Baltagi, Russel Davidson, Cheng Hsiao, and Jeff Wooldridge for offering the opportunity to learn more and more during the CIDE Summer Schools (Bertinoro, Forlì). Any error is our fault (sorry for that).

relazione fra tasso di interesse e tasso di inflazione). Tuttavia, la sola teoria non basta per definire compiutamente tutti gli elementi di cui si compone un modello econometrico stimabile; pertanto, sono necessarie ulteriori ipotesi di specificazione, quali la scelta della forma funzionale ed eventuali trasformazioni delle variabili. Spesso la forma funzionale ipotizzata è quella lineare (o log-lineare).
La relazione è, inoltre, stocastica per la presenza di un termine di errore (in genere additivo), ossia di una variabile casuale che serve a cogliere qualsiasi effetto omesso, spesso non osservabile, che rende non esatta la relazione ipotizzata dalla teoria.
Il modello classico di regressione lineare si fonda sulle seguenti 5 ipotesi di specificazione:

1. Relazione lineare nei parametri, supposta vera;
2. Le variabili esplicative sono deterministiche (non stocastiche, fisse) oppure esogene, ossia non correlate con il termine di errore (nelle sezioni successive vedremo varie nozioni di esogeneità);
3. Gli errori hanno media nulla;
4. Gli errori hanno varianza costante;
5. Gli errori non sono tra loro correlati.

Le ipotesi 3-5 sono sintetizzate da: $\varepsilon \sim IID(0, \sigma_\varepsilon^2)$

La **stima** dei parametri del modello ha lo scopo di assegnare specifici valori ai parametri (sconosciuti) del problema di interesse. La disponibilità di stime permette di quantificare la relazione di causalità fra le variabili esplicative (nel caso più semplice una sola x) e la variabile dipendente (definita con y). Un metodo largamente utilizzato per la stima del modello parametrico è quello dei minimi quadrati ordinari (OLS), che attribuisce ai parametri della relazione quei valori che minimizzano il quadrato delle distanze fra le osservazioni disponibili e la corrispondente retta di regressione; tali distanze sono anche dette residui. Dall'imposizione delle condizioni (necessarie) per il minimo si ottiene il sistema delle equazioni normali, dalla cui soluzione si ottengono i valori stimati.

Se le cinque ipotesi di specificazione del modello classico di regressione sono vere, si dimostra (Teorema di Gauss-Markov) che lo stimatore OLS è BLUE (Best Linear Unbiased Estimator), vale a dire il miglior stimatore lineare corretto (non distorto); il migliore perché è quello che ha varianza minima (proprietà dell'efficienza) fra tutti gli stimatori lineari e corretti. E' possibile dimostrare che la proprietà della correttezza si basa sulle ipotesi 1-3 e quella dell'efficienza sulle ipotesi 4-5.
Se non vale l'ipotesi 2 e gli errori sono correlati con le esplicative, gli OLS sono distorti e inconsistenti; il metodo di stima consistente è quello delle variabili strumentali (IV – 2SLS oppure GMM).
Il venire meno delle ipotesi 4-5 comporta la perdita della proprietà dell'efficienza degli OLS; il problema può essere superato in fase di stima mediante l'impiego del metodo dei minimi quadrati generalizzati (GLS), che sono BLUE pur in presenza di eteroschedasticità e/o autocorrelazione degli errori (Teorema di Aitken).

La fase dei **test** risponde a due blocchi di domande fondamentali.

Il primo ha a che vedere con le scelte di specificazione del modello: come faccio a sapere se ho effettuato scelte corrette (nel senso di scelte di specificazione coerenti con i dati del problema che sto affrontando)? La risposta a questo quesito si ottiene impiegando test di scorretta specificazione (ad esempio di autocorrelazione e di eteroschedasticità), che si concentrano sull'analisi dei residui della regressione.

Il secondo blocco di domande si concentra invece sui valori assegnati ai parametri. La stima puntuale può essere vista come un'estrazione da un'urna di un particolare valore; in tal caso, quali sono la forma e i parametri caratteristici (ad esempio i momenti primo e secondo) della

distribuzione dello stimatore (urna)? A questi quesiti si risponde mediante l'utilizzo dei test di significatività basati sulle stime dei parametri del modello.
Per essere concretamente effettuata, la fase di test del modello deve essere fondata statisticamente e, quindi, richiede l'individuazione di un insieme di ipotesi distributive sulla variabile casuale errore del modello.

Se alle precedenti 5 ipotesi di specificazione si aggiunge anche la: 6. Normalità della distribuzione degli errori;
è possibile calcolare specifici intervalli di stima per i parametri del modello ed effettuare test di verifica di ipotesi sui parametri, utilizzando i valori critici delle distribuzioni t e F.

Le ipotesi 3-6 sono sintetizzate da: $\varepsilon \sim IIN(0, \sigma_\varepsilon^2)$

1.2. Dati cross-section e serie storiche

L'insieme dei dati osservabili, include le realizzazioni delle variabili del modello (y e x), o di proxy (nel caso in cui le variabili teoriche di interesse non siano osservabili).
Le osservazioni campionarie possono essere classificate in:

(i) dati sezionali (**cross-section**); le osservazioni si riferiscono a diversi individui (paesi, famiglie, imprese, ecc.) rilevate nello stesso periodo (ad esempio, nel 1990);

(ii) **serie storiche**; le osservazioni sono relative allo stesso individuo, o allo stesso aggregato, misurate in diversi periodi (ad esempio: l'inflazione in Italia dal 1980 al 1998);

(iii) pooling di dati sezionali e serie storiche; le osservazioni sono bidimensionali, in quanto variano sia per individuo (o aggregato) sia nel periodo temporale di rilevazione. I **panel data** appartengono a questa categoria di dati.

A seconda del tipo di osservazioni campionarie di cui si dispone, la fase di stima incontra problematiche in parte diverse. Con riferimento ai casi (i) e (ii), possono essere sottolineati i seguenti tratti distintivi.

L'utilizzo di dati cross-section per la stima è spesso accompagnato da problemi di eteroschedasticità dei residui, a causa della mancata spiegazione della elevata varianza delle osservazioni. Queste difficoltà sono riconducibili al fatto che la variabilità nei dati tende a concentrarsi per gruppi (famiglie ricche e famiglie povere, imprese di piccole e imprese di medio-grandi dimensioni, paesi europei a bassa e alta inflazione).

In questi casi l'utilizzo del metodo di stima GLS può condurre a risultati non soddisfacenti, dato che l'eteroschedasticità dei residui può indicare problemi di scorretta specificazione del modello (soprattutto scorretta forma funzionale ed omissione di variabili rilevanti).

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 3)

La principale innovazione di metodo introdotta dalle serie storiche rispetto alle tecniche tipiche per dati cross-section è l'enfasi sull'importanza dell'ordine delle osservazioni (realizzazioni del processo stocastico). Le osservazioni di una serie storica sono rigorosamente ordinate secondo il tempo e, quindi, non è possibile supporle indipendenti le une dalle altre: fenomeni economici quali il trend, il ciclo e la stagionalità contraddicono tale eventualità.
Questa peculiarità delle serie storiche da un lato rende spesso inefficace l'impiego del modello statico:

$y_t = a + b\, x_t + \varepsilon_t$

In altri termini, se $\varepsilon_i \sim IID(0, \sigma_\varepsilon^2)$, si ha che $y_t \sim IID(a+b\ x_t\ ,\ \sigma_y^2)$ cioè le y_t sono "indipendentemente distribuite"; si tratta di un'ipotesi spesso irreale per serie economiche.
Sono, pertanto, preferibili modelli dinamici. Fra questi, il modello:

$$y_t = a + b_1 x_t + b_2 x_{t-1} + b_3 y_{t-1} + \varepsilon_t$$

è largamente utilizzato in letteratura. Si tratta del modello autoregressivo (perché include y_{t-1} fra le esplicative) e a ritardi distribuiti (perché presenta sia x_t sia x_{t-1}); è un ARDL(1,1) con dinamica del primo ordine, perché i ritardi sono sino a t-1, sia in x, sia in y.
Una scorretta specificazione dinamica spesso si tramuta in autocorrelazione dei residui. Come per l'eteroschedasticità con dati cross-section, anche per residui autocorrelati si potrebbe utilizzare il metodo di stima GLS, ipotizzando che gli errori del modello seguano un processo autoregressivo.
D'altro canto, però, molti studi hanno sottolineato la scorrettezza di tale procedura che, se seguita acriticamente, il più delle volte tende a nascondere significativi problemi di specificazione dinamica.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 4)

Parte della letteratura ritiene che i dati cross-section riflettano comportamenti di lungo periodo (tutti gli individui del campione sono affetti nello stesso modo dal quadro macroeconomico), mentre le serie storiche enfatizzino gli effetti di breve periodo (studi seminali di Tobin negli anni '50 e successivi approfondimenti di Maddala di fine'60).

Anche se gli studi sui legami di lungo periodo nell'ambito di modelli con meccanismo a correzione dell'errore per serie storiche hanno molto attenuato la portata di questa dicotomia, è chiaro che la combinazione, nello stesso ambito, dell'informazione cross-section e time series ed una crescente attenzione alla specificazione dinamica permettono di formulare e stimare migliori modelli interpretativi degli eventi economici (Cfr Baltagi e Griffin, 1984).

1.3. Dati che combinano serie storiche e cross-section

Supponiamo di disporre di osservazioni temporali t, relative ad un certo numero di unità statistiche di base o più semplicemente 'individui' i (imprese, famiglie, paesi, regioni, settori, titoli ...), per le variabili economiche di interesse y e x. Il panel di dati è formato da osservazioni indicate con y_{it} e x_{it} dove:
i = 1, 2, ..., N (individui) e t = 1, 2, ..., T (periodo storico).
La disponibilità di informazioni di questo tipo permette di specificare modelli più flessibili.

Un modello è sempre una semplificazione della realtà (tende a presentare poche variabili esplicative):
$\mathbf{a + b\ x_{it}}$ dove a e b sono parametri.
La scelta di adottare un'unica pendenza b per tutti gli individui è giustificata nei casi in cui si disponga di poche osservazioni temporali per fare inferenza sui singoli individui. Se i comportamenti degli individui si assomigliano, c'è un vantaggio nel fare pooling (almeno parziale) perché alla scarsa conoscenza del comportamento di un individuo si può sopperire con i dati per gli altri individui (che si comportano in un modo analogo). Il rischio da evitare è quello di imporre a priori arbitrarie regolarità; per questo nella sezione [A] approfondiremo i legami fra modelli alternativi e le restrizioni che ne stanno alla base. → **to pool or not to pool?**

I dati $\mathbf{y_{it}}$ che vogliamo spiegare sono però influenzati da molti fattori. Da questo discende il fondamentale problema per l'economista quantitativo: trovare una lista adeguata di variabili esplicative. Se non ci riesce, incappa nel problema di omissione di fattori esplicativi rilevanti.

Nessun modello può dirsi al riparo dal rischio di avere omesso fattori esplicativi di diverso tipo: (i) che variano per individuo, ma sono costanti nel tempo (ad esempio, il genere, le caratteristiche individuali, il background socio-economico); (ii) che variano solo nel tempo (ad esempio, il ciclo economico che è lo stesso per tutti gli individui); (iii) che variano in entrambe le direzioni (ad esempio, i salari individuali, lo stock di capitale).

Supponiamo che sia possibile evidenziare due fattori di errore (fonti di ignoranza): $\boldsymbol{\mu_i}$ riconducibile alla tipologia (i) di omissioni; $\boldsymbol{\varepsilon_{it}}$ riconducibile alle tipologie (ii) e (iii). L'effetto dei fattori esplicativi omessi viene rappresentato con:

$$\mathbf{v_{it} = \mu_i + \varepsilon_{it}} \qquad [\text{ipotesi: } Cov(\mu_i , \varepsilon_{it}) = 0]$$

Mettendo assieme tutti i pezzi, si ha il modello:

$$\underset{\text{dato}\uparrow}{\mathbf{y_{it}}} \quad = \underset{\uparrow\text{parte spiegata}}{\mathbf{a + b\, x_{it}}} \quad + \quad \underset{\uparrow\text{ignoranza}}{\mathbf{\mu_i + \varepsilon_{it}}}$$

Un'ipotesi fondamentale dell'approccio uniequazionale è quella di esogenità delle variabili esplicative incluse: $Cov[x_{it} , (\mu_i +\varepsilon_{it})] = 0$.
Il modello panel presenta un'articolazione ed una serie di alternative che permettono di sfruttare meglio la variabilità individuale dei dati per tenere conto di effetti omessi: → **fixed or random effects?**

Problema: specificare un modello che dia conto di possibili diversità di comportamento fra individui e nel tempo in modo da poter utilizzare assieme (pool) tutte le osservazioni disponibili per misurare la causalità (teoria economica) da x a y.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 5)

Quindi, l'obiettivo principale del corso sarà quello di chiarire sia le diverse ipotesi di specificazione (Cfr sezioni [A]-[B]) che stanno alla base delle alternative modellazioni di y_{it}, sia le procedure di stima-inferenza statistica (Cfr sezione [C]) che permettono di verificare l'accettabilità di quelle modellazioni.

In generale:

(1) $$y_{it} = a_{it} + b_{it}\, x_{it} + \varepsilon_{it}$$

E' un modello che dà conto di diverse forme di eterogeneità: i parametri a_{it} (intercetta) b_{it} (pendenza) e l'errore ε_{it} possono variare fra individui e nel tempo.
L'eterogeneità dei comportamenti individuali, se non è colta dalle variabili esplicative del modello (causalità economica), finisce con l'essere spiegata da coefficienti che variano per individuo e nel tempo (a_{it} e b_{it}), oppure dai termini di errore (ε_{it}).

Ipotesi:
- il 'vero' legame fra y e x è statico e lineare;
- x è esogena;

Il modello (1) ha solo uno scopo descrittivo (no stima, inferenza e previsione). Affinché acquisisca capacità esplicativa e sia operativo è necessario strutturarlo con ulteriori ipotesi riguardo:

[A] il grado di variabiltà dei coefficienti;
[B] le proprietà statistiche delle componenti non osservabili;
[C] le relazioni fra variabili esplicative e termini di errore (stima ottimale e test di specificazione).

[A] Modelli alternativi per dati panel

Imponendo vincoli su a_{it} e b_{it} si ottengono specificazioni di interesse per gli scopi della ricerca applicata.

Specificazione [1]: $\quad$ $\mathbf{a_{it} = a}$ $\quad$ $\mathbf{b_{it} = b}$

(1.1) $\quad y_{it} = a + b\, x_{it} + \varepsilon_{it}$

intercetta e pendenza sono costanti (specificazione pooled). Pertanto, è ε_{it} a catturare tutte le eventuali differenze fra individui e nel tempo. Possono, pertanto, derivare i tipici problemi:

$E(\varepsilon_{it}^2) = \sigma_i^2$	eteroschedasticità individuale
$E(\varepsilon_{it}\, \varepsilon_{it-k}) \neq 0$	auto-correlazione di ordine k
$E(\varepsilon_{it}\, \varepsilon_{jt}) \neq 0$	cross-correlazione ($i \neq j$)

Se così fosse, la specificazione (1.1) potrebbe essere scorretta (non è vera la poolability imposta) e lo stimatore OLS per a e b potrebbe essere distorto ed inconsistente.
→ il modello (1.1) è spesso eccessivamente vincolato. Seguendo una logica dal-generale-al-particolare (LSE) è preferibile iniziare da modelli meno vincolati e poi verificare l'ammissibilità delle restrizioni implicate dal modello (1.1).

Specificazione [2]: $\quad$ $\mathbf{a_{it} = a_i = a + \mu_i}$ $\quad$ $\mathbf{b_{it} = b}$

(1.2) $\quad y_{it} = a_i + b\, x_{it} + \varepsilon_{it}$

l'intercetta varia solo per individuo (one-way individual); la pendenza è costante.
Questione: μ_i è deterministico o stocastico?

Se deterministico	→	modello con effetti fissi (dummy variable model)
Se stocastico	→	modello con effetti casuali (error components model)

Specificazione [3]: $\quad$ $\mathbf{a_{it} = a + \mu_i + \tau_t}$ $\quad$ $\mathbf{b_{it} = b}$

(1.3) $\quad y_{it} = a_{it} + b\, x_{it} + \varepsilon_{it}$

l'intercetta varia per individuo e nel tempo (two-ways individual and temporal); la pendenza è costante.
Questione: μ_i e τ_t sono deterministici o stocastici?

Se deterministici	→	modello con effetti fissi (dummy variable model)
Se stocastici	→	modello con effetti casuali (error components model)

Specificazione [4]: $\mathbf{a_{it} = a_i = a + \mu_{a,i}}$ $\mathbf{b_{it} = b_i = b + \mu_{b,i}}$

(1.4) $y_{it} = a_i + b_i x_{it} + \varepsilon_{it}$

intercetta e pendenza variano entrambe per individuo.

Questione: $\mu_{a,i}$ e $\mu_{b,i}$ sono deterministici o stocastici?

Se deterministici $\rightarrow$ modello SUR (seemingly unrelated regression)

Se stocastici $\rightarrow$ modello con parametri stocastici

Specificazione [5]: $\mathbf{a_{it} = a_t = a + \tau_{at}}$ $\mathbf{b_{it} = b_t = b + \tau_{bt}}$

(1.5) $y_{it} = a_t + b_t x_{it} + \varepsilon_{it}$

intercetta e pendenza variano entrambe nel tempo.

$\rightarrow$ Time varying parameter models

Discussione

I 5 casi riportati sopra non esauriscono l'insieme dei modelli particolari che è possibile ricavare dal caso generale (1), ma sono comunque quelli più rilevanti per la letteratura empirica.

Le osservazioni individuali (imprese, famiglie) presentano tipicamente N grande e T piccolo ("longitudinal data"). In questi casi, le tecniche si concentrano sulla variabilità fra individui (cross-section). Perciò, quando in letteratura si parla di 'modelli panel', ci si riferisce a casi riconducibili alla specificazione [2]. Al più, si introduce la specificazione [3] con τ_t deterministici, allo scopo di 'depurare' i dati da effetti ciclici e/o trend che si suppone giochino lo stesso ruolo per tutti gli individui (però problemi di previsione). Nei modelli [2]-[3] l'omogeneità delle pendenze è più una necessità che non un consolidato ex ante.

Quando $N \approx T$ e T grande, come tipicamente avviene per dati in cui gli individui sono settori, regioni, paesi, ecc., si parla di "data fields". In questi casi, il modello [4] diviene più attraente, in quanto l'accresciuta dimensione temporale permette di stimare i parametri a_i e b_i individuo per individuo e di approfondire il tema dell'eterogeneità delle pendenze b_i . Inoltre, un rilevante periodo temporale permette di introdurre nel modello elementi di dinamica.

La specificazione dinamica ha dimostrato la sua rilevanza nel campo dei modelli per serie storiche. Parte di queste note estenderà al caso dei dati panel le problematiche della modellazione dinamica.

Le specificazioni [5] a parametri variabili nel tempo hanno ricevuto più attenzione a livello di serie storiche ("pooled data" o "several time series"), e perciò vanno oltre lo scopo del presente corso.

[B] Le ipotesi di specificazione delle componenti non osservabili

Per completare la specificazione dei modelli econometrici in [A] è necessario formulare ipotesi statistiche sulle loro componenti non osservabili, da aggiungere alle ipotesi di linearità ed esogenità.

Con ε_{it} si indica il 'tradizionale' termine di <u>errore stocastico</u> del modello classico di regressione lineare per il quale si ipotizza che: $E(\varepsilon_{it}) = 0$ (media zero); $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$ (varianza costante);

$E(\varepsilon_{it}\, \varepsilon_{js}) = 0$ (covarianza nulla quando $i \neq j$ e $t \neq s$). Sinteticamente:

$$\varepsilon_{it} \sim IID\ (0, \sigma_\varepsilon^2)\ ; \qquad \forall\ i \in N,\ t \in T$$

Se ε è sempre considerato stocastico, non è detto che debbano esserlo anche le altre componenti non osservabili (μ e τ) che, al pari di ε , rappresentano la nostra 'ignoranza', in quanto

approssimano, con spostamenti di intercetta [modelli (1.2) e (1.3)] e di intercetta-pendenza (1.4), l'effetto di fattori esplicativi non modellati.
Nei modelli(1.2), (1.3) e (1.4) μ e τ sono ipotizzati deterministici o stocastici. Tale scelta di specificazione ha rilevanti ripercussioni sulle proprietà statistiche degli stimatori dei loro parametri.

[B.1] Il modello con effetti fissi

Nel modello con effetti fissi (fixed effects model, FE) si ipotizza che l'intercetta sia deterministica e vari da individuo a individuo (Cfr. equazione 1.2'), oppure per individuo e nel tempo (Cfr. equazione 1.3').

(1.2') $y_{it} = a + b\, x_{it} + \sum_{j=1}^{N-1} \mu_j D_{ji} + \varepsilon_{it}$

(1.3') $y_{it} = a + b\, x_{it} + \sum_{j=1}^{N-1} \mu_j D_{ji} + \sum_{s=1}^{T-1} \mu_s D_{st} + \varepsilon_{it}$

Nel primo caso vengono utilizzate N-1 variabili dummy individuali D_{ji} , nel secondo (N-1)+(T-1) dummy D_{ji} e D_{st} tali che: $D_{ji}=1$ se i=j, $D_{ji}=0$ se $i\neq j$; $D_{st}=1$ se t=s, $D_{st}=0$ se $t\neq s$.

Gli N-1 parametri μ_i misurano gli scostamenti delle intercette di N-1 individui da quella dell'individuo preso come base.

Allo stesso modo, i T-1 parametri τ_t rappresentano gli scostamenti delle intercette di T-1 periodi dal periodo-base.

Nota: se si inseriscono nei modelli (oltre all'intercetta) N dummy individuali e/o T dummy temporali, qualsiasi stima è resa impossibile dalla perfetta collinearità con l'intercetta, in quanto $\sum_{j=1}^{N} D_{ji} = 1$ ($\sum_{s=1}^{T} D_{st} = 1$) .

[B.2] Il modello con effetti casuali

Una alternativa al modello FE è il modello con effetti casuali (random effects model, RE) in cui μ_i (e/o τ_t) vengono considerate stocastiche.

In particolare, si suppone che: $E(\mu_i) = 0$; $E(\mu_i^2) = \sigma_\mu^2$; $E(\mu_i\, \mu_j) = 0$ (quando $i\neq j$). Ipotesi analoghe (media nulla, omoschedasticità e incorrelazione) sono formulate per gli effetti temporali τ_t . In breve:

$$\mu_i \sim IID\,(0, \sigma_\mu^2)\ ; \qquad \tau_t \sim IID\,(0, \sigma_\tau^2)$$

Inoltre, nel modello con effetti casuali si suppone che i movimenti stocastici di μ e τ siano incorrelati con tutti gli altri regressori e con l'errore ε dell'equazione.

I modelli con effetti casuali possono essere riscritti nel seguente modo:

(1.2") $y_{it} = a + b\, x_{it} + v_{2it}$ $\qquad v_{2it} = \varepsilon_{it} + \mu_i$

(1.3") $y_{it} = a + b\, x_{it} + v_{3it}$ $\qquad v_{3it} = \varepsilon_{it} + \mu_i + \tau_t$

dove a e b sono parametri costanti (per individuo e nel tempo) e i movimenti stocastici dell'intercetta sono inglobati nei termini stocastici di errore composto v_{2it} (per il caso [2], modello one-way, two error components) e v_{3it} (per il caso [3], modello two-ways. three error components).

Per questa ragione il modello RE viene talvolta definito come 'error components model'.

[C] Stimatori ottimali e test di specificazione

[C.0] Notazione

Le medie individuali (per individuo) di x e y si definiscono:
$x_{i.} = (\Sigma_t\, x_{it}) / T$
$y_{i.} = (\Sigma_t\, y_{it}) / T$

Le medie complessive (totali):
$x_{..} = (\Sigma_i\, \Sigma_t\, x_{it}) / (N\,T) = (\Sigma_i\, x_{i.}) / N$
$y_{..} = (\Sigma_i\, \Sigma_t\, y_{it}) / (N\,T) = (\Sigma_i\, y_{i.}) / N$

La doppia dimensione (individuale e temporale) dei dati panel permette di scomporre la deviazione tra ciascuna osservazione (ad esempio x_{it}) e la media complessiva $x_{..}$ in due parti: la componente (solo temporale) **dentro** l'individuo "within" e quella (solo individuale) **fra** individui "between":

(2) $(x_{it} - x_{..}) \equiv (x_{it} - x_{i.}) + (x_{i.} - x_{..})$

$(x_{it}-x_{i.})$ è la trasformazione within (dentro, intra-individuals) di x_{it}
$(x_{i.} - x_{..})$ è la trasformazione between (fra, inter-individuals) di x_{it}

Le sommatorie di trasformazioni within sono definite:
$W_{xx} = \Sigma_i\, \Sigma_t\, (x_{it} - x_{i.})^2$ devianza di x dentro i
$W_{yy} = \Sigma_i\, \Sigma_t\, (y_{it} - y_{i.})^2$ devianza di y dentro i
$W_{xy} = \Sigma_i\, \Sigma_t\, [(x_{it} - x_{i.})\,(y_{it} - y_{i.})]$ codevianza di x e y dentro i

Le sommatorie di trasformazioni between sono definite:
$B_{xx} = T\, \Sigma_i\, (x_{i.} - x_{..})^2$ devianza di x fra i
$B_{yy} = T\, \Sigma_i\, (y_{i.} - y_{..})^2$ devianza di y fra i
$B_{xy} = T\, \Sigma_i\, [(x_{i.} - x_{..})\,(y_{i.} - y_{..})]$ codevianza di x e y fra i

Le sommatorie di scarti rispetto alle medie complessive sono infine definite:
$T_{xx} = \Sigma_i\, \Sigma_t\, (x_{it} - x_{..})^2$ devianza totale di x
$T_{yy} = \Sigma_i\, \Sigma_t\, (y_{it} - y_{..})^2$ devianza totale di y
$T_{xy} = \Sigma_i\, \Sigma_t\, [(x_{it} - x_{..})\,(y_{it} - y_{..})]$ codevianza totale di x e y

Prendendo la somma in i e t dei quadrati della definizione (2) si ottiene una utile identità:

$$\Sigma_i\, \Sigma_t\, (x_{it} - x_{..})^2 = \Sigma_i\, \Sigma_t\, [(x_{it} - x_{i.})+(x_{i.} - x_{..})]^2 =$$
$$= \Sigma_i\, \Sigma_t\, [(x_{it} - x_{i.})^2+(x_{i.} - x_{..})^2+2\,(x_{it} - x_{i.})(x_{i.} - x_{..})] =$$
$$= \Sigma_i\, \Sigma_t\, (x_{it} - x_{i.})^2 + \Sigma_i\, \Sigma_t\, (x_{i.} - x_{..})^2$$

da cui, utilizzando la precedente notazione si può scrivere:

$$T_{xx} = W_{xx} + B_{xx}$$
(3)
$$T_{yy} = W_{yy} + B_{yy}$$
$$T_{xy} = W_{xy} + B_{xy}$$

La devianza (codevianza) **totale** è scomposta nella devianza (codevianza) **dentro** l'individuo e in quella **fra** individui.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6)

[C.1] Modelli con pendenze costanti

[C.1.1] Modello pooled

(1.1) $y_{it} = a + b\, x_{it} + \varepsilon_{it}$

ipotesi di specificazione:

- (1.1) è il 'vero' legame fra y e x;
- x è esogena in senso contemporaneo (la relazione è ristretta allo stesso periodo temporale): $E(\varepsilon_{it}|x_{it})=0,\ \forall\ i=1,...,N,\ \forall\ t=1,...,T$;
- $\varepsilon_{it} \sim IID\ (0, \sigma_\varepsilon^2)$;

→ Metodo di stima: OLS (pooled)

$$\hat{b}_{POLS} = \frac{T_{xy}}{T_{xx}} = \frac{W_{xy} + B_{xy}}{W_{xx} + B_{xx}}$$

$$\hat{a}_{POLS} = y_{..} - \hat{b}_{POLS}\ x_{..}$$

Se valgono le precedenti ipotesi di specificazione (completa poolability), lo stimatore OLS è BLUE.

In presenza di <u>effetti individuali correlati con le variabili esplicative</u>, il modello (1.1) è scorretto e le stime OLS sono distorte e inconsistenti.

Esempio con dati artificiali: "montec.xls"

DGP: $y_{it} = a_i + b\, x_{it} + \varepsilon_{it}$ $x_{it} = \xi_{it} + 2\, a_i$
campione: $i = 1, 2, 3$ $t = 1, 2, ..., 10$
parametri: $a_1 = 1$; $a_2 = 2$; $a_3 = 3$ $b = 0.5$
estrazioni: $\varepsilon_{it} \sim IIN(0, 1)$ $\xi_{it} \sim IIN(0, 1)$
Quindi il modello vero è un modello con effetti individuali, mentre specifico e stimo un modello pooled.
<u>Nota</u>: nell'esperimento $E(x_{it}\, a_i) \neq 0$.

Dependent Variable: Y
Method: Least Squares
Sample: 1 30
Included observations: 30

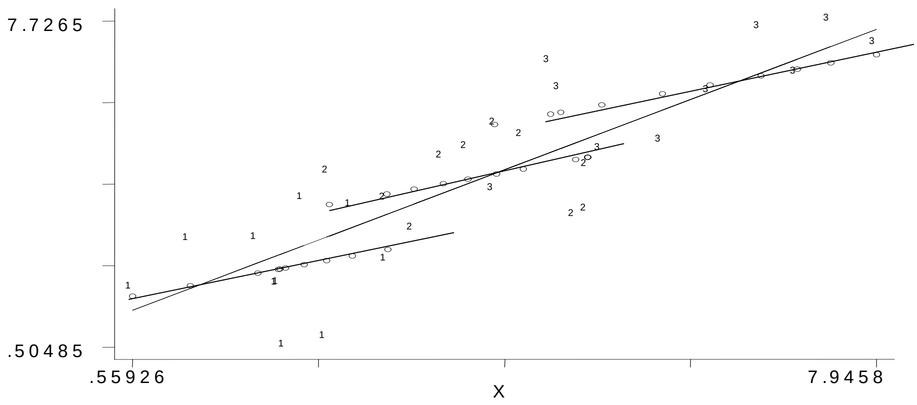
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.847623	0.464185	1.826048	0.0785
X	0.842679	0.104280	8.080926	0.0000

R-squared	0.699897	Mean dependent var	4.240601
Adjusted R-sq	0.689179	S.D. dependent var	1.944447
S.E. of regression	1.084055	Akaike info criterion	3.063634
Sum squared resid	32.90488	Schwarz criterion	3.157047
Log likelihood	-43.95451	F-statistic	65.30136
Durbin-Watson stat	2.284252	Prob(F-statistic)	0.000000

La stima OLS del modello pooled è pesantemente distorta (b_{OLS}=0.84 contro b=0.5) per la mancata specificazione degli effetti individuali correlati con le esplicative (gli effetti individuali fanno slittare la retta di regressione): la stima OLS pooled tende a sovrastimare (o sottostimare se $Corr\ (x_{it}, a_i) < 0$) la pendenza.

La Figura 1 riporta il caso simulato.

Fig. 1 - Spezzata vera e interpolante pooled (falsa)



Lo stimatore OLS pooled non dà conto del fatto che T osservazioni temporali per N individui diversi non è detto siano la stessa cosa che NT diversi individui: dando lo stesso peso alle due fonti di variabilità within e between, lo stimatore OLS pooled ignora la struttura panel dei dati.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6.1)

[C.1.2] Modello con effetti fissi

(1.2') $y_{it} = a + b\, x_{it} + \sum_{j=1}^{N-1} \mu_j D_{ji} + \varepsilon_{it}$

oppure:

(1.2*) $y_{it} = a_i + b\, x_{it} + \varepsilon_{it}$

ipotesi di specificazione:

- (1.2') o (1.2*) rappresentano il 'vero' legame fra y e x;
- x è esogena in senso stretto (il termine di errore è non correlato con le variabili esplicative in tutti i periodi temporali): $E(\varepsilon_{it}|x_{is})=0$, $\forall$ i=1,...,N, $\forall$ t, s=1,...,T;
- $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$;

→ Metodo di stima: OLS con variabili dummy (least squares with dummy variables, LSDV)

→ Metodo di stima: stimatore within (fixed-effects or within-groups estimator, FE)

$$\hat{b}_{LSDV} \equiv \hat{b}_{FE} = \frac{W_{xy}}{W_{xx}}$$

$$\hat{a}_{LSDV,i} = y_{i.} - \hat{b}_{LSDV}\, x_{i.}$$

$$\hat{a}_{LSDV} = \hat{a}_{FE} = \frac{1}{N}\sum_i \hat{a}_{LSDV,i}$$

Il simbolo ≡ in prima equazione indica che la stima della pendenza coincide nei due metodi LSDV e FE: dal Teorema di Frisch-Waugh-Lovell[1], la stima LSDV della (1.2') può essere riprodotta stimando Qy = QXb + Qε, dove Q è la matrice (NT×NT) che produce i residui di una regressione con esplicative le N variabili dummy. Per esempio, stimando OLS l'equazione y = Dμ + ε_y, si ottiene $\hat{\varepsilon}_y = y - D\hat{\mu} = y - D(D'D)^{-1}D'y = \left[I_{NT} - D(D'D)^{-1}D'\right]y = Qy$, dove D è la matrice (NT×N) delle variabili dummy individuali,

$$D = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 1 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix} = I_N \otimes e_T,$$

con I_N matrice (N×N) identità ed e_T vettore (T×1) di 1.[2] Esplicitando alcuni calcoli[3], deriva che:

$$D(D'D)^{-1}D' = (I_N \otimes e_T)[I_N \otimes (1/T)](I_N \otimes e_T)' = I_N \otimes e_T(1/T)(I_N \otimes e_T)' = I_N \otimes e_T e_T'(1/T)$$

matrice (NT×NT) che, premoltiplicata per una variabile, ne genera le medie individuali; $Q = I_{NT} - I_N \otimes e_T e_T'(1/T)$ matrice (NT×NT) che, premoltiplicata per una variabile, ne genera gli scostamenti dalle medie individuali. Pertanto, la stima Qy = QXb + Qε equivale alla stima OLS sui dati trasformati within: $(y_{it}-y_{i.}) = b(x_{it}-x_{i.}) + (\varepsilon_{it}-\varepsilon_{i.})$. Questa trasformazione ha il vantaggio di ridurre il carico computazionale: nella stima LSDV occorre invertire una matrice (K+N)×(K+N), dove K è il numero di esplicative (nel nostro caso K=1); con dati trasformati within la dimensione scende a K×K.

L'assenza del simbolo $\hat{a}_{FE,i}$ nella seconda equazione indica che, dallo stimatore within generalmente attuato dai software econometrici, non si ottiene esplicitamente la stima degli N-1 effetti individuali, così come invece avviene nel caso di stima LSDV[4].

La sostituzione del simbolo ≡ con il simbolo = in terza equazione indica che la coincidenza nella stima della costante tra i due metodi di stima si verifica solo nel caso in cui lo stimatore LSDV includa tutte le N variabili dummy (la costante va poi calcolata come media delle N stime), invece che la costante ed N-1 variabili dummy.

Se valgono le precedenti ipotesi di specificazione, lo stimatore LSDV-FE è BLUE. Inoltre, $\hat{b}_{LSDV}$ è consistente per N→∞ e/o T→∞ ; $\hat{a}_{LSDV,i}$ è consistente per T→∞.

---

1 Frisch R. and Waugh F. V. (1933) Partial Time Regressions as Compared with Individual Trends, Econometrica, 1, 387-401; Lovell M. C. (1963) Seasonal Adjustment of Economic Time Series, Journal of the American Statistical Association, 58, 993-1010.

2 Il simbolo ⊗ indica il prodotto di Kronecker, in base al quale ciascun elemento di una matrice A (N×L) viene moltiplicato per l'intera matrice B (R×P), per cui A ⊗ B = (NR×LP).

3 $D'D = (I_N \otimes e_T)'(I_N \otimes e_T) = I_N \otimes e_T' e_T = I_N \otimes T$, in quanto la matrice identità è idempotente (cioè: I'I=I). $(D'D)^{-1} = I_N \otimes (1/T)$.

4 Gli N-1 effetti individuali possono comunque essere calcolati in un secondo tempo, sulla base dell'espressione in seconda equazione.

Se (1.1) è il DGP, ma si stima per errore il modello (1.2'), lo stimatore LSDV è ancora corretto e consistente, ma inefficiente (stima N-1 parametri irrilevanti).

Il modello LSDV-FE non dà nessun peso alla variabilità fra (between) individui perché le componenti individuali $y_{i.}$ e $x_{i.}$ sono sottratte completamente dalle osservazioni y_{it} e x_{it} : ciò che rileva sono gli scostamenti $\rightarrow$ lo stimatore LSDV-FE usa solo la variabilità interna (within) a ciascun individuo.

Se x_{it} non ha variabilità temporale (ad esempio, il genere) la sua trasformazione within assume sempre valore zero ed il corrispondente coefficiente non può essere stimato. Fattori individuali quali il settore di appartenenza e l'abilità manageriale delle imprese o l'istruzione dei lavoratori possono essere inclusi in x_{it} e stimati solo se hanno una minima variabilità nel tempo, almeno per alcuni individui.[5]

Il modello FE è un modello di "bias-reducing" perché l'omissione di variabili z_i time-invariant è sopperita dall'inclusione di effetti fissi. Il prezzo da pagare è la non considerazione della variabilità between (comunque distorta a causa dell'omissione di z_i).

Una trasformazione alternativa a quella within è la differenza prima (Δ):

$$
\begin{array}{llllc}
y_{it} & = a_i & + b\, x_{it} & + \varepsilon_{it} & - \\
y_{it-1} & = a_i & + b\, x_{it-1} & + \varepsilon_{it-1} & = \\
\hline
\Delta y_{it} & = & b\, \Delta x_{it} & \Delta\varepsilon_{it} &
\end{array}
$$

Il calcolo della differenza prima fa perdere N osservazioni delle NT complessive; in altri termini, si perde l'osservazione temporale iniziale di ciascuna cross-section, che rimane con T-1, invece che T, osservazioni disponibili.

Le trasformazioni within e Δ portano allo stesso risultato: eliminano la presenza di effetti individuali nei dati e/o l'influenza di qualsiasi variabile omessa time-invariant. Quindi, lo stimatore OLS applicato a dati trasformati Δ o within è sempre corretto e consistente $\rightarrow$ stime molto diverse sono perciò interpretabili come evidenza che alcune ipotesi del modello FE non sono valide.
Casi di scorretta specificazione: (i) le variabili esplicative x_{it} sono misurate con errore; (ii) i dati sono affetti da dynamic selection.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6.2)

[C.1.3] Modello con effetti casuali

(1.2'') $y_{it} = a + b\, x_{it} + v_{2it}$ $\qquad$ $v_{2it} = \varepsilon_{it} + \mu_i$

ipotesi di specificazione:

- (1.2'') è il 'vero' legame fra y e x;
- x è esogena in senso stretto: $E(\varepsilon_{it}|x_{is})=0$, $\forall$ i=1,...,N, $\forall$ t, s=1,...,T;
- $\mu_i \sim IID\ (0, \sigma_\mu^2)$;
- $Cov(x_{it}, \mu_i) = 0$;
- $\varepsilon_{it} \sim IID\ (0, \sigma_\varepsilon^2)$;
- $Cov(\varepsilon_{it}, \mu_i) = 0$

[5] Esistono metodi IV o GMM che permettono la stima di effetti time-invariant. Si vedano: Hausman J. A.-Taylor W. E. (1981) Panel data and unobservable individual effects, Econometrica, 52, 1219-1240; Amemiya T.-MaCurdy T. (1986) Instrumental-variable estimation of an error-components model, Econometrica, 54, 869-881; Breusch T., Mizon G. and Schmidt P. (1989) Efficient estimation using panel data, Econometrica, 57, 695-700; Arellano M.-Bover O. (1995) Another look at the instrumental variable estimation of error-components models, Journal of Econometrics, 68, 29-51.

da queste ipotesi si derivano le seguenti proprietà statistiche per l'errore composto v_{2it}:

- $E(v_{2it}) = 0$;
- $E(v_{2it}^2) = \sigma_\varepsilon^2 + \sigma_\mu^2$;
- $Cov(v_{2it}, v_{2is}) = \sigma_\mu^2$ $(t \neq s)$: gli errori dell'i-esimo individuo in diversi istanti temporali sono correlati
- $Cov(v_{2it}, v_{2jt}) = 0$ $(i \neq j)$: gli errori di diversi individui sono sempre incorrelati

Lo stimatore FE-LSDV su dati generati da (1.2") è corretto e consistente, ma inefficiente in quanto stima N-1 parametri inutili e la matrice di varianze-covarianze dell'errore composito v_2 non è una matrice identità a meno di uno scalare.

Infatti, supponendo per semplicità N=2 e T=3, $E(v_2 v_2')$ è una matrice diagonale a blocchi:

i, t	1,1	1,2	1,3	2,1	2,2	2,3
1,1	$\sigma_\varepsilon^2+\sigma_\mu^2$	σ_μ^2	σ_μ^2	0	0	0
1,2	σ_μ^2	$\sigma_\varepsilon^2+\sigma_\mu^2$	σ_μ^2	0	0	0
1,3	σ_μ^2	σ_μ^2	$\sigma_\varepsilon^2+\sigma_\mu^2$	0	0	0
2,1	0	0	0	$\sigma_\varepsilon^2+\sigma_\mu^2$	σ_μ^2	σ_μ^2
2,2	0	0	0	σ_μ^2	$\sigma_\varepsilon^2+\sigma_\mu^2$	σ_μ^2
2,3	0	0	0	σ_μ^2	σ_μ^2	$\sigma_\varepsilon^2+\sigma_\mu^2$

Principali caratteristiche di $E(v_2 v_2')$:

- al contrario del modello AR stazionario, la correlazione a coppie fra disturbi di diversi periodi non decresce a mano a mano che la distanza t-s aumenta, ma rimane costante perché v_{2it} e v_{2is} includono entrambi μ_i → forte autocorrelazione;
- ogni blocco individuale non dipende da i e, quindi, è lo stesso per tutti gli individui → omoschedasticità.

→ Metodo di stima: GLS

$$\hat{a}_{GLS} = \hat{a}_{RE} = y_{..} - \hat{b}_{GLS}\, x_{..}$$

(4) $$\hat{b}_{GLS} = \hat{b}_{RE} = (W_{xy} + \theta\, B_{xy}) / (W_{xx} + \theta\, B_{xx})$$

dove:

(4') $$\theta = \sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + T\, \sigma_\mu^2)$$

Se valgono le precedenti ipotesi di specificazione, lo stimatore RE-GLS è BLUE.

Lo stimatore GLS del modello con effetti casuali utilizza informazioni sia sulla variabilità <u>within</u> sia su quella <u>between</u>. Il parametro θ è il peso che lo stimatore RE-GLS attribuisce alla variabilità fra individui; infatti, dalla (4) si nota che:

$$\theta = 0 \quad \rightarrow \quad \hat{b}_{GLS} \equiv \hat{b}_{LSDV}$$

$$\theta = 1 \quad \rightarrow \quad \hat{b}_{GLS} \equiv \hat{b}_{OLS}$$

A seconda dei valori di θ , $\hat{b}_{GLS}$ si avvicina maggiormente alla stima con effetti fissi o a quella <u>pooled</u> (senza effetti individuali).
Stimatori OLS pooled e FE-LSDV sono modi "all-or-nothing" di utilizzare l'informazione fra individui: OLS tratta indistintamente tutte ("all") le fonti di variabilità, mentre FE-LSDV non dà nessun peso ("nothing") alla variabilità between.

Lo stimatore RE-GLS rappresenta perciò il caso intermedio fra due approcci estremi nel considerare gli effetti individuali: tutti uguali $\rightarrow$ poolability completa (OLS) o tutti diversi $\rightarrow$ completamente non poolable (FE-LSDV).

Alla luce della (4'), $\theta \cong 0$ implica una variabilità degli effetti individuali σ_μ^2 molto maggiore di σ_ε^2, oppure $T\rightarrow\infty$. In questi casi, $\hat{b}_{GLS} \cong \hat{b}_{LSDV}$. I GLS sono efficienti rispetto a LSDV, ma il guadagno di efficienza tende a zero per T che tende all'infinito.

Viceversa, quando $\theta \cong 1$ la variabilità degli effetti individuali non rileva significativamente e

$\hat{b}_{GLS} \cong \hat{b}_{OLS}$.

Contrariamente ai metodi di stima OLS e LSDV, la formula (4)-(4') non è applicabile ai dati (da cui la terminologia "unfeasible GLS") perché non si conosce θ $\rightarrow$ deve essere stimato sulla base di stime di σ_ε^2 e σ_μ^2 (la letteratura suggerisce possibilità alternative di "feasible GLS", da cui diversi valori di $\hat{b}_{GLS}$).

Se s_ε^2 e s_μ^2 sono stime consistenti di σ_ε^2 e σ_μ^2 , gli stimatori feasible GLS alternativi sono tutti asintoticamente equivalenti ai GLS.

Una possibilità[6] è (i) stimare il modello FE e dai residui $\hat{\varepsilon}_{LSDV,it}$ ottenere una stima di σ_ε^2 :

$s_\varepsilon^2 = \Sigma_i \Sigma_t \hat{\varepsilon}_{LSDV,it}{}^2 /(NT\text{-}N\text{-}K)$

dove K è il numero di regressori esclusa l'intercetta (nel nostro caso K=1). Dopodiché, (ii.a) stimare σ_μ^2 a partire dagli effetti fissi ottenuti al passo (i); tuttavia, s_μ^2 è distorto ed inconsistente. L'alternativa è (ii.b) utilizzare la stima s_v^2 della varianza degli errori della regressione between.

Il modello between (BE) è derivato prendendo le medie individuali del modello (1.2"):

$$y_{i.} = a + b\, x_{i.} + v_{i.} \qquad\qquad v_{i.} = \varepsilon_{i.} + \mu_i$$

dove: $Var(v_{i.}) = \sigma_v^2 = \sigma_\varepsilon^2/T + \sigma_\mu^2$. Dato che gli errori $v_{i.}$ sono incorrelati si usa la stima OLS:

$\hat{b}_{BE} = B_{xy} / B_{xx}$.

Ottenuti i residui $\hat{v}_{BE,i.}$ si ha che: $s_v^2 = \Sigma_i\, \hat{v}_{BE,i.}{}^2 /(N\text{-}K)$.

Date la stima s_ε^2 ottenuta al passo (i) e la s_v^2 , la stima s_μ^2 è ricavata dalla formula: $s_\mu^2 = s_v^2 - s_\varepsilon^2/T$.[7]

Dalla procedura di stima, emerge che il modello RE è un modello di 'pooling' di alternative fonti di informazione e non di bias-reducing.

Fuller-Battese (1973, 1974)[8] dimostrano che lo stimatore GLS (4)-(4') equivale allo stimatore OLS applicato ai dati trasformati:

$(y_{it} - \lambda\, y_{i.})$ e $(x_{it} - \lambda\, x_{i.})$, dove $\lambda = 1-\theta^{1/2}$.

Si definiscano $y_{it}^* = [y_{it}-\lambda y_{i.}-(1-\lambda)y_{..}]$ e $x_{it}^* = [x_{it}-\lambda x_{i.}-(1-\lambda)x_{..}]$. La stima OLS del modello sarà:

---

[6] Swamy-Arora feasible GLS estimator; Swamy P.A.V.B. and Arora S.S. (1972) The exact finite sample properties of the estimators of coefficients in the error components regression models, Econometrica, 40, 261-275.

[7] Non sempre la stima s_μ^2 viene positiva; possibili soluzioni sono sostituire le stime negative con zero oppure utilizzare una stima di θ data da s_ε^2 su T s_v^2. 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, 68, 324-328, dimostrano con simulazioni che la scelta di un metodo particolare per ottenere i feasible GLS non influenza significativamente le stime.

[8] Fuller W.A. and Battese G.E. (1973) Transformations for estimation of linear models with nested error structure, Jaournal of the American Statistical Association, 68, 626-632; (1974) Estimation of linear models with cross-error structure, Journal of Econometrics, 2, 67-78.

$\hat{b}_{GLS} = [\Sigma_i \Sigma_t (x_{it}^* y_{it}^*)]/(\Sigma_i \Sigma_t x_{it}^{*2}) = \{W_{xy}+\theta T\Sigma_i[(x_{i.}-x_{..})(y_{i.}-y_{..})]\}/[W_{xx}+\theta T\Sigma_i(x_{i.}-x_{..})^2]$

Cioè come nella (4). [hint: sostituisci $\lambda=(1-\theta^{½})$ e $(1-\lambda)=\theta^{½}$ e sviluppa le sommatorie ...].

Il parametro λ indica quanta parte individuale sottraggo ad ogni osservazione effettiva prima di stimare OLS che tratta tutte le osservazioni allo stesso modo (poolable). → λ rende omogenei i dati relativi ad individui diversi.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6.3, SECTION 6.4)

[C.1.4] Meglio i modelli panel con effetti fissi o casuali?

Alcune argomentazioni favoriscono l'impiego del modello FE:

- è facile e immediato da stimare (no approssimazioni come in GLS);
- è robusto all'omissione di variabili esplicative time-invariant;
- è adatta per stimare effetti specifici del campione (panel di paesi, di settori, di regioni);
- è consistente anche quando le caratteristiche individuali sono correlate con le esplicative;
- anche nei casi in cui è valido il modello RE, è comunque consistente (perde solo l'efficienza).

Altre argomentazioni favoriscono invece l'impiego del modello RE:

- è metodologicamente coerente: tratta l'effetto individuale come ill termine di errore;
- permette di risparmiare molti gradi di libertà (specialmente per N grande);
- permette di studiare l'influenza di esplicative time-invariant (ad es. di selection bias);
- è adatto per campioni casuali e inferisce sulla popolazione (previsioni individui out-of-sample);
- tiene conto della varianza between e non solo di quella within: efficiente uso di entrambe.

Discussione

La scelta di trattare gli effetti individuali come fissi o come casuali è delicata e può essere ricondotta ad una serie di fattori:

- Le determinanti degli effetti individuali: se sono motivati da un elevato numero di circostanze casuali e non osservabili, è più indicato il modello random.
- Il numero di individui: per N grande e T piccolo, il modello FE ha pochi gradi di libertà (stime non troppo affidabili). Se l'interesse è soprattutto sulle pendenze, meglio filtrare le differenze individuali (senza stimarle esplicitamente) con il modello RE.
- La natura del campione: quando il campione è chiuso ed esaustivo (come nel caso di paesi o di settori), gli effetti fissi sono i naturali candidati. Quando il campione è aperto (N individui sono estratti da una popolazione), la specificazione random è più interessante.
- Il tipo di inferenza: spetta al ricercatore scegliere se desidera fare inferenza sulle caratteristiche della popolazione (perché interessato anche al comportamento degli individui esclusi dal campione) mediante inferenza non condizionale, o concentrarsi sugli effetti presenti nel campione (perché l'interesse è proprio su quelli) mediante inferenza condizionale ai soli individui nel campione.

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6.5)

Inferenza

La verifica statistica del quesito è effettuata mediante il **test di Hausman** che è un test del confronto fra il risultato di stimatori alternativi, secondo lo schema:

Ipotesi:	Modello:	effetti random (RE)	effetti fissi (FE)
H_0: Cov(x_{it} , μ_i) = 0		**consistente ed efficiente**	consistente inefficiente
H_1: Cov(x_{it} , μ_i) $\neq$ 0		incosistente	**consistente**

Sotto H_0 il modello RE è il migliore (la stima GLS è BLUE), mentre, sotto H_1, le proprietà statistiche dello stimatore GLS del modello RE vengono meno. La stima LSDV del modello FE è consistente sia sotto H_0 che sotto H_1, ma non è efficiente sotto H_0. Perciò, sotto H_0, le stime saranno statisticamente simili e, quindi, la scelta ricade su quella GLS del modello RE; viceversa sotto l'alternativa H_1: la scelta ricade si FE che, almeno, è efficiente.

Formalmente, si definisce la differenza fra stime: $\hat{q} = \hat{b}_{LSDV} - \hat{b}_{GLS}$, che può essere riscritta anche come: $\hat{b}_{LSDV} = \hat{q} + \hat{b}_{GLS}$. L'idea è quella di verificare se q è significativamente diversa da zero: il non rifiuto dell'ipotesi nulla implica che plim $\hat{q}$ = plim $\hat{b}_{LSDV}$ - plim $\hat{b}_{GLS}$ = $b - b = 0$ perché entrambi gli stimatori sono consistenti.

Per definizione, si ha che: var($\hat{b}_{LSDV}$) = var($\hat{q}$) + var($\hat{b}_{GLS}$) + 2cov($\hat{q}, \hat{b}_{GLS}$). Il contributo fondamentale di Hausman (1978) è dimostrare che, sotto l'ipotesi nulla, Cov($\hat{q}$, $\hat{b}_{GLS}$) = 0, da cui var($\hat{q}$) = var($\hat{b}_{LSDV}$) - var($\hat{b}_{GLS}$). In generale, in un modello con K stime da confrontare, la statistica-test è quindi: $m = q' [Var(q)]^{-1} q$ che si distribuisce, sotto H_0, come $\chi^2_{(K)}$.

Da tutto ciò deriva però un aspetto negativo del test: in campioni finiti, la stima della varianza di $\hat{b}_{GLS}$ può essere numericamente maggiore di quella di $\hat{b}_{LSDV}$, da cui una var($\hat{q}$) negativa!

La strategia per provare che Cov($\hat{q}$, $\hat{b}_{GLS}$) = 0 nel caso di K=1 muove da un qualsiasi stimatore di b consistente che, sotto H_0, viene definito come $\tilde{b} = \hat{b}_{GLS} + \lambda\hat{q}$ (dove λ è una qualsiasi costante $\neq$ 0), la cui varianza è per definizione var($\tilde{b}$) = var($\hat{b}_{GLS}$) + λ^2 var($\hat{q}$) + 2λ cov($\hat{b}_{GLS}, \hat{q}$). Dato che, sotto H_0, $\hat{b}_{GLS}$ è anche efficiente, deve essere che var($\tilde{b}$) $\geq$ var($\hat{b}_{GLS}$) e, quindi, la seguente diseguaglianza: λ^2 var($\hat{q}$) + 2λ cov($\hat{b}_{GLS}, \hat{q}$) ≥ 0 per ogni λ. Ma quando cov($\hat{b}_{GLS}, \hat{q}$) > 0, un λ negativo compreso nell'intervallo -2cov($\hat{b}_{GLS}, \hat{q}$)/var($\hat{q}$) < λ < 0 implica che la diseguaglianza sopra non è soddisfatta ($\hat{b}_{GLS}$ non sarebbe più uno stimatore efficiente). Allo stesso modo, quando cov($\hat{b}_{GLS}, \hat{q}$) < 0 con λ positivo compreso nell'intervallo 0 < λ < -2cov($\hat{b}_{GLS}, \hat{q}$)/var($\hat{q}$). Pertanto, la validità della precedente diseguaglianza implica che cov($\hat{b}_{GLS}, \hat{q}$) non può essere maggiore o minore di zero, ma solo zero. Infatti, quando cov($\hat{b}_{GLS}, \hat{q}$) = 0, la diseguaglianza è verificata per qualsiasi λ.

Il modello RE "supera" il test anche quando la non significatività dello scarto fra stime dipende da una loro bassa significatività. Perciò, si suggerisce di "osservare attentamente" anche le stime puntuali (soprattutto quando sono interpretabili in senso economico).

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6.3)

[C.1.5] La dicotomia mancata: il punto di vista di Mundlak

Supponiamo che gli effetti stocastici a_i dipendano linearmente da $x_{i.}$ (media per individuo di x_{it}) più una componente casuale w_i . La linearità della relazione dipende dall'ulteriore ipotesi di normalità dell'effetto individuale e di x_{it} . In pratica, si collega una parte della variabilità di a_i alla variabilità delle $x_{i.}$:

$$a_i = \pi\, x_{i.} + w_i$$

L'esclusione dell'intercetta semplifica l'algebra senza modificare la sostanza dei risultati (è come se si supponesse che y e x abbiano media zero).

Ad esempio, in una funzione di produzione specificata come RE, a_i è l'effetto individuale sconosciuto all'econometrico, ma noto all'imprenditore che, pertanto, utilizza questa informazione nella scelta dei fattori produttivi x_{it} (in base alle caratteristiche del proprio impianto in a_i). L'errore di specificazione nel considerare a_i e x_{it} non correlati causa la distorsione dell'approccio RE e non l'appropriatezza o meno dell'inferenza RE (Cfr la discussione sopra, relativa al test di Hausman).

Sostituendo la nuova definizione di effetto casuale nella (1.2) si ottiene:

$$(1.2^{M1}) \qquad y_{it} = \pi\, x_{i.} + b\, x_{it} + v_{M1,it}$$

dove $v_{M1,it} = [\, w_i + \varepsilon_{it}\,]$ sono gli errori stocastici.

Utilizzando la formulazione di Fuller-Battese, si ha che la stima GLS del modello (1.2^{M1}) equivale alla stima OLS del modello:

$$(y_{it}\text{-}\lambda\, y_{i.}) = \pi\, (x_{i.}\text{-}\lambda\, x_{i.}) + b\, (x_{it}\text{-}\lambda\, x_{i.}) + v_{M2,it}$$

da cui, riparametrizzando:

$$(1.2^{M2}) \qquad (y_{it}\text{-}\lambda\, y_{i.}) = b\, (x_{it}\text{-}x_{i.}) + \delta\, x_{i.} + v_{M2,it}$$

dove: $v_{M2,it} = (v_{M1,it} \text{-}\lambda\, v_{M1,i.}\,)$; $\delta = (\pi+b)\,(1\text{-}\lambda)$. Dato che $(x_{it}\text{-}x_{i.})$ e $x_{i.}$ sono ortogonali e che Cov $[(x_{it}\text{-}x_{i.})\,(y_{it}\text{-}\lambda y_{i.})] = W_{xy}$ [perché $\Sigma_{i,t}\, (x_{it}\text{-}x_{i.})\theta^{½}y_{i.} = 0$], lo stimatore OLS della (1.2^{M2}) sarà:

$$\hat{b}_{GLS} = W_{xy} / W_{xx} \equiv \hat{b}_{LSDV}$$

Se gli effetti casuali sono correlati con le medie individuali delle variabili esplicative, la stima $\hat{b}_{GLS}$ del modello con effetti casuali è la stessa del modello con effetti fissi.

Mundlak (1978) dimostra che il precedente risultato vale anche quando gli effetti individuali sono funzione di altre variabili z_i esplicative time-invariant.

Il precedente risultato non vale invece se nel vettore π ci sono alcuni parametri nulli (non tutte le esplicative del vettore x_{it} influenzano a_i), a meno che le esplicative associate ai π=0 siano ortogonali a quelle associate ai $\pi \neq 0$.

Se come ritiene Mundlak, c'è sempre correlazione fra le variabili esplicative e gli effetti individuali, allora la scelta fra modelli RE e FE è inutile ed arbitraria perché, quando il modello è correttamente specificato, c'è un solo stimatore: quello ad effetti fissi.

Per una visione meno estrema, si ricordi che il presupposto dei risultati di Mundlak, $Cov(a_i\, , x_{it}) \neq 0$, può sempre essere verificato col test di Hausman.

In pratica, il punto non è discriminare fra effetti fissi e casuali, ma fra modelli con effetti individuali correlati o non correlati con le variabili esplicative.

→ La condizione di esogenità delle esplicative rispetto a tutte le componenti di errore stocastico:

$$Cov(a_i\, , x_{it}) = 0 \qquad \text{e} \qquad Cov(\varepsilon_{it}\, , x_{it}) = 0$$

garantisce la consistenza degli stimatori OLS, LSDV e GLS di b.

[C.1.6] Meglio i modelli panel o pooled con parametri fissi?

Due test di specificazione rispondono alla questione, a seconda che gli effetti individuali del panel (modello non vincolato) siano fissi o casuali.

Se nel panel con effetti fissi le differenze fra intercette non sono significative, le N componenti individuali collassano in una sola $\rightarrow$ il modello OLS pooled è più parsimonioso. Il **test di significatività delle differenze** fra le N intercette è una F e richiede le stime dei modelli vincolato e non vincolato. Si stima il modello non vincolato (1.2') con LSDV e si salvano i residui $\hat{\varepsilon}_{LSDV,it}$; si stima il modello vincolato (1.1) con OLS pooled e si salvano i residui $\hat{\varepsilon}_{OLS,it}$. Si costruisce:

$$F_0=\frac{\sum_i \sum_t \hat{\varepsilon}^2_{OLS,it} \quad \hat{\varepsilon}^2_{LSDV,it}}{\sum_i \sum_t \hat{\varepsilon}^2_{LSDV,it}} \frac{NT-N-K}{N-1}$$

Sotto H_0: $a_i = a$ $\forall i$ (cioé $\mu_i = 0$ per i=1, 2, ..., N-1), la statistica F_0 si distribuisce come una: $F(N-1, NT-N-K)$.

Se nel modello con effetti casuali si ha che $\sigma_\mu^2 = 0$ (ipotesi nulla), gli effetti individuali collassano nel modello vincolato OLS pooled.

Il **test di Breusch-Pagan** è di tipo LM perché richiede solo la stima vincolata. La statistica è ottenuta dai residui $\hat{\varepsilon}_{OLS,it}$ del modello pooled (1.1):

$S_1 = \Sigma_i (\Sigma_t \hat{\varepsilon}_{OLS,it})^2$ e $S_2 = \Sigma_i \Sigma_t \hat{\varepsilon}_{OLS,it}{}^2$

da cui si ottiene:

$$LM=\frac{NT}{2\times(T-1)}\left(\frac{S_1}{S_2}-1\right)^2$$

che sotto H_0 si distribuisce come un χ^2 con 1 grado di libertà (1 vincolo).

☺A: LECTURE_PANEL_STATIC_APPLICATION (SECTION 6.2, SECTION 6.3)

[C.2] Modelli con pendenze specifiche (individuali)

[C.2.1] Modello di regressioni apparentemente non collegate

(1.4) $y_{it} = a_i + b_i x_{it} + \varepsilon_{it}$

ipotesi di specificazione:

- (1.4) è il 'vero' legame fra y e x;
- x è esogena in senso contemporaneo;
- $E(\varepsilon_{it}) = 0$;
- $E(\varepsilon_{it}^2) = \sigma_i^2$;
- $Cov(\varepsilon_{it}, \varepsilon_{is}) = 0$ $(t \neq s)$;
- $Cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}$ $(i \neq j)$.

I parametri di diversi individui sono tra loro diversi e gli errori sono indipendenti dal tempo. Tuttavia, la covarianza σ_{ij} diversa da zero implica che lo stimatore OLS individuo per individuo non sia efficiente; lo è, invece, il metodo GLS (usa l'informazione derivante dalla correlazione fra gli errori di diversi individui).

$\rightarrow$ Metodo di stima: SUR

Zellner (1962) suggerisce questo approccio (di tipo GLS) per modellare i dati panel. L'idea di base è che la covarianza σ_{ij} sia diversa da zero a causa di un certo numero di variabili (omesse, non misurabili) comuni per tutti gli individui (regressioni).

Nel caso semplice di N=2 e T=3 E(εε') ha la seguente forma:

i, t	1,1	1,2	1,3	2,1	2,2	2,3
1,1	σ_1^2	0	0	σ_{12}	0	0
1,2	0	σ_1^2	0	0	σ_{12}	0
1,3	0	0	σ_1^2	0	0	σ_{12}
2,1	σ_{12}	0	0	σ_2^2	0	0
2,2	0	σ_{12}	0	0	σ_2^2	0
2,3	0	0	σ_{12}	0	0	σ_2^2

La matrice di varianza-covarianza ha una forma alquanto dissimile da quella del modello con effetti casuali (caratterizzata da equi-correlazione).

Alcuni fattori non osservabili (inclusi nel termine di errore) influenzano gli individui simultaneamente e determinano covarianze contemporanee fra i disturbi di due diversi individui. Quindi, la struttura delle covarianze dei residui introduce nel modello una relazione di interdipendenza.

La procedura di stima SUR è effettuata per passi.
Passo 1: stima OLS separata di ciascuna delle N relazioni (1.4) e salvataggio dei residui $\hat{\varepsilon}_{OLS,it}$
Passo 2: calcolo (stima) delle covarianze $s_{OLS,ij} = (\Sigma_t\ \hat{\varepsilon}_{OLS,it}\ \hat{\varepsilon}_{OLS,jt}) / (T-K)$, dove K è il numero dei parametri stimati per individuo (nel nostro caso K = 2)
Passo 3: stima GLS simultanea (per tutti gli individui)
Passo 4: eventuali iterazioni (se le ε sono normali, le iterazioni convergono alle stime di massima verosimiglianza

Questo approccio non è consigliabile nel caso di panel longitudinali dove T è piccolo (poca informazione al passo 1) e N è grande (funge da moltiplicatore del numero di parametri: N(K + ½ (N+1)).

L'econometrico oscilla fra desiderio di modellare le diversità individuali e necessità di preservare gradi di libertà (stime affidabili). Il modello OLS pooled è molto parsimonioso ma non considera nessun tipo di specificità individuale, mentre il modello SUR è l'opposto.

Secondo il principio della parsimonia, il metodo SUR è appropriato solo quando i dati sono generati da una popolazione eterogenea, altrimenti è inefficiente (per i molti parametri individuali).

Quando le correlazioni dipendono dall'omissione di variabili rilevanti, non è detto che l'approccio SUR-GLS porti a risultati accettabili (possibili problemi di specificazione del modello).

[C.2.2] Stime OLS separate o stima SUR?

In due casi le stime OLS separate (individuo per individuo) sono uguali alle stime SUR.

- gli errori non sono correlati e, quindi, nulla collega gli individui tra loro;
- anche se gli errori sono tra loro correlati, in ogni relazione individuale compaiono le stesse variabili esplicative con le stesse osservazioni.

In queste due situazioni le stime OLS separate sono ottimali perché utilizzano tutta l'informazione.
Nel caso in cui le osservazioni relative alle esplicative differiscano, si può sottoporre a verifica l'ipotesi di correlazione nulla fra gli errori. Dalle stime delle covarianze $s_{OLS,ij}$ e delle varianze $s_{OLS,i}^2$

si ottengono i coefficienti di correlazione al quadrato, $r_{OLS,ij}^2 = s_{OLS,ij}^2 / (s_{OLS,i}^2 \; s_{OLS,j}^2)$, da cui la statistica test: $\chi^2_{N(N-1)/2} = T \sum_{i=2}^{N} \sum_{j=1}^{i-1} r_{OLS,ij}^2$ che verifica l'ipotesi H_0: $\sigma_{ij} = 0$ per tutte le $i \neq j$.

☺A: LECTURE_SURE_APPLICATION

[C.2.3] Modello con parametri stocastici

Equivale al modello con effetti casuali in cui si ipotizza che anche i parametri b_i siano stocastici. Per semplificare la notazione, tralasciamo l'intercetta (si suppone che le medie siano nulle):

(1.4') $y_{it} = b_i x_{it} + \varepsilon_{it}$ $\qquad b_i = b + \mu_i$

da cui si ottiene:

(1.4") $y_{it} = b\, x_{it} + v_{it}$ $\qquad v_{it} = \varepsilon_{it} + \mu_i x_{it}$

ipotesi di specificazione:

- (1.4') o (1.4") è il 'vero' legame fra y e x;
- x è esogena in senso stretto: $E(\varepsilon_{it}|x_{is})=0$, $\forall$ i=1,...,N, $\forall$ t, s=1,...,T;
- $\varepsilon_{it} \sim IID(0, \sigma_i^2)$;
- $\mu_i \sim IID(0, \sigma_\mu^2)$;
- $Cov(\mu_i , \varepsilon_{it}) = 0$;
- $E(v_{it}) = 0$;
- $Cov(v_{it} , v_{is}) = \sigma_i^2 + \sigma_\mu^2 x_{it}^2$ se $(t=s)$;
- $Cov(v_{it} , v_{is}) = \sigma_\mu^2 x_{it} x_{is}$ se $(t \neq s)$;
- $Cov(v_{it} , v_{js}) = 0$ se $(i \neq j)$;

Il modello è eteroschedastico e, quindi,
→ Metodo di stima: GLS

Senza scendere nei dettagli, si può dimostrare che la stima GLS dell'equazione (1.4") è una media ponderata (con pesi w_i) delle N stime $\hat{b}_{OLS,i}$ (OLS delle equazioni individuali):

$$\hat{b}_{GLS} = \Sigma_i \, w_i \, \hat{b}_{OLS,i}$$

dove i pesi w_i sono definiti da:

$$w_i = 1/(\sigma_\mu^2 + \sigma_{b,i}^2) / \Sigma_i \, [1/(\sigma_\mu^2 + \sigma_{b,i}^2)]$$

dove $\sigma_{b,i}^2 = \sigma_i^2 / \Sigma_t \, x_{it}^2$ sono le varianze di $\hat{b}_{OLS,i}$.

I pesi tendono ad essere minori per i parametri di individui caratterizzati da una maggiore $\sigma_{b,i}^2$. La variabilità complessiva $(\sigma_\mu^2 + \sigma_{b,i}^2)$ varia per individuo a causa delle sole σ_i^2 (dentro $\sigma_{b,i}^2$).

Se la varianza dei disturbi stocastici ai parametri b_i (la stessa per tutti gli i) è molto superiore alle varianze degli stimatori OLS, i pesi sono circa gli stessi per tutti gli individui e la media ponderata è simile alla media semplice: se i disturbi stocastici ai parametri sono la maggiore fonte di variazione, non è utile discriminare l'informazione di individui diversi.

Per rendere "feasible" le stime GLS bisogna disporre di stime dei parametri (sconosciuti) σ_μ^2 e σ_i^2 per calcolare i pesi w_i . Come nel caso del modello con effetti casuali, alternative procedure di stima portano a risultati diversi. Per una rassegna, si veda Maddala et al (1997).
Una possibilità è quella di:

(i) effettuare N stime OLS separate $\hat{b}_{OLS,i}$;

(ii) sulla base dei residui $\hat{\varepsilon}_{OLS,it}$ calcolare le stime di σ_μ^2 e σ_i^2, rispettivamente:

$s_\mu^2 = \Sigma_i(\hat{b}_{OLS,i} - \hat{b}_{OLS,.})^2/(N-1)$

$s_i^2 = \Sigma_t\ \hat{\varepsilon}_{OLS,it}{}^2 / T$

Nota che il modello con effetti casuali (1.2”) e quello con parametri stocastici (1.4”) stimano entrambi K parametri (nel nostro caso K=2: a, b).

Il numero di gradi di libertà è invece molto diverso quando K è grande: il modello RE richiede infatti $T \geq 2$ (per le trasformazioni within), mentre il modello con parametri stocastici necessita della stima OLS individuo per individuo (che implica $T \geq K$).

[C.3] Sintesi: l’importanza della poolability

La modellazione di dati che variano per individuo e nel tempo si riconduce al tema della poolability.

A sua volta, la poolability può essere vista come il confronto fra effetti fissi e casuali e può essere schematizzata nel seguente modo.

[Si considera ancora il caso semplice di una sola variabile esplicativa e di effetti individuali. L’estensione a più esplicative e/o ad effetti temporali è immediata e complica l’algebra senza introdurre nuovi concetti.]

Se l’eterogenità è catturata dalle variabili esplicative del modello statico, allora si ha lo stesso modello (di poolability completa, vale a dire non condizionale) per tutte le NT osservazioni:

(5) $\mathbf{y_{it} = a + b\ x_{it} + u_{it}}$ con: $\mathbf{Cov(x_{it}, u_{it}) = 0}$

Il modello (5) non preclude la possibilità che i parametri varino fra individui, a patto che queste variazioni siano interpretabili come una estrazione puramente casuale dalla stessa distribuzione di probabilità ($\forall$i,t), Cfr Tabella 1.

Tab. 1 – Errori composti per modelli poolable

Errore (*)	Modello con effetto causale	Stimatore BLUE di a, b
$u_{it} = \varepsilon_{it}$	(1.1)	OLS
$u_{it} = \mu_i + \varepsilon_{it}$	(1.2)	GLS
$u_{it} = \mu_{a,i} + \mu_{b,i}\ x_{it} + \varepsilon_{it}$	(1.4)	GLS

(*) nota: $\varepsilon_{it} \sim IID\ (0, \sigma_\varepsilon^2)$

Sotto le condizioni (5), qualsiasi stimatore del modello statico è consistente a prescindere dalle caratteristiche dell’errore u_{it}. Questo fatto è dovuto alla natura stocastica dell’effetto individuale.

Se invece l’eterogeneità non è confinata al solo errore stocastico, i parametri del modello sono tra loro diversi (in tutto o in parte, Cfr Tabella 2):

(6) $\mathbf{y_{it} = a_i + b_i\ x_{it} + \varepsilon_{it}}$ con: $\mathbf{Cov(x_{it}, \varepsilon_{it}) = 0}$

Tab. 2 – Modelli non (o parzialmente) poolable

Parametri del modello (6)	Modello con effetti fissi	Stimatore BLUE di a_i, b_i
$a_i = a + \mu_i$; $b_i \equiv b$	(1.2)	LSDV
$a_i = a_i$; $b_i = b_i$	(1.4)	SUR

Qualsiasi stimatore pooled (in Tabella 1) del modello (6) è distorto e non consistente perché la non poolability implica effetti condizionali a (collegati con) x_{it} .

Qualsiasi stimatore non pooled (in Tabella 2) del modello (5) è consistente, ma inefficiente perché la poolability permetterebbe di ridurre il numero di parametri.

Supponiamo di voler prevedere $y_{i,T+1}$ data una ipotesi esogena per $x_{i,T+1}$. Dal modello (5) si ha che: $y_{i,T+1}^{(5)} = a_{GLS} + b_{GLS}\, x_{i,T+1}$; invece, dal modello (6) one way, si ha che: $y_{i,T+1}^{(6)} = a_{LSDV,i} + b_{LSDV}\, x_{i,T+1}$. E' evidente che la previsione $y_{i,T+1}^{(5)}$ non è condizionale agli individui del campione, mentre $y_{i,T+1}^{(6)}$ è una previsione condizionale all'individuo i-esimo (prevedo proprio quell'individuo perché uso la sua intercetta). Se il modello (6) fosse un modello con effetti two ways non potrei prevedere $y_{i,T+1}$ perché mancherebbe la stima dell'intercetta $a_{i,T+1}$.
→ Il modello (6) ha un senso solo quando l'interesse è specifico per l'individuo i-esimo (ad esempio, un paese, un settore, una regione, ecc.).

[D] Variabili esplicative correlate con l'errore

In questa sezione andremo a modificare le prime due ipotesi di specificazione di tutti i modelli analizzati finora: esogenità delle esplicative [Cfr D.1] e modello statico [Cfr D.2]. La trattazione sarà comune in quanto, come vedremo meglio, la presenza della dipendente ritardata fra le covariate del modello panel dinamico è condizione sufficiente per generare correlazione fra esplicative e termini di errore e, quindi, inconsistenza di tutti gli stimatori.

[D.1] L'approccio delle variabili strumentali

Come visto nella lecture_IV, se nel modello classico di regressione lineare viene meno l'ipotesi di incorrelazione delle variabili esplicative con il termine di errore, lo stimatore OLS (o GLS a seconda delle ipotesi di specificazione degli errori del modello) perde le proprietà della correttezza e della consistenza. Due casi di insorgenza di correlazione tra regressori ed errore (errori di misura nelle covariate e simultaneità) sono riscontrabili in modelli di cross-section, di serie storiche e di panel. In quest'ultimo caso, in letteratura esistono stimatori corrispondenti agli OLS pooled, LSDV e GLS da applicare quando le variabili esplicative sono correlate con i termini di errore.

In pratica, dopo avere scelto il modello panel (il grado di poolability) con cui trattare l'eterogeneità (pooled completo, FE, RE), se si sospettano problemi di correlazione fra le esplicative ed i termini di errore, si usano i corrispondenti stimatori IV. Il ricercatore deve aggiungere alle ipotesi di specificazione anche l'elenco delle variabili strumentali. Si ricordi che la condizione (necessaria) di identificazione dei parametri è che il numero di strumenti (esogeni) sia uguale o superiore al numero di variabili endogene esplicative (da strumentare).

Il terzo caso di insorgenza di correlazione tra covariate ed errore è quello di modelli dinamici con errori autocorrelati; pertanto, richiede che i dati abbiano dimensione temporale (serie storiche o panel).

Tuttavia, contrariamente alle serie storiche, nel caso panel la dinamica implica l'insorgere di inconsistenza degli stimatori senza richiedere la contemporanea presenza di errori autocorrelati. Tutto ciò è vero sia nel caso di modelli FE, sia per modelli RE. Questo fatto solleva importanti questioni in quanto le relazioni di comportamento spesso implicano componenti dinamiche. L'argomento è trattato nella seguente Section [D.2]. Con le dovute modifiche, i metodi di stima presentati di seguito per modelli dinamici possono essere applicati anche al caso di modelli statici affetti da problemi di correlazione tra covariate e termine di errore.

[D.2] Modelli dinamici con dati panel

In generale, elementi di dinamica nei modelli panel possono essere introdotti (i) ipotizzando che gli errori siano processi AR (dinamica degli errori), (ii) aggiungendo come regressori alcuni ritardi della variabile dipendente (dinamica dell'equazione). Dato che il caso (i) è ritenuto un approccio meno soddisfacente dal punto di vista metodologico (vedi la discussione preliminare sub 1.2), supponiamo che il modello 'vero' sia dinamico:

(7) $y_{it} = \rho\, y_{it-1} + a_i + \varepsilon_{it}$ $\qquad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$

dove a_i sono gli <u>effetti fissi</u>.

<u>Avvertenza</u>: in ciò che segue, ci riferiremo a modelli con effetti one way (solo individuali), in cui l'unica variabile esplicativa è la dipendente ritardata di un periodo. L'estensione a casi più articolati non modifica la sostanza di quanto enunciato.

In questo modello è possibile assumere:

- x è predeterminata o esogena in senso sequenziale: $E(\varepsilon_{it}|x_{is})=0\ \forall\ t \geq s$ and $E(\varepsilon_{it}|x_{is})\neq 0\ \forall\ t < s$, $\forall$ i=1,...,N, $\forall$ t, s=1,...,T. Soltanto le esplicative passate e presenti non sono correlate con gli errori correnti. Infatti, essendo nel modello AR $x_{it} = y_{it-1}$, segue che $E(x_{it}|\varepsilon_{it}) = E(x_{it-1}|\varepsilon_{it}) = 0$, ma $E(x_{it+1}|\varepsilon_{it}) \neq 0$ (e anche $E(x_{it}|\varepsilon_{it-1}) \neq 0$); quindi, la condizione di esogeneità stretta è violata.

La trasformazione within equivale a fare le seguenti elaborazioni (cfr Nickell, 1981, Econometrica).

Sottraggo $y_{i.} = \rho\, y^*_{i.} + a_i + \varepsilon_{i.}$ $\qquad$ (con $y^*_{i.} = T^{-1}\sum_{t=0}^{T-1} y_{it} \neq y_{i.}$) dalla (7)

$\rightarrow (y_{it} - y_{i.}) = \rho\,(y_{it-1} - y^*_{i.}) + (\varepsilon_{it} - \varepsilon_{i.})$ $\qquad$ in cui:

$Cov[(y_{it-1} - y^*_{i.}), (\varepsilon_{it} - \varepsilon_{i.})] = Cov[y_{it-1}, (\varepsilon_{it} - \varepsilon_{i.})] = -Cov(y_{it-1}, \varepsilon_{i.}) = -[\sigma_\varepsilon^2 (1-\rho^{t-1})]/[T(1-\rho)] \neq 0$

[ricorda che $\varepsilon_{i.} = (\varepsilon_{i1} + \varepsilon_{i2} + \varepsilon_{i3} + \ldots + \varepsilon_{iT})/T$]

$\rightarrow$ La trasformazione within crea una correlazione fra esplicative e residui nel modello che scompare solo asintoticamente ($T\rightarrow\infty$) Ricapitolando:

T grande $\rightarrow$ Cov = 0 $\rightarrow$ incorrelazione asintotica

T piccolo $\rightarrow$ Cov $\neq$ 0 $\rightarrow$ distorsione negativa

Nota che i precedenti risultati valgono nonostante nessuna autocorrelazione sia stata ipotizzata per gli ε_{it}

$\rightarrow$ è necessario "strumentare" y_{it-1}, ma come?

La sostanza non cambia quando il modello dinamico presenta <u>effetti casuali</u> $\mu_i \sim IID(0, \sigma_\mu^2)$:

(8) $y_{it} = a + \rho\, y_{it-1} + (\mu_i + \varepsilon_{it})$ $\qquad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$

perché l'esistenza di covarianza non nulla fra le esplicative ed i termini di errore ($\mu_i+\varepsilon_{it}$) la si ottiene per per definizione a causa della presenza dell'effetto individuale μ_i sia in y_{it} sia in y_{it-1}.

[D.2.1] L'approccio IV per panel dinamici: Anderson-Hsiao

La trasformazione in differenze prime delle precedenti equazioni (7) e (8):

(9) $\Delta y_{it} = \rho\, \Delta y_{it-1} + \Delta\varepsilon_{it}$

elimina gli effetti individuali e presenta:

$\Delta y_{it-1} = (\mathbf{y_{it-1}} - y_{it-2})$

$\Delta\varepsilon_{it} = (\varepsilon_{it} - \boldsymbol{\varepsilon_{it-1}})$

$\rightarrow$ esplicative ed errore sono ancora correlati, per definizione, nelle componenti in grassetto.

Tuttavia, qualsiasi variabile datata prima di t-1 è uno strumento valido (se $\varepsilon_{it} \sim$ IID).

Anderson-Hsiao (1981) suggeriscono di utilizzare Δy_{it-2} come strumento per Δy_{it-1} e di stimare ρ nella (7) col metodo IV-2SLS.

Arellano (1989) nota che y_{it-2} (in livelli) è, per definizione, uno strumento *rilevante* (non-debole) perché correlato con l'esplicativa Δy_{it-1} e, allo stesso tempo, *esogeno* perché incorrelato con l'errore $\Delta\varepsilon_{it}$. Inoltre, y_{it-2} permette di risparmiare un'osservazione (guadagno in termini di gradi di libertà).

Se il modello di partenza include variabili esplicative time-invariant (è come aggiungere una componente "b z_i"), queste vengono azzerate dal calcolo delle differenze prime (al pari degli effetti individuali) e il parametro *b* non può essere stimato. Per stimare *b*:

(i) si stima IV il modello in differenze (quindi escludendo *b*);
(ii) si salvano i residui e se ne calcola la media individuale;
(iii) si stima *b* da una regressione cross-section in cui la dipendente sono le medie individuali dei residui sub (ii) e l'esplicativa è z_i .

Questa stima di *b* è consistente per $N\rightarrow\infty$, inconsistente se N è fisso e $T\rightarrow\infty$.

[D.2.2] L'approccio GMM-diff per panel dinamici: Arellano-Bond

Lo stimatore IV dell'equazione in differenze (9) ha alcune pregevoli caratteristiche: è molto semplice da implementare e, soprattutto, è consistente. D'altro canto, non è efficiente perché:

- non usa tutte le possibili condizioni di ortogonalità
- non tiene conto della struttura degli errori $\Delta\varepsilon_{it}$

Una risposta ad entrambi i precedenti punti viene fornita dall'approccio del metodo generalizzato dei momenti (GMM-diff) proposto da Arellano-Bond (1991), uno stimatore per panel dinamici quando T è piccolo e N è (molto) grande. Nei casi in cui la dimensione temporale è elevata, il semplice approccio IV (solo consistente) di Anderson-Hsiao [Cfr D.2.1] costituisce una semplice e attraente alternativa.

Arellano-Bond (1991) notano che addizionali strumenti possono essere ottenuti utilizzando le condizioni di ortogonalità fra la dipendente ritardata e i disturbi ε_{it} . Il loro approccio viene illustrato dal semplice modello AR(1):

$$y_{it} = \rho\, y_{it-1} + v_{it} \qquad \text{con: } v_{it} = \mu_i + \varepsilon_{it}$$

dove: $\mu_i \sim \text{IID}(0, \sigma_\mu^2)$ e $\varepsilon_{it} \sim \text{IID}(0, \sigma_\varepsilon^2)$. Allo scopo di ottenere uno stimatore consistente di ρ per T fisso (piccolo) e $N\rightarrow\infty$ si calcolano le differenze prime per eliminare gli effetti individuali μ_i (il risultato è quello già riportato nella precedente (9):

(10) $$\Delta y_{it} = \rho\, \Delta y_{it-1} + \Delta\varepsilon_{it}$$

Si noti che l'errore della (10) è un modello MA(1) non invertibile. Se le osservazioni iniziano in t=1, il primo periodo in cui la relazione (10) può essere osservata è in t=3. La Tabella 3 sintetizza gli strumenti utilizzabili per ogni osservazione.

Tab. 3 – Gli strumenti utilizzabili

t =	equazione (10):	strumenti:
3	$y_{i3} - y_{i2} = \rho\,(y_{i2} - y_{i1}) + \varepsilon_{i3} - \varepsilon_{i2}$	y_{i1}
4	$y_{i4} - y_{i3} = \rho\,(y_{i3} - y_{i2}) + \varepsilon_{i4} - \varepsilon_{i3}$	y_{i1} y_{i2}
…	…	…
T	$y_{iT} - y_{iT-1} = \rho\,(y_{iT-1} - y_{iT-2}) + \varepsilon_{iT} - \varepsilon_{iT-1}$	y_{i1} y_{i2} … $y_{i\,T-3}$ $y_{i\,T-2}$

I precedenti strumenti possono essere organizzati nella seguente matrice W_i di dimensione $(T\text{-}2)\times[(T\text{-}1)(T\text{-}2)/2]$:

$$W_i = \begin{bmatrix} y_{i1} & 0 \;\; 0 & \dots & 0 \;\; 0 \;\dots\; 0 \;\; 0 \\ 0 & y_{i1} \;\; y_{i2} & \dots & 0 \;\; 0 \;\dots\; 0 \;\; 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 \;\; 0 & \dots & y_{i1} \;\; y_{i2} \;\dots\; y_{i\,T-3} \;\; y_{i\,T-2} \end{bmatrix}$$

da cui, sovrapponendo le N matrici W_i , si ottiene W (di N(T-2) righe) che rappresenta l'insieme degli strumenti per tutte le NT osservazioni:

$$W = (W_1' , \dots , W_N')'$$

Una procedura IV non tiene conto della particolare struttura della matrice di varianze-covarianze di $\Delta\varepsilon_{it}$:

$$E(\Delta\varepsilon_i\, \Delta\varepsilon_i') = \sigma_\varepsilon^2\, (I_N \otimes G) = \Phi$$

dove: $\Delta\varepsilon_i' = (\varepsilon_{i3}\text{-}\varepsilon_{i2} , \varepsilon_{i4}\text{-}\varepsilon_{i3} , \dots , \varepsilon_{iT}\text{-}\varepsilon_{iT-1})'$, e:

$$G = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 & -1 & 0 \\ 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$

G è una matrice $(T\text{-}2)\times(T\text{-}2)$ la cui particolare forma dipende dal fatto che $\Delta\varepsilon_i$ è un processo MA(1) non invertibile.
Le condizioni di ortogonalità (equazioni dei momenti) fra gli strumenti ed i termini di errore che abbiamo descritto sono: $E(W_i'\Delta\varepsilon_i) = 0$ con i = 1, 2, …, N. Il vec della (10) è premoltiplicato per W′ allo scopo di imporre le precedenti condizioni di ortogonalità:

(11) $$W'\Delta y = W'(\Delta y_{-1})\rho + W'\Delta\varepsilon$$

Per applicare lo stimatore GLS, si definisce la matrice di pesi
$\Omega^{-1} = [W'\Phi W]^{-1}$ (→ feasible perché la sua struttura è nota) e si ottiene lo stimatore (preliminare e consistente) "one-step" di Arellano-Bond:

$$\begin{aligned}\hat{\rho}_{GLS} &= \{(\Delta y_{-1})'W\Omega^{-1}W'(\Delta y_{-1})\}^{-1}\{(\Delta y_{-1})'W\Omega^{-1}W'\Delta y\} \\ &= \{(\Delta y_{-1})'W[W'(I_N \otimes G)W]^{-1}W'(\Delta y_{-1})\}^{-1} \\ &\quad \times \{(\Delta y_{-1})'W[W'(I_N \otimes G)W]^{-1}W'\Delta y\}\end{aligned}$$

$\hat{\rho}_{GLS}$ non è però ottimale. Lo stimatore ottimale di ρ lo si ottiene nel seguente modo. Dai residui della stima "one-step" $\Delta\hat{\varepsilon}_{GLS}$ è possibile ottenere una stima della matrice delle varianze-covarianze:

$$V_{N,GLS} = \sum_{i=0}^{N} W_i'(\Delta\hat{\varepsilon}_{GLS,i})(\Delta\hat{\varepsilon}_{GLS,i})'W_i$$

da utilizzare al posto di Ω per ottenere lo stimatore "two-step" che realizza lo stimatore GMM di Hansen (!982):

$$\hat{\rho}_{GMM} = \{(\Delta y_{-1})'W[V_{N,GLS}]^{-1}W'(\Delta y_{-1})\}^{-1}\{(\Delta y_{-1})'W[V_{N,GLS}]^{-1}W'\Delta y\}$$

I due stimatori $\hat{\rho}_{GLS}$ e $\hat{\rho}_{GMM}$ sono asintoticamente equivalenti se vale la condizione: $\varepsilon_{it} \sim IID(0,\sigma_\varepsilon^2)$.

La varianza asintotica è:

$a\text{Var}(\hat{\rho}_{GMM}) = \{(\Delta y_{-1})'W[V_{N,GMM}]^{-1}W'(\Delta y_{-1})\}^{-1}$

Lo stimatore GMM non richiede conoscenze a priori su: condizioni iniziali e distribuzione di ε_{it} e μ_i

In uno studio che confronta la performance di alternativi stimatori per modelli panel, Judson-Owen (1999) trovano che:

- quando T è grande, il calcolo della stima GMM è di difficile attuazione;
- lo stimatore GMM "one step" funziona meglio del "two step";
- riducendo arbitrariamente la dimensione (il numero di ritardi) della matrice degli strumenti W non peggiora sostanzialmente la performance dei GMM.

Infine, Judson-Owen effettuano un confronto, per panel dinamici, degli stimatori di Anderson-Hsiao (AH), di Arellano-Bond ("one-step", GMM1) e i LSDV la cui distorsione è corretta secondo le indicazioni di Kiviet, 1995 (LSDVC). I loro risultati sono sintetizzati in Tabella 4.

Tab. 4 – I metodi raccomandati dall'esperimento

panel:	T≤10	T=20	T=30
bilanciato	LSDVC	LSDVC	LSDVC
non bilanciato	GMM1	GMM o AH	LSDV

→ suggerimento: non c'è "il metodo" che funziona al meglio in qualsiasi situazione ....

[D.2.3] L'approccio GMM-sys per panel dinamici: Blundell-Bond

L'ottimalità della trasformazione in differenze strumentata con ritardi nei livelli viene però meno se l'esplicativa del modello è una variabile integrata. Infatti, in generale, se l'esplicativa x_{it} è I(1), il coefficiente di correlazione fra Δx_{it} e x_{it-1} è quasi nullo perché le differenze prime di una variabile integrata, per definizione, *non possono* essere correlate con i suoi livelli (intuizione: si pensi alla specificazione del modello univariato per effettuare il test di Dickey-Fuller); mentre Δx_{it} e Δx_{it-1} *possono* essere correlate, a meno che x_{it} sia random walk. Infine, si sottolinea che il coefficiente di correlazione fra x_{it} e x_{it-1} sarà sempre quasi unitario per qualsiasi x_{it} integrato.

Per questo motivo, Arellano-Bover (1995) e Blundell-Bond (1998) propongono un approccio GMM-sys in cui, accanto alle equazioni originali in differenze strumentate con livelli (tipiche dell'approccio GMM-diff) aggiungono anche equazioni in livelli strumentate con le differenze. In altri termini, all'interno del sistema "diff"&"lev" sono strumentate con ritardi di "lev"&"diff".

Pro: nel complesso, GMM-sys si adatta a qualsiasi tipologia statistica
contro: strumenti deboli, poca identificazione, distorsione in campioni finiti

☺A: LECTURE_PANEL_DYNAMIC_APPLICATION

[E] Non stazionarietà e dinamica nei panel

Negli anni recenti, lo studio dell'interrelazione fra test di radici unitarie e dati panel si è sviluppata in modo considerevole collegando:

(i) l'area di studi della integrazione e della cointegrazione per serie storiche;
(ii) l'econometria dei panel.

Scopo dichiarato: combinare l'informazione proveniente dalle serie storiche con quella dalle cross-section, nella speranza che l'inferenza su radici unitarie e cointegrazione possa essere più precisa, grazie alla dimensione individuale.

Gli sforzi lungo questa direzione sono di solito giustificati dalla bassa potenza dei test per serie storiche. Tutto ciò accade specialmente quando le serie storiche sono brevi, ma c'è la possibilità di raccogliere informazioni lungo la dimensione cross-section (paesi, regioni, imprese e settori).

Un altro filone di ricerca (dinamica ed eterogeneità) prende le mosse da Pesaran-Shin (1995) che dimostrano l'inconsistenza degli stimatori di panel eterogenei dinamici con dati aggregati oppure pooled e raccomandano l'uso di stimatori "group-mean".

[E.1] Test di integrazione

Fin dall'apparizione dei papers di **Levin-Lin** (1992, 1993), d'ora in poi LL, i test panel di radici unitarie sono diventati molto popolari presso gli economisti applicati.

Oramai è dato per acquisito che i test di Dickey-Fuller (DF) e ADF sono poco potenti nel discriminare fra processi integrati ed alternative stazionarie, e che l'uso di dati panel è un modo per incrementare la potenza dei test di radici unitarie per serie storiche.

La potenza del test non è però tutto: si tenga presente che le ipotesi H_0 e H_1 differiscono nei due tipi di test:

ipotesi:	serie storiche	Panel, LL
H_0	$\pi=0$	$\pi_i=0\ \forall i$
H_1	$\pi<0$	$\pi_i<0\ \forall i$

→ i due test (serie storiche e panel) non rispondono alla stessa domanda e, perciò, non è detto che siano interscambiabili per decidere sulla questione di interesse.

Il primo modello ad affrontare in modo articolato il tema delle radici unitarie nei panel è di LL:

(12) $$\Delta y_{it} = a_i + b_i\, t + \tau_t + \pi\, y_{it-1} + \varepsilon_{it}$$

in cui si dà conto dell'esistenza di effetti fissi (a_i); effetti temporali specifici (b_i) e generici (τ_t). Inizialmente LL suppongono che $\varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2)$.

A partire dalla (12), LL ricavano un vasto numero di sottomodelli, il cui tratto comune è che sono tutti stimati OLS pooled.

La presenza di effetti fissi ha un ruolo importante nello spiegare l'eterogeneità, visto che il coefficiente π è vincolato ad essere lo stesso per tutti gli individui.

Il test LL verifica: H_0: $\pi = 0$ contro H_1: $\pi < 0$; con ulteriori ipotesi sotto H_0 e H_1 riferite ai coefficienti (a_i b_i τ_t) delle componenti deterministiche (ad esempio, nel modello dove $b_i=\tau_t=0$, si verifica H_0: $\pi=0$ e $a_i=0\ \forall i$; contro H_1: $\pi< 0$ e a_i non vincolati).

Alcune estensioni in Levin-Lin (1993) permettono una struttura più generale dell'errore (eteroschedasticità e autocorrelzione), ma mantengono l'ipotesi di indipendenza cross-section fra gli individui (ipotesi che talvolta crea problemi ai test, come vedremo).

Il limite principale del test LL è che il parametro π è lo stesso per tutte le osservazioni. Ciò che convince meno sono le sue implicazioni sotto l'alternativa.

Ad esempio, per verificare col test LL l'ipotesi di convergenza della crescita di N paesi, H_0 prevede che nessun paese converga (π=0 per tutti i paesi), mentre sotto H_1 tutti i paesi convergono alla stessa velocità.

I lavori di LL rivestono una fondamentale importanza perché introducono i principali temi dibattuti nella letteratura sulle radici unitarie, cioè:

(a) dimostrano che la normalità asintotica degli stimatori è legata ad opportune trasformazioni e correzioni;
(b) sottolineano la necessità di considerare casi alternativi di tendenza all'infinito di T e N;
(c) sollevano il tema "omogeneità vs eterogeneità" sotto le ipotesi nulla e alternativa (i loro modelli sono fra i più restrittivi perché richiedono che π sia lo stesso per tutti gli individui del campione);
(d) fondano le loro analisi nell'ipotesi di indipendenza fra gli individui;
(e) introducono modifiche dei test più elementari per dare conto di residui non sferici ed endogenità dei regressori.

I punti (a)-(b) costituiscono una importante base di partenza per nuovi studi. Gli altri tre temi sono stati ripresi da alcuni autori che hanno proposto importanti estensioni.

Il test di **Im-Pesaran-Shin** (1997) (IPS) è una estensione delle ipotesi relative ai casi (c)-(d) di LL, molto restrittivi. Il punto (e) è importante nel contesto dei test di cointegrazione per panel (si veda la [E.2]).

Il test di IPS estende il test LL per dare conto di eterogeneità dei coefficienti π_i :

$$\Delta y_{it} = a_i + \pi_i\, y_{it-1} + \varepsilon_{it}$$

Si effettuano N test di significatività di π di tipo DF-ADF individuo per individuo ($t_{i,T}$). Mediante esperimenti di Monte Carlo per panel bilanciati, IPS tabulano i momenti primo E($t_{i,T}$) e secondo V($t_{i,T}$) delle statistiche $t_{i,T}$, e dimostrano che:

$$\sqrt{N}\,\frac{\bar{t}_{N,T} - E(t_{i,T})}{\sqrt{V(t_{i,T})}} \Rightarrow \text{Normal}(0,1) \text{ con: } \bar{t}_{N,T} = \frac{1}{N}\sum_{i=1}^{N} t_{i,T}$$

In pratica, il test IPS rappresenta un modo di combinare l'evidenza (eterogenea) di N test di radici unitarie calcolate per N individui. I momenti primo e secondo di $\bar{t}_{i,T}$ sono tabulati ad hoc.

Sotto l'ipotesi nulla, tutte le N serie storiche individuali hanno una radice unitaria:

$$\pi_i = 0 \qquad \forall i$$

mentre, sotto l'ipotesi alternativa, gli individui sono raggruppati in due insiemi complementari: N_S di processi non stazionari e N_{1-S} di processi stazionari:

$$\pi_i = 0 \qquad \forall i \in N_S$$
$$\pi_i < 0 \qquad \forall i \in N_{1-S}$$

tali che: $\lim_{N\to\infty} (N_{1-S} / N) = \delta$, con $0 < \delta \leq 1$

sulla base della quale si possono riformulate le ipotesi:

H_0: $\delta = 0$
H_1: $\delta > 0$

"The main failing of panel unit root tests is that the null hypothesis (usually that *all* the series in the panel are realisations of unit root processes) will be violated even if a very small number of series in the panel are stationary, and rejection cannot help us calculate the entity of that number." (Taylor and Sarno, 1998, p. 283).

Il problema dell'eterogeneità dei panel (Cfr Pesaran-Smith, 1995) è forse risolto nella fase di stima, ma non nelle fasi di inferenza.

Alcuni fenomeni (ad esempio le analisi cross-country), caratterizzati da eterogeneità, introducono una asimmetria fra H_0 e H_1 non presente nei modelli per serie storiche: la stessa H_0 è imposta a tutte le i , mentre H_1 può variare lungo la i.

In panel eterogenei con N elevato e T ≈ 15 è possibile formulare test informativi in "senso medio": per inferire se H_0 può essere rifiutata per una significativa frazione di individui. Una misura (stima) per δ richiede invece T elevato, ma in tal caso l'esigenza di pooling sarebbe meno sentita.

IPS dimostrano, con un esperimento di Monte Carlo, la migliore performance in campioni finiti del loro test rispetto a quello di LL.

☺A: EXAMPLE FROM UNIT ROOTS TESTS APPLIED TO CORPORATE FINANCE FOR ITALIAN MANUFACTURING COMPANIES

Bontempi-Golinelli (2005) è un esempio di applicazione dei test di radici unitarie a dati panel e si affrontano alcuni dei punti discussi mediante l'approfondimento del tema dell'eterogeneità.

[E.2] Test di cointegrazione

Un filone di analisi di cointegrazione per dati panel è stato sviluppato mediante test che, sotto l'ipotesi nulla, hanno l'assenza di cointegrazione. Tali test sono l'analogo panel del test di Engle-Granger (1987) per serie storiche, perché usano i residui di regressioni statiche per costruire test di radici unitarie le cui distribuzioni sono tabulate ad hoc.

Il test di cointegrazione di **Pedroni** (1999) si fonda su una regressione statica:

(13) $$y_{it} = a_i + b_i x_{it} + c_i t + \varepsilon_{it}$$

che ammette una considerevole eterogeneità nel panel: pendenze eterogenee, effetti fissi e trend individuali sono tutti possibili. Le variabili esplicative possono essere più di una e non serve ipotizzare la loro esogenità. Infine, si impone indipendenza cross-section, a meno di comuni disturbi aggregati.

L'approccio di Pedroni prevede sette modelli alternativi a partire dalla (13) che possono essere classificati in due categorie: (i) quattro modelli sono sviluppati nell'ambito del pooling lungo la dimensione within (i primi tre sono di tipo Phillips-Perron, PP, il quarto di tipo ADF), (ii) tre lungo la dimensione betwen (due PP e uno ADF).

Se definiamo con ρ_i il coefficiente autoregressivo dei residui OLS dell'i-esimo individuo, la categoria (i) di test, definiti "panel statistic" verifica le ipotesi:

H_0: $\rho_i = 1 \quad \forall i$; H_1: $\rho_i = \rho < 1 \quad \forall i$

la categoria (ii) di test "group statistic" verifica:

H_0: $\rho_i = 1 \quad \forall i$; H_1: $\rho_i < 1 \quad \forall i$

L'analogia con i test LL nel caso (i) e IPS nel caso (ii) è evidente, in termini di eterogeneità consentita sotto l'ipotesi alternativa.

Concentriamo l'attenzione sulla procedura di calcolo della "group t-statistic", la terza della categoria (ii), considerata una estensione del test di IPS:

passo 1: stima OLS della (13) per ogni individuo → $\hat{\varepsilon}_{OLS,it}$

passo 2: calcolo il test individuale di tipo ADF usando $\hat{\varepsilon}_{OLS,it}$

passo 3: media semplice delle statistiche ADF per individuo

Calcolata una delle sette statistiche test $P_{N,T}$, bisogna standardizzarla mediante: $(P_{N,T} - \mu N^{½})/\nu^{½} \Rightarrow N(0,1)$

I momenti primo μ e secondo ν di $P_{N,T}$, ottenuti mediante simulazione, sono tabulati ad hoc e variano: (a) per statistica, (b) per numero di esplicative, (c) per nucleo deterministico.

Nota: l'esito dei test di integrazione interagisce con le proprietà di cointegrazione fra individui nel panel (Cfr Banerjee-Marcellino, 2000): la violazione dell'ipotesi di incorrelazione cross-section (cointegrazione fra alcuni i e j) implica size empirici molto maggiori di quelli nominali.
$\rightarrow$ quando vera, la H_0 di radici unitarie è rifiutata troppo spesso.

[E.3] Dinamica, cointegrazione e poolability

Quando la dimensione di T è elevata e dello stesso ordine di grandezza di N, può essere utile concentrare l'attenzione sui legami di lungo periodo e sulla velocità di aggiustamento a tali relazioni.

I parametri dei modelli per dati panel di questo tipo possono essere analizzati mediante 2 procedure (ai due estremi):

- Averaging: stima di equazioni separate per individuo (group) ed esame della distribuzione dei coefficienti fra individui. Lo stimatore "mean group" (MGE) è una stima consistente della media dei parametri (eterogenei).
- Pooling: utilizzo di un modello panel (con effetti fissi o random, FE o RE), dove le pendenze e le varianze dell'errore sono ipotizzate uguali.

Il lavoro di **Pesaran-Shin-Smith** (1999, JASA) considera uno stadio intermedio fra i due precedenti: gli stimatori "pooled mean group" (PMGE) che, come vedremo, implicano averaging e pooling.

Lo stimatore PMGE si applica a modelli che prevedono diverse (da individuo a individuo) intercette, parametri di breve periodo e varianza dell'errore, ma che vincolano i coefficienti di lungo periodo ad essere gli stessi.

Nota: la non imposizione di vincoli sui parametri di breve permette diverse specificazioni dinamiche per individuo.

I vantaggi principali dell'approccio di Pesaran-Shin-Smith sono:

- stima legami comuni di lungo periodo senza imporre implausibili dinamiche comuni;
- si applica al caso di regressori sia stazionari sia non stazionari;
- permette di discriminare fra stimatori MGE e PMGE.

Esempio: Lo studio della domanda di moneta nell'area dell'euro (Golinelli-Pastorello, 2002, EJF). Variabili di interesse m (moneta), y (PIL), r (tasso di interesse a lungo).

Approccio: stima di dati aggregati per l'area e per paese.

L'approccio ARDL(p,p) a livello di area (Pesaran and Shin, 1998):

$$\Delta m_t = \mu + \phi m_{t-1} + \pi' x_t + \sum_{i=1}^{p-1} \lambda_i^* \Delta m_{t-i} + \sum_{i=1}^{p-1} \psi_i^{*\prime} \Delta x_{t-i} + \varepsilon_t$$

dove: μ, ϕ, π, λ e ψ sono parametri; $x_t = (y_t, r_t)'$ e ε_t è l'errore. Sotto ipotesi standard per ε_t e se $\phi < 0 \rightarrow$ esiste una relazione stabile di lungo periodo fra m_t e x_t, e la si definisce con:

$$m_t = \theta' x_t + \eta_t$$

dove $\theta = -\pi/\phi$ e $\eta_t \sim I(0)$.

Our ARDL(2,2) model for m, y and r can be simplified in the final model for the Euro area:

$$\Delta m_t = \underset{(0.04)}{-0.16} + \underset{(0.08)}{0.52}\, \Delta m_{t-1} - \underset{(0.10)}{0.31}\, \Delta r_{t-1} - \underset{(0.03)}{0.11}\, (m - \underset{(0.06)}{1.37}\, y + \underset{(0.37)}{0.68}\, r)_{t-1}$$

Main characteristics of the model:

- very parsimonious model
- data congruent
- stable parameter estimates
- forecasting ability 98q1-99q3: p-value 20.1%

Le stesse funzioni di domanda di moneta possono essere studiate per paese (**MGE approach**), specificando un modello ARDL per ogni paese (h) con *m* variabile dipendente e *y* e *r* variabili esplicative.

The results are in Golinelli-Pastorello (2002, Table 6, p. 390). The comparison between the results from aggregate and disaggregated models shows the better properties of the first. The main findings are:

- the regression standard error of the national models is at least double w.r.t the area model;
- the diagnostic tests for national models signal some specification problems;
- a long run relationship is hardly ever detected;
- almost always the estimates of long run elasticity of income are bigger than 1, while the elasticity to the long rate are much more dispersed (they spread from –4 to 5).

This worsening in performance when shifting from the area to the national model can depend on two causes:

- specification problems at country level (our national models are too simplified);
- the disturbing effect of shocks and outliers on the estimation of national money demands (sampling error).

→ Nel secondo dei due casi, è possibile effettuare il "pool" dei parametri di lungo periodo di individui diversi, allo scopo di migliorare la performance del modello.

Seguendo l'approccio di Pesaran, Shin and Smith, è possibile verificare la poolability dei parametri di lungo periodo (long run parameter homogeneity test),

H_0: $\theta_h = \theta \quad \forall h$ ipotesi di poolability

Le stime θ sono ottenute:

- calcolando le medie semplici per paese (MGE, Mean Group Estimator)
- imponendo il vincolo di pooling (PMGE, Pooled Mean Group Estimate).

Sotto la nulla di poolability, entrambi gli stimatori sono consistenti, ma PMGE è anche efficiente.

L'ipotesi nulla può essere verificata mediante due approcci:

- una statistica-test di tipo Hausman (in piccoli campioni non è detto che la differenza fra varianze sia positiva definita);
- un test del rapporto di verosimiglianza, di confronto fra la log-likelihood vincolata (sotto l'ipotesi di pooling) e non vincolata (la somma delle log-likelihood nei modelli per individuo); in piccoli campioni questo test tende a rifiutare in eccesso la nulla.

We made pooled estimates for <u>4 groups</u> of countries:

1. Euro area without Luxembourg;
2. Core area (Austria, Belgium, France, Germany, Netherlands);
3. Core plus Italy;
4. Core plus Spain.

The results can be summarised as follows:

1. Euro area without Luxembourg:
- long run parameter poolability Hausman test is not rejected; the likelihood ratio test is rejected;
- previous results can be due to the huge dispersion of the unrestricted estimates;
- the fitting of the pooled model is only slightly worse than the unrestricted, and we have no worsening in diagnostic tests;

2. Core area
- long run parameter poolability Hausman test can not be computed; the likelihood ratio test is not 5% rejected;
- the fitting of the pooled model is broadly the same as the unrestricted;

3. and 4. Core + Italy, and Core + Spain
- the purpose is not to verify the possibility of coexistence of Italy and Spain in a monetary union with the countries of the "core";
- the Hausman test can not be computed; the likelihood ratio test 5% rejects the poolability in the core + Italy subset (but does not reject at 1%) and does not 5% reject in the core + Spain subset.

Evidentemente, l'approccio PMGE in parte conferma gli esiti aggregati (confronto solo qualitativo) e in parte fornisce indicazioni sul grado di omogeneità dei singoli paesi/individui (confronto mediante test).

General references

Arellano, Manuel, *Panel Data Econometrics*, Oxford University Press, 2003.
Baltagi, Badi H., *Econometric Analysis of Panel Data*, John Wiley & Sons, 2005.
Hsiao, Cheng, *Analysis of Panel Data*, Cambridge University Press, 2003.
Manera, Matteo and Galeotti Marzio, *Microeconometria. Metodi ed applicazioni*, Carocci Editore, 2005.
Sevestre Patrick, *Économétrie des Données de Panel*, Dunod, Paris, 2002.
Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.

Relevant review-articles:

Handbook of Econometrics:
Capitolo 22, *Panel Data*, di Gary Chamberlin.
Capitolo 29, Econometric Analysis of Longitudinal Data, di Heckman e Singer.

Interesting:

Matyas, L. and P. Sevestre (eds.), *The Econometrics of Panel Data: A Handbook of Theory and Applications*, Kluwer Academic Publishers, 1996.
Maddala, G.S. (ed.), *The Econometrics of Panel Data*, Vol. I e II, Edward Elgar Publishing, Cheltenham, 1993.
Baltagi, Badi H. and Baldev Raj (eds.), *Panel Data Analysis*, Physica-Verlag, Heidelberg, 1992.
Dielman, T.E., *Pooled Cross-Sectional and Time-Series Data Analysis*, Marcel Dekker, 1989.

Good chapters on panel:

Greene, *Econometric Analysis*, Prentice Hall, 1997, ch. 14.
Wooldridge, Jeffrey M., *Introductory Econometrics. A modern approach*, Thomson South-Western, 2003, ch. 13, 14.
Hayashi, Fumio, *Econometrics*, Princeton University Press, 2000, ch. 5.

Ruud, Paul A., *An Introduction to Classical Econometric Theory*, Oxford University Press, 2000, ch. 24.

Basic:

Judge, Griffiths, Hill, et. al., *The Theory and Practice of Econometrics*, Wiley, 1985.
Johnston, Di Nardo, *Econometric Methods*, McGraw-Hill, 1998.
Davidson, Russell and MacKinnon James G., *Estimation and Inference in Econometrics*, Oxford University Press, 1992.

Applications:

Baltagi-Raj (1992): Parte III, e Matyas-Sevestre (1996): Parte III.

Two special issues of the **Journal of Econometrics**: Carraro, Peracchi e Weber (ed.), 1993, Vol. 59, pp. 1-211; and Baltagi (ed.), 1995, Vol. 68, pp. 1-268.