Panel Data

Giulio Palomba

Agosto 2008

I dati in formato panel combinano le informazioni relative alle caratteristiche di N individui nello stesso istante temporale con quelle rilevate per gli stessi individui in T diversi periodi di tempo. Nei modelli di tipo panel i dati disponibili hanno perciò entrambe le caratteristiche di

- Dati Cross Section: per un dato istante sono osservate le caratteristiche di più individui,
- Dati Time Series: per un dato collettivo di individui sono rilevate le diverse caratteristiche in diversi istanti[1].

La seguente matrice mostra la disposizione dei dati in formato panel relativi ad una variabile Y; ogni colonna si riferisce ad un diverso individuo per cui la variabile è stata rilevata, mentre per riga sono disposte le diverse osservazioni nel tempo. Ovviamente la variabile Y è composta di NT osservazioni.

$$\underset{(N\times T)}{Y} = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{i1} & \cdots & y_{N1} \\ y_{12} & y_{22} & \cdots & y_{i2} & \cdots & y_{N2} \\ \vdots & \vdots & & \vdots & & \vdots \\ y_{1t} & y_{2t} & \cdots & y_{it} & \cdots & y_{Nt} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ y_{1T} & y_{2T} & \cdots & y_{iT} & \cdots & y_{NT} \end{bmatrix} \tag{1}$$

Poiché i dati cross section e quelli time series hanno ciascuno le proprie peculiarità, essi portano con sé tutte le complicazioni soprattutto per quanto riguarda il venir meno di alcune ipotesi classiche del modello di regressione lineare

$$Y = X\beta + \varepsilon. \tag{2}$$

Attraverso la (2) è possibile introdurre la notazione. Il vettore Y di dimensione $(NT \times 1)$ è ottenuto applicando l'operatore *vec* alla matrice (1) e rappresenta la variabile dipendente, la matrice dei regressori X ha dimensione $(NT \times k)$, mentre il vettore k-dimensionale β contiene i parametri incogniti da stimare. Il termine di disturbo ε ha le stesse dimensioni della variabile dipendente.

In molti testi spesso è utilizzata una notazione meno compatta rispetto all'equazione (2): molto spesso i modelli per i dati panel vengono presentati nella formulazione che tiene conto della singola osservazione, quindi l'equazione del modello lineare di regressione diventa

$$y_{it} = x'_{it}\beta + \varepsilon_{it}, \tag{3}$$

dove tutte le variabili si riferiscono all'osservazione relativa all'i-esimo individuo nell'istante t-esimo di tempo; in questo contesto y_{it} e ε_{it} sono scalari, mentre la matrice dei regressori è data da un vettore riga con k componenti.

A volte può capitare di imbattersi in una notazione che accorpa tutte le osservazioni relative all'i-esimo individuo per il quale vengono rilevate T osservazioni. L'equazione che ne scaturisce è perciò la seguente:

$$\underset{(T\times 1)}{y_i} = \underset{(T\times k)}{x_i}\underset{(k\times 1)}{\beta} + \underset{(T\times 1)}{\varepsilon_i},$$

[1] Talvolta i termini "cross section" e "time series" sono tradotti rispettivamente con "cross-sezionali" e "serie storiche".

Nelle pagine che seguiranno, salvo alcune eccezioni, sarà utilizzata la notazione compatta introdotta nell'equazione (2).

La matrice delle varianze e delle covarianze del termine di errore del modello panel è quadrata, simmetrica ed ha dimensione $(NT \times NT)$. Essa è definita come

$$\Omega = Var(\varepsilon) = E(\varepsilon\varepsilon')$$

La convenienza dell'utilizzo dei modelli di tipo panel risiede soprattutto nel guadagno di efficienza della stima perché il maggior numero di osservazioni che si ha rispetto alla sola dimensione cross section o time series genera uno stimatore con variannza più piccola.

1 Modelli per serie storiche pooled

Questa sezione consiste in una rassegna dei principali modelli di regressione lineare per serie storiche pooled man mano che le ipotesi classiche si fanno sempre meno stringenti.

Le serie storiche pooled consistono in una combinazione di pochi individui osservati attraverso un campione di T osservazioni ritenuto sufficientemente ampio da consentire regressioni di tipo time series per ciascun individuo. Questo tipo di modelli permette l'ottenimento di stime più efficienti rispetto al caso delle singole regressioni perché utilizza un set informativo maggiore dovuto alla presenza di più individui.

1.1 Modello lineare classico

Data l'equazione (2), devono essere rispettate le ipotesi classiche

1. $E(\varepsilon \mid X) = 0$,
2. La matrice X ha rango pieno pari a k,
3. $E(X'\varepsilon) = 0$,
4. $\Omega = Var(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_{NT}$: quest'ultima ipotesi (di omoschedasticità) implicitamente assume che
   (a) la varianza di ciascuna osservazione σ^2_{it} è costante per $\forall\ i$ e $\forall\ t$,
   (b) $E(\varepsilon_{it}\varepsilon_{is})$ per ogni $t \neq s$, cioè non c'è correlazione tra le osservazioni relative allo stesso individuo in istanti diversi,
   (c) $E(\varepsilon_{it}\varepsilon_{jt})$ per ogni $i \neq j$, cioè non c'è correlazione istantanea tra le osservazioni relative ad individui diversi.

Sotto queste condizioni lo stimatore OLS risulta essere non distorto, consistente, BLUE.

1.2 Modello con eteroschedasticità pura

Rispetto al modello lineare classico di cui sopra viene rimossa l'ipotesi per la quale la varianza è costante lungo la diagonale principale della matrice Ω. In particolare, si assume che ciascun individuo all'interno del campione conserva l'ipotesi di omoschedasticità nel periodo di tempo considerato, ma può presentare una varianza differente rispetto agli altri individui.

L'eteroschedasticità pura si configura perciò come una situazione in cui la matrice Ω resta diagonale, ma con varianze che variano ogni T osservazioni. Analiticamente si ha

$$\Omega = \begin{bmatrix} \sigma_1^2 I_T & 0 & \dots & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_T & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & & \vdots \\ 0 & 0 & & \sigma_i^2 I_T & & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \sigma_N^2 I_T \end{bmatrix}. \tag{4}$$

La presenza di eteroschedasticità pura è condizione necessaria affinché si utilizzi lo stimatore GLS

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{5}$$

Ovviamente, data la forma diagonale di Ω, lo stimatore GLS in pratica è uno stimatore dei minimi quadrati ponderati (stimatore WLS) in quanto può essere ottenuto attraverso la regressione OLS di $\Omega^{-1/2}Y$ su $\Omega^{-1/2}X$, con $\Omega^{-1/2}$ matrice diagonale i cui elementi (pesi) sono dati da N sequenze di lunghezza T con valori pari a $1/\sigma_i$.

Poiché gli N parametri σ_i^2 non sono noti, occorre una loro stima consistente. La soluzione a questo problema risiede in due strade alternative e non equivalenti:

- si stima un modello OLS su tutte le NT osservazioni, si salvano i residui $\hat{\varepsilon}$ (vettore di dimensione NT),
- si stimano N regressioni del tipo

$$\underset{(T\times 1)}{y_i} = \underset{(T\times k)}{x_i}\underset{(k\times 1)}{\beta_i} + \underset{(T\times 1)}{\varepsilon_i}.$$

In entrambi i casi, per ciascuno degli N individui, si calcola la statistica

$$\hat{\sigma}_i = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_i}{T-k}.$$

Naturalmente, una volta ottenuta la stima $\hat{\Omega}$, lo stimatore (5) diviene "feasible" (FGLS) con le usuali proprietà di non distorsione e consistenza. Inoltre, per $T \to \infty$, esso risulta asintoticamente efficiente.

1.3 Modello con eteroschedasticità pura e correlazione tra individui

Rispetto all'approccio precedente viene rimossa l'assunzione di incorrelazione contemporanea tra gli individui. In pratica si ha

$$E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}^2$$

per ogni i e j, quindi la matrice delle varianze e delle covarianze del termine di disturbo diventa

$$\Omega = \Sigma \otimes I_T \tag{6}$$

dove

$$\underset{(N\times N)}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \dots & \sigma_{NN} \end{bmatrix}.$$

La struttura della matrice Ω di fatto consiste nell'accostamento di N^2 matrici diagonali quadrate di dimensione $T \times T$, struttura coerente con il modello "Seemingly Related Regression" (SUR) dato da

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} &= \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \\ Y &= (\underset{(T\times k)}{X_i} \otimes I_N)\underset{(Nk\times 1)}{\beta} + \varepsilon, \end{aligned}$$

sotto l'ipotesi che $\beta_1 = \beta_2 = \ldots = \beta_N$ (quindi in tutto k parametri da stimare). Dato che la matrice Ω non rispetta l'ipotesi di omoschedasticità, anche in questo caso lo stimatore FGLS risulta essere il più appropriato e le covarianze stimate $\hat{\sigma}_{ij}$ possono essere ottenut attraverso i due metodi introdotto nel precedente paragrafo.

Una volta ottenuta $\hat{\Sigma}$, quindi $\hat{\Omega} = \hat{\Sigma} \otimes I_T$, lo stimatore FGLS diventa

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y \tag{7}$$

con $X = (X_i \otimes I_N)$. Lo stimatore FGLS è non distorto, consistente ed asintoticamente efficiente per $T \to \infty$, dato N.

1.4 Modello con eteroschedasticità e correlazioni pure

In questo caso sono le correlazioni ad essere pure e non l'eteroschedasticità: ciò significa che la matrice delle varianze e delle covarianze per ciascun individuo tiene conto del fatto che c'è autocorrelazione tra le osservazioni, mentre tra diversi individui tale autocorrelazione è inesistente.

- eteroschedasticità pura: $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma^2_{ij}$ (nello stesso istante c'è correlazione tra diversi individui),
- correlazioni pure: $E(\varepsilon_{it}\varepsilon_{is}) = \rho_{i,t-s}$ (per lo stesso individuo c'è correlazioni per le osservazioni in diversi istanti).

Considerando il vettore $(T \times 1)$ ε_i, si ha perciò che

$$E(\varepsilon_i\varepsilon_i') = \sigma_i^2\Sigma_i \tag{8}$$

con

$$\underset{(T\times T)}{\Sigma_i} = \begin{bmatrix} 1 & \rho_i & \rho_i^2 & \dots & \rho_i^{T-1} \\ \rho_i & 1 & \rho_i & \dots & \rho_i^{T-2} \\ \rho_i^2 & \rho_i & 1 & \dots & \rho_i^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \dots & 1 \end{bmatrix}.$$

Gli elementi extradiagonali della matrice Σ_i sono ottenuti ricorsivamente mediante un modello AR(1) calcolato sugli errori relativi all'i-esimo individuo (ε_i), cioè

$$\varepsilon_{i,t} = \rho_i\varepsilon_{i,t-1} + u_{i,t}$$

con $i = 1, 2, \dots, N$ e $t = 2, 3, \dots, T$. Per il calcolo di tutte le altre autocorrelazioni si procede mediante sostituzioni ricorsive[2]. Dal punto di vista analitico la matrice diagonale di cui alla (4) diventa diagonale a blocchi in quanto le matrici identità I_T (diagonali) vengono rimpiazzate con le matrici Σ_i (piene), quindi si ha

$$\Omega = \begin{bmatrix} \sigma_1^2\Sigma_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \sigma_2^2\Sigma_2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & & \vdots \\ 0 & 0 & & \sigma_i^2\Sigma_i & & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \sigma_N^2\Sigma_N \end{bmatrix}. \tag{9}$$

Il modello pertanto va stimato in due stadi: nel primo si effettua una regressione OLS di Y su X per ottenere i residui $\hat{\varepsilon}$. A questo punto, per ciascun individuo, si effettua un ulteriore OLS $\hat{\varepsilon}_{i,t} = \rho_i\hat{\varepsilon}_{i,t-1} + u_{i,t}$ per ottenere la stima consistente del parametro $\hat{\rho}_i$.

Il secondo step consiste in una stima WLS analoga alla (5) nella quale la matrice dei pesi è nota come Trasformazione di Prais e Winsten definita come segue

$$z_{i,t} = \hat{\rho}_i z_{i,t-1}$$

[2] In particolare, per l'autocorrelazione di ordine 2 risulta

$$\begin{aligned} \varepsilon_{i,t} &= \rho_i\varepsilon_{i,t-1} + u_{i,t} \\ &= \rho_i(\rho_i\varepsilon_{i,t-2} + u_{i,t-1}) + u_{i,t} \\ &= \rho_i^2\varepsilon_{i,t-2} + \rho_i u_{i,t-1} + u_{i,t}. \end{aligned}$$

Generalizzando, per l'autocorrelazione di ordine s si ha

$$\varepsilon_{i,t} = \rho_i^s\varepsilon_{i,t-s} + \sum_{r=0}^{s-1}\rho_i^r u_{i,t-r}.$$

Ovviamente il coefficiente ρ_i^s è quello che va immesso all'interno della matrice Σ_i.

dove $z_{i,t} = y_{i,t}, x_{i,t}$. Inoltre, per evitare la perdita della prima osservazione, si moltiplica $z_{i,1}$ per la quantità $\sqrt{1-\hat{\rho}_i}$.

Anche in questo caso lo stimatore ottenuto ha le usuali proprietà della non distorsione, della consistenza e dell'efficienza asintotica per $T \to \infty$.

1.5 Modello con eteroschedasticità e correlazioni pura e con correlazione tra individui

Questo modello è il più generale di tutti quelli proposti finora in quanto

- c'è eteroschedasticità dei termini di errore tra gli individui,
- c'è correlazione istantanea tra i diversi individui,
- c'è autocorrelazione tra le osservazioni relative ad ogni individuo.

La logica conseguenza di queste assunzioni è che la matrice Ω sia piena, quindi assuma la forma

$$\Omega = \begin{bmatrix} \sigma_1^2\Sigma_{11} & \sigma_{12}\Sigma_{12} & \dots & \sigma_{1i}\Sigma_{1i} & \dots & \sigma_{1N}\Sigma_{1N} \\ \sigma_{21}\Sigma_{21} & \sigma_2^2\Sigma_{22} & \dots & \sigma_{2i}\Sigma_{2i} & \dots & \sigma_{2N}\Sigma_{2N} \\ \vdots & \vdots & \ddots & & & \vdots \\ \sigma_{i1}\Sigma_{i1} & \sigma_{i2}\Sigma_{i2} & & \sigma_i^2\Sigma_{ii} & & \sigma_{iN}\Sigma_{iN} \\ \vdots & \vdots & & & \ddots & \vdots \\ \sigma_{N1}\Sigma_{N1} & \sigma_{N2}\Sigma_{N2} & \dots & \sigma_{Ni}\Sigma_{Ni} & \dots & \sigma_i^2\Sigma_{ii} \end{bmatrix}. \tag{10}$$

in questo caso il metodo di stima adottato sostanzialmente ricalca quello presentato nel precedente paragrafo.

2 Modelli per dati longitudinali

Quando si parla di dati longitudinali si intende una struttura come quella illustrata dalla matrice (1) nella quale generalmente la numerosità degli individui è elevata, mentre quella relativa alla dimensione temporale è piuttosto contenuta.

Si tenga presente che, qualora le ipotesi circa la matrice delle varianze e delle covarianze Ω e la costante (qualora ci fosse) rispettino quelle proprie dei modelli di serie storiche pooled, questi divengono automaticamente utilizzabili in questo contesto semplicemente scambiando gli indici relativi agli individui e al tempo.

2.1 Modello ad effetti fissi

Considerando l'i-esimo individuo, il modello ad effetti fissi si configura come segue

$$y_i = \alpha_i + x_i\beta + \varepsilon_i, \tag{11}$$

dove y_i e ε_i hanno dimensione $(T \times 1)$, x_i ha dimensione $(T \times k)$ e β è il vettore contenente k parametri da stimare. La peculiarità della (11) riguarda la costante che si configura come un vettore di T elementi costanti pari ad α_i: questa caratteristica indica innanzi tutto che per ciascun individuo occorre stimare un solo valore della costante e che, se $\alpha_i \neq \alpha_j$ per ogni $i \neq j$, tale costante misura l'EFFETTO INDIVIDUALE, cioè quell'insieme di caratteristiche specifiche proprie di ciascun individuo che però restano immutate nel tempo. In pratica, nel modello ci sono in tutto $k + N$ parametri da stimare, k contenuti nel vettore β ed N costanti per i diversi individui. Queste costanti rappresentano l'eterogeneità presente tra gli individui nel sistema, caratteristica peculiare dei panel data.

Generalizzando la (11) riscrivendola in forma matriciale si ottiene:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \\ y_N \end{bmatrix} = \begin{bmatrix} \iota_T & 0 & \dots & 0 & X_1 \\ 0 & \iota_T & \dots & 0 & X_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & X_{N-1} \\ 0 & 0 & \dots & \iota_T & X_N \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{N-1} \\ \varepsilon_N \end{bmatrix}, \tag{12}$$

dove ι_T è un vettore contenente T elementi pari a 1. In forma compatta si ha perciò

$$Y = [\ (I_N \otimes \iota_T) \quad X\ ] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon \tag{13}$$

oppure

$$\underset{(NT\times 1)}{Y} = \underset{(NT\times N)}{(I_N \otimes \iota_T)} \underset{(N\times 1)}{\alpha} + \underset{(NT\times k)}{X} \underset{(k\times 1)}{\beta} + \underset{(NT\times 1)}{\varepsilon} \tag{14}$$

Poiché i valori del vettore α non sono osservabili essi entrerebbero a pieno titolo all'interno dell'errore del modello ma, se così fosse, essi potrebbero essere correlati con le variabili esplicative X_i e la stima risulterebbe distorta.

Le formulazione (14) permette di stimare il modello attraverso l'OLS in quanto tutte le ipotesi classiche sono rispettate. Il modello prende il nome di MODELLO A VARIABILI DUMMY poiché occorre costruire N (nuerosità degli effetti individuali) variabili dummy da inserire all'interno della matrice dei regressori. Lo stimatore che si ottiene è non distorto, consistente e BLUE. La sua forma analitica è ottenibile come

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} (I_N \otimes \iota_T)'(I_N \otimes \iota_T) & (I_N \otimes \iota_T)'X \\ X'(I_N \otimes \iota_T) & X'X \end{bmatrix}^{-1} \begin{bmatrix} (I_N \otimes \iota_T)'Y \\ X'Y \end{bmatrix}$$

Dato che per le proprietà del prodotto di Kronecker vale $(I_N \otimes \iota_T)'(I_N \otimes \iota_T) = I_N \otimes \iota_T'\iota_T = TI_N$, risulta

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} TI_N & (I_N \otimes \iota_T)'X \\ X'(I_N \otimes \iota_T) & X'X \end{bmatrix}^{-1} \begin{bmatrix} (I_N \otimes \iota_T)'Y \\ X'Y \end{bmatrix}.$$

Per invertire la matrice contenuta all'interno dell'espressione dello stimatore OLS si ricorre ad un noto risultato sulle matrici partizionate e, dopo alcuni calcoli[3] si arriva a

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \frac{1}{T}(I_N \otimes \iota_T)'(Y - X\hat{\beta}) \\ \\ (X'MX)^{-1}X'MY \end{bmatrix}, \tag{15}$$

dove $M = I_{NT} - P$ è la matrice di proiezione che, applicata ad una variabile, per ogni individuo restituisce lo scostamento dalla media aritmetica temporale. Tale matrice, per definizione, risulta essere quadrata ($NT \times NT$), diagonale a blocchi, simmetrica ed idempotente[4].

[3]In particolare ci si riferisce alla seguente inversione

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S_2A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S2 \\ -S_2A_{21}A_{11}^{-1} & S_2 \end{bmatrix},$$

dove $S_2 = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$. L'Appendice A-2 contiene tutta la derivazione analitica dello stimatore del modello ad effetti fissi.

[4]Definizione e proprietà delle matrici P e M sono discusse nell'Appendice A-1.

2.2 Stimatore within

Prendendo in considerazione lo stimatore $\hat{\beta}$ determinato nell'equazione (15) e tenendo presente la proprietà di idempotenza della matrice M si ha

$$\begin{aligned}\hat{\beta} &= (X'MX)^{-1}X'MY \\ &= (\dot{X}\dot{X})^{-1}\dot{X}\dot{Y}. \end{aligned} \tag{16}$$

Tale stimatore è perciò ottenibile anche attraverso la regressione OLS di $\dot{Y} = MY$ su $\dot{X} = MX$; in pratica si tratta di applicare il modello lineare classico dove sia la variabile dipendente, sia la matrice dei regressori sono espressa in deviazione dalle corrispondenti medie individuali calcolate rispetto al tempo[5].

Lo stimatore $\hat{\beta}$ prende perciò il nome di STIMATORE WITHIN in quanto tiene conto degli effetti individuali grazie alla trasformazione effettuata attraverso la matrice M, ma li elimina[6] dal modello utilizzando per ciascun individuo l'informazione derivante dalle variazioni temporali (variazioni "nei gruppi").

Lo stimatore within e lo stimatore a variabili dummy producono sempre gli stessi valori numerici.

Una volta ottenuto lo stimatore within, gli effetti individuali esclusi dal suo computo possono essere sfruttati attraverso l'equazione (14), infatti

$$\begin{aligned}&(I_N \otimes \iota_T)\alpha = Y - X\hat{\beta} \\ &\frac{1}{T}(I_N \otimes \iota_T)'(I_N \otimes \iota_T)\alpha = \frac{1}{T}(I_N \otimes \iota_T)'(Y - X\hat{\beta}) \\ &\frac{1}{T}(I_N \otimes \iota_T')(I_N \otimes \iota_T)\alpha = \frac{1}{T}(I_N \otimes \iota_T)'(Y - X\hat{\beta}) \\ &\frac{1}{T}(I_N \otimes T)\alpha = \frac{1}{T}(I_N \otimes \iota_T)'(Y - X\hat{\beta}) \\ &\hat{\alpha} = \frac{1}{T}(I_N \otimes \iota_T)'(Y - X\hat{\beta}). \end{aligned} \tag{17}$$

L'equazione (17) mostra che, per ogni singolo individuo, la costante è pari alla differenza tra la media individuale della variabile dipendente e le medie individuali dei regressori ponderate per lo stimatore within. Dal punto di vista dell'individuo, analiticamente si ha

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i\hat{\beta} \tag{18}$$

Le costanti α_i con $i = 1, 2, \ldots, N$ catturano l'effetto di quelle variabili che variano tra individuo e individuo, ma restano immutate nel tempo; lo stimatore within perciò tiene conto solo dell'eterogeneità tra gli individui.

Il limite più evidente di questo approccio consiste nell'impossibilità di includere nel modello regressori che assumano un valore costante all'interno delle osservazioni relative al singolo individuo: dal punto di vista statistico, questa impossibilità deriva dal fatto che una variabile esplicativa con questa caratteristica risulterebbe collineare con $(I_N \otimes \iota_T)$ nell'equazione (14), mentre dal punto di vista algebrico calcolare lo scostamento di queste variabili dal loro valore medio individuale (attraverso la matrice M) produrrebbe colonne di zeri nella matrice dei regressori che quindi non avrebbe rango pieno. In questo caso il metodo OLS non sarebbe perciò applicabile.

Per la verifica di ipotesi relativa all'assenza di eterogeneità tra gli individui il test t di azzeramento delle costanti α_i non è di alcuna utilità pratica. È invece possibile costruire un test F nel quale l'ipotesi nulla è $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_N$ ($N - 1$ vincoli in tutto); la statistica test è

$$\frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}} \cdot \frac{NT - N - K - 1}{N - 1} \sim F_{N-1, NT-N-K-1}, \tag{19}$$

dove $\tilde{\varepsilon}$ e $\hat{\varepsilon}$ sono i residui rispettivamente del modello vincolato e di quello libero, mentre lo stimatore corretto e consistente per la varianza è

$$\hat{\sigma}^2_\varepsilon = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{NT - N - K - 1}.$$

[5] Si tenga presente che, per l'ipotesi classica $E(\varepsilon) = 0$, quindi risulta $M\varepsilon = \varepsilon$

[6] È ovvio che il prodotto $M(I_N \otimes \iota_T) = 0$ quindi le costanti del modello sono rimosse attraverso il calcolo dello stimatore within.

Alla luce di questo risultato si ha inoltre $Var(\hat{\beta}) = \hat{\sigma}_\varepsilon^2 (X'MX)^{-1}$. Si noti infine che, sotto H_0, di fatto lo stimatore within coincide con lo stimatore pooled.

Lo stimatore within è

- BLUE,
- consistente per $NT \to \infty$,
- asintoticamente normale poiché
$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{\text{d}} N\left(0, \sigma_\varepsilon^2 Q^{-1}\right),$$
dove $Q = \lim_{NT \to \infty} \left(\frac{1}{NT} X'MX \right)^{-1}$.

2.3 Modello ad effetti casuali

Il modello ad effetti casuali tratta gli effetti individuali come parte del termine di errore, quindi li considera come componenti stocastiche sicuramente incorrelate con i regressori: in questo modo è possibile includere all'interno della matrice X variabili che cambiano tra soggetto e soggetto, pur rimenendo costanti all'interno delle T osservazioni relative al singolo individuo. Con il modello ad effetti fissi questa opportunità era preclusa.

Considerando l'i-esimo individuo, la forma analitica del modello ad effetti casuali è

$$\begin{aligned} y_i &= \alpha_i + x_i \beta + \varepsilon_i \\ y_i &= \alpha + x_i' \beta + \mu_i + \varepsilon_i \end{aligned} \tag{20}$$

dove il vettore $(T \times 1)$ relativo alla costante $\alpha_i = \alpha + \mu_i$ è dato dalla somma di una componente indipendente da i e da t e da un'altra che varia da individuo ad individuo. Ovviamente, dato i, α_i è un vettore di costanti. Affinché si ottengano stime consistenti con quest'approccio, la condizione necessaria è l'incorrelazione tra α_i e la matrice dei regressori x_i per ogni i.

Rispetto al modello ad effetti fissi il termine di errore ε_i ha esattamente tutte le stesse proprietà, mentre occorre introdurre alcune ipotesi aggiuntive riguardo alla componente μ_i.

1. $E(\mu_i) = 0$,
2. $Var(\mu_i) = \sigma_\mu^2$ per ogni $i = 1, 2, \ldots, N$,
3. $E(\mu_i, \mu_j) = 0$ per ogni $i \neq j$ (incorrelazione tra gli effetti individuali),
4. $E(\mu_i, \varepsilon_{j,t}) = 0$ per ogni i, j, t (incorrelazione tra effetti individuali e disturbi).

Riscrivendo il modello in forma compatta si ha

$$\underset{(NT\times 1)}{Y} = \underset{(NT\times 1)}{\alpha} + \underset{(NT\times k)}{X}\underset{(k\times 1)}{\beta} + \underset{(NT\times 1)}{(\mu \otimes \iota_T)} + \underset{(NT\times 1)}{\varepsilon} \tag{21}$$

dove μ di dimensione N è il vettore contenente gli effetti individuali. Definendo inoltre il vettore $U = (\mu \otimes \iota_T) + \varepsilon$ si nota immediatamente che l'errore del modello ad effetti casuali è composto di una componente che varia tra gli individui, ma resta costante nel tempo, ed un'altra che varia stocasticamente tra gli individui e nel tempo.

Date le ipotesi aggiuntive di cui sopra, la matrice delle varianze e delle covarianze di U ricopre un ruolo determinante. Essa è definita come

$$\begin{aligned} \Omega &= Var(U) \\ &= E(UU') \\ &= E\{[(\mu \otimes \iota_T) + \varepsilon][(\mu \otimes \iota_T) + \varepsilon]'\} \\ &= E[(\mu \otimes \iota_T)(\mu \otimes \iota_T)' + \varepsilon\varepsilon'] \\ &= E(\mu\mu' \otimes \iota_T \iota_T' + \varepsilon\varepsilon') \\ &= E(\mu\mu' \otimes \iota_T \iota_T') + E(\varepsilon\varepsilon'). \end{aligned}$$

Dato che $E(\mu\mu') = \sigma^2_\mu I_N$, la matrice $E(\mu\mu' \otimes \iota_T\iota'_T)$ assume una struttura diagonale a blocchi quindi, tenendo presente anche che $E(\varepsilon\varepsilon') = \sigma^2_\varepsilon I_{NT}$, si ottiene

$$\Omega = \sigma^2_\mu(I_N \otimes \iota_T\iota'_T) + \sigma^2_\varepsilon I_{NT} = I_N \otimes (\sigma^2_\mu \iota_T\iota'_T + \sigma^2_\varepsilon I_T). \tag{22}$$

La matrice Ω è anch'essa diaginale a blocchi e ciascun blocco è dato da

$$\underset{(T\times T)}{\Omega_i} = \begin{bmatrix} \sigma^2_\mu + \sigma^2_\varepsilon & \sigma^2_\mu & \dots & \sigma^2_\mu \\ \sigma^2_\mu & \sigma^2_\mu + \sigma^2_\varepsilon & \dots & \sigma^2_\mu \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_\mu & \sigma^2_\mu & \dots & \sigma^2_\mu + \sigma^2_\varepsilon \end{bmatrix}.$$

La matrice Ω_i mostra che l'errore composto (U) ha autocorrelazione non nulla e costante nel tempo e soprattutto che la struttura di autocorrelazione non varia da individuo ad individuo (la matrice è priva degli indici i e t).

Poiché tale matrice delle varianze e delle covarianze è diagonale a blocchi, il modello ad effetti casuali deve essere stimato attraverso il metodo GLS, quindi si ha

$$\hat{b} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{23}$$

dove $\hat{b} = [\ \hat{\alpha} \quad \hat{\beta}\ ]'$ ha dimensione $(k+1)$.

La matrice inversa Ω^{-1} è data da

$$\begin{aligned} \Omega^{-1} &= (I_N \otimes \Omega_i)^{-1} \\ &= I_N \otimes \Omega_i^{-1} \\ &= I_N \otimes (\sigma^2_\mu \iota_T\iota'_T + \sigma^2_\varepsilon I_T)^{-1}. \end{aligned}$$

Aggiungendo e togliendo $P_\iota\sigma^2_\varepsilon$ si ottiene

$$\begin{aligned} \Omega^{-1} &= I_N \otimes [(T\sigma^2_\mu + \sigma^2_\varepsilon)P_\iota + \sigma^2_\varepsilon(I_T - P_\iota)]^{-1} \\ &= I_N \otimes [(T\sigma^2_\mu + \sigma^2_\varepsilon)P_\iota + \sigma^2_\varepsilon M_\iota]^{-1} \\ &= [(T\sigma^2_\mu + \sigma^2_\varepsilon)P + \sigma^2_\varepsilon M]^{-1}. \end{aligned}$$

Ponendo $\sigma^2 = (T\sigma^2_\mu + \sigma^2_\varepsilon)$, per le proprietà delle matrici P e M si ha[7]

$$\Omega^{-1} = \frac{1}{\sigma^2}P + \frac{1}{\sigma^2_\varepsilon}M \tag{24}$$

e quindi

$$\Omega^{-1/2} = \frac{1}{\sigma}P + \frac{1}{\sigma_\varepsilon}M. \tag{25}$$

Da questa definizione emerge che lo stimatore GLS per il modello ad effetti casuali coincide con lo stimatore OLS della regressione di $\dot{Y} = \Omega^{-1/2}Y$ su $\dot{X} = \Omega^{-1/2}X$. Le proprietà di questo stimatore sono

1. se σ^2_ε e σ^2_μ sono noti, lo stimatore GLS è consistente per $N \to \infty$ e $T \to \infty$,
2. per T dato, lo stimatore GLS è più efficiente dello stimatore within; per $N \to \infty$ tale efficienza tende a svanire,
3. se $\Omega^{-1} \equiv M$ lo stimatore GLS coincide con lo stimatore within, quindi il modello ad effetti casuali coincide con quello ad effetti fissi: ciò può accadere se l'unica fonte di variabilità deriva dagli effetti individuali μ_i. Analiticamente deve perciò risultare che
   - $\sigma^2_\varepsilon = 0$ (vettore ε costante per ogni i e t),

[7]Si veda l'Appendice A-1.

- $T \to \infty$ (per definizione $\hat{\sigma}_\varepsilon^2 = 0$): in questo caso gli effetti individuali diventano osservabili[8],

,

4. se $\Omega^{-1} \equiv I_{NT}$ il modello ad effetti casuali diventa un modello OLS standard e coincide con un modello di serie storiche pooled; in questo caso naturalmente $\sigma_\mu^2 = 0$ quindi non ci sono effetti individuali e tutta la variabilità dipende dal termine di disturbo ε.

2.4 Stimatore between

Considerando il modello ad effetti casuali di cui alla (21), la trasformazione Between consiste nell'esprimere le variabili attraverso le medie temporali di ciascun individuo; in pratica algebricamente si tratta di premoltiplicare l'intera equazione per la matrice P,

$$\begin{aligned} PY &= P\alpha + PX\beta + P[(\mu \otimes \iota_T) + \varepsilon] \\ &= PXb + Pu. \end{aligned}$$

Lo stimatore che si allpica è perciò un GLS che si configura come un modello OLS della regressione di $\dot{Y} = PY$ su $\dot{X} = PX$, infatti

$$\begin{aligned} \hat{b} &= (X'P^{-1}X)^{-1}X'P^{-1}Y \\ &= (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y} \end{aligned} \tag{26}$$

dove $\hat{b} = [\ \hat{\alpha} \quad \hat{\beta}\ ]'$ ha dimensione $(k+1)$. Lo stimatore di cui alla (26) risulta essere non distorto e consistente per $N \to \infty$.

Analogamente allo stimatore within, lo stimatore between determina una perdita di informazione poiché si basa sul calcolo delle medie temporali di ciascun individuo. Per definizione, tale trasformazione produce una perdita di efficienza.

Mentre lo stimatore within sfrutta la variazione che avviene all'interno delle osservazioni relative a ciascun individuo (deviazioni dalle medie o variazioni "nei gruppi"), lo stimatore between sfrutta quelle derivanti dalla variabilità delle osservazioni tra diversi individui (variazioni "tra i gruppi"), in quanto opera una regressione di N medie su un set di regressori nel quale sono state calcolate le N medie corrispondenti.

2.5 Stimatore GLS, within e between

I tre stimatori visti finora possono essere messi in relazione in quanto lo stimatore GLS è una media ponderata degli stimatori within e between; considerando i parametri $a_1 \in [0,1]$ e $a_2 = 1 - a_1$ e le due trasformazioni within e between si ha

$$(a_1P + a_2M)Y = (a_1P + a_2M)X\beta + a_1P\varepsilon_{bet} + a_2M\varepsilon_{wit}.$$

Lo stimatore GLS che ne risulta è

$$\begin{aligned} \hat{\beta}_{GLS} &= [X'(a_1P + a_2M)'(a_1P + a_2M)X]^{-1}X'(a_1P + a_2M)'(a_1P + a_2M)Y \\ &= [X'(a_1^2P + a_2^2M)X]^{-1}X'(a_1^2P + a_2^2M)Y. \end{aligned} \tag{27}$$

È perciò possibile esprimere lo stimatore GLS semplicemente imponendo $\Omega^{-1} = (a_1P + a_2M)$. Poiché dall'equazione (25) risulta $a_1 = 1/\sigma$ e $a_2 = 1/\sigma_\varepsilon$, dove $\sigma = (T\sigma_\mu^2 + \sigma_\varepsilon^2)^{1/2}$, si hanno i seguenti scenari:

- se $\sigma_\varepsilon^2 = 0 \Rightarrow a_2 \to \infty$ (peso infinito assegnato allo stimatore within),
- se $T \to \infty \Rightarrow a_1 = 0$ (lo stimatore GLS coincide con lo stimatore within, gli effetti individuali sono osservabili),
- se $\sigma_\mu^2 = 0 \Rightarrow \sigma = \sigma_\varepsilon$, $a_1 = a_2$ (lo stimatore GLS in realtà è uno stimatore OLS, omoschedasticità).

[8]Considerando il modello per la singola osservazione $y_{it} - \alpha - x_{it}'\beta = \mu_i + \varepsilon_{it}$, se $T \to \infty$ significa che il valore atteso della componente ε_{it} è davvero nullo quindi l'espressione a sinistra del segno di uguaglianza rappresenta la singola osservazione per μ_i. In questo caso lo stimatore GLS è consistente.

2.6 Stimatore FGLS

Quando σ^2_ε e σ^2_μ sono osservabili in pratica lo stimatore GLS può essere applicato senza alcun problema; nella pratica questa situazione capita raramente.

Per ovviare a questo inconveniente si ricorre allo stimatore "Feasible GLS" (FGLS). Innanzi tutto si ricorre ai residui dello stimatore within $\hat{\varepsilon}_{wit}$ per ottenere lo stimatore

$$\hat{\sigma}^2_\varepsilon = \frac{\hat{\varepsilon}'_{wit} M \hat{\varepsilon}_{wit}}{NT - N - k}, \tag{28}$$

dove la correzione per i gradi di libertà è data dal numero dei parametri da stimare che ammonta a $N + k$.[9]

Successivamente si ricorre al modello ad effetti casuali e si considera il modello relativo all'i-esima media individuale rispetto al tempo $y_i - \alpha - \beta x_i = \mu_i + \varepsilon_i$; la varianza rispetto allo scalare $u_i = \mu_i + \varepsilon_i$ è data da

$$\begin{aligned} Var(u_i) &= Var(\mu_i + \varepsilon_i) \\ &= Var(\mu_i) + Var\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right) \\ &= \sigma^2_\mu + \frac{\sigma^2_\varepsilon}{T} \\ &= \sigma^2_R. \end{aligned}$$

Considerando perciò l'i-esimo individuo, uno stimatore corretto e consistente per σ^2_R è

$$\hat{\sigma}^2_R = \frac{\hat{u}'_i \hat{u}_i}{N - k}, \tag{29}$$

dove $\hat{u}_i$ sono i residui del modello e k indica il numero dei regressori escludendo la costante. Data la definizione analitica di σ^2_R è immediato stimare indirettamente la varianza degli effetti individuali attraverso l'equazione

$$\hat{\sigma}^2_\mu = \hat{\sigma}^2_R - \frac{\hat{\sigma}^2_\varepsilon}{T} \tag{30}$$

Attraverso questa relazione è quindi possibile stimare il modello col metodo GLS (che diviene *feasible*). L'unico inconveniente di questo metodo è determinato dal fatto che, in campioni finiti, può accadere che la (30) restituisca un valore negativo.

2.7 Test statistici

Per decidere se è preferibile la stima di un modello ad effetti fissi o uno ad effetti casuali è possibile utilizzare alcune procedure di test. I più famosi sono il test di Breusch e Pagan (1980) e quello di Hausman (1978).

2.7.1 Test di Breusch e Pagan

Il test di Breusch e Pagan (test BP) è uno dei test diagnostici più popolari per valutare la presenza di eteroschedasticità all'interno del modello lineare di regressione $Y = X\beta + \varepsilon$ con $\varepsilon \sim N(0, \sigma^2\Omega)$. L'ipotesi nulla del test è l'assenza di eteroschedasticità quindi, poiché vale l'assunzione

$$Var(\varepsilon) = \sigma^2 f(Z\gamma) = \sigma^2 f(\gamma_0 + \gamma_1 Z_1 + \ldots + \gamma_q Z_q),$$

essa si struttura come

$$H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_q = 0 \qquad (q \text{ vincoli}), \tag{31}$$

[9] Se si considerasse lo scenario relativo a ciascun individuo si avrebbero $N(T - k - 1)$ g.d.l. in tutto, quindi una stima in eccesso del loro numero.

dove Z è una matrice dove ciascuna delle $(q+1)$ colonne costituisce una variabile esplicativa per la varianza del termine di errore. La statistica test, nella sua forma generale, si configura come un test LM e risulta essere

$$LM_{BP} = \frac{1}{2\hat{\gamma_0}^2}(\hat{\varepsilon}^2 - \hat{\gamma}_0)' Z(Z'Z)^{-1} Z'(\hat{\varepsilon}^2 - \hat{\gamma}_0), \tag{32}$$

dove $\hat{\gamma}_0 = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$ è lo stimatore OLS non corretto della varianza, mentre n è il numero totale delle osservazioni. In pratica, la statistica test (32) è esprimibile come[10]

$$LM_{BP} = nR^2$$

dove l'indice R^2 è quello relativo alla regressione di $(\hat{\varepsilon}^2/\hat{\gamma}_0 - 1)$ su Z. La distribuzione limite della statistica test BP è $LM_{BP} \sim \chi^2_q$. Per il calcolo di questa statistica occorre procedere come segue:

- stima OLS del modello $Y = X\beta + \varepsilon$,
- calcolo dello stimatore $\hat{\gamma}_0$,
- stima della regressione ausiliaria,
- calcolo dell'indice R^2.

Nell'ambito dei modelli panel data è possibile ricorrere al test BP per sottoporre a verifica di ipotesi la significatività degli effetti individuali. L'ipotesi nulla impone il solo vincolo

$$H_0 : \sigma^2_\mu = 0, \tag{33}$$

che garantisce omoschedasticità, quindi la matrice Ω diagonale. Il test BP necessita solo dei residui del modello vincolato che in questo contesto è dato dal modello ad effetti fissi, quindi la statistica test assume la forma

$$\begin{aligned} LM_{BP} &= \frac{NT}{2(T-1)} \left[\frac{\hat{\varepsilon}'_{wit}(I_N \otimes \iota_T)(I_N \otimes \iota_T)'\hat{\varepsilon}_{wit} - \hat{\varepsilon}'_{wit}\hat{\varepsilon}_{wit}}{\hat{\varepsilon}'_{wit}\hat{\varepsilon}_{wit}} \right]^2 && (34) \\ &= \frac{NT}{2(T-1)} \left[\frac{\hat{\varepsilon}'_{wit}(I_N \otimes \iota_T \iota'_T)\hat{\varepsilon}_{wit}}{\hat{\varepsilon}'_{wit}\hat{\varepsilon}_{wit}} - 1 \right]^2, && (35) \end{aligned}$$

dove $\hat{\varepsilon}_{wit}$ è il residuo del modello stimato attraverso lo stimatore within. Poiché in questo caso l'ipotesi nulla impone solo un vincolo, la distribuzione limite della statistica test è data da una χ^2_1.

2.7.2 Test di Hausman

Un'altra procedura di test per la scelta del modello panel da adottare è data dal test di Hausman (1978); lo stimatore withi è costoso in termini di variabili da inserire nel modello e ciò genera una perdita di g.d.l., mentre lo stimatore ad effetti casuali deve avere la prerogativa che gli effetti individuali devono essere incorrelati coi regressori altrimenti lo stimatore stesso è inconsistente.

Ponendo $u = \mu \otimes \iota_T + \varepsilon$, il test di Hausman si occupa perciò di testare l'ipotesi nulla

$$\begin{cases} H_0 : E(X'u) = 0 \\ H_1 : E(X'u) \neq 0. \end{cases}$$

Considerando gli stimatori within (OLS) e GLS si hanno i seguenti scenari:

	H_0	H_1
$\hat{\beta}_{OLS}$	consistente inefficiente	consistente
$\hat{\beta}_{GLS}$	consistente efficiente	inconsistente

[10] Si veda l'Appendice A-3 per la dimostrazione.

Naturalmente il test è basato sulla differenza $\hat{q} = \hat{\beta}_{OLS} - \hat{\beta}_{GLS}$: se questa risulta essere statisticamente irrilevante è preferibile l'utilizzo degli effetti casuali, mentre se $\hat{q}$ è diversa da zero lo stimatore within è preferibile[11].

La statistica test è data da

$$H = \hat{q}'[Var(\hat{q})]^{-1}\hat{q} \tag{36}$$

dove

$$Var(\hat{q}) = Var(\hat{\beta}_{OLS}) + Var(\hat{\beta}_{GLS}) + 2Cov(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}).$$

Sotto H_0 si può dimostrare che la covarianza tra i due stimatori OLS e GLS è nulla, infatti basta considerare lo stimatore $\tilde{\beta}$ definito dalla seguente combinazioe lineare

$$\tilde{\beta} = \hat{\beta}_{GLS} + \lambda\hat{\beta}_{OLS},$$

dove λ è uno scalare diverso da zero; calcolando la sua varianza si ottiene

$$\begin{aligned} Var(\tilde{\beta}) &= Var(\hat{\beta}_{GLS}) + \lambda^2 Var(\hat{\beta}_{GLS}) + 2\lambda Cov(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}) \\ Var(\tilde{\beta}) - Var(\hat{\beta}_{GLS}) &= \lambda^2 Var(\hat{\beta}_{GLS}) + 2\lambda Cov(\hat{\beta}_{OLS}, \hat{\beta}_{GLS}). \end{aligned}$$

Poiché $Var(\tilde{\beta}) - Var(\hat{\beta}_{GLS}) \geq 0$ per definizione, occorre necessariamente che anche l'equazione di secondo grado spuria al secondo membro sia maggiore o uguale a zero, cioè

$$\lambda[\lambda Var(\hat{\beta}_{GLS}) + 2Cov(\hat{\beta}_{OLS}, \hat{\beta}_{GLS})] \geq 0.$$

Le soluzioni per questa disequazione sono $\lambda \leq 0$ e $\lambda \geq -2\dfrac{Cov(\hat{\beta}_{OLS}, \hat{\beta}_{GLS})}{Var(\hat{\beta}_{OLS})}$. Ovviamente, la condizione di positività $Var(\tilde{\beta}) - Var(\hat{\beta}_{GLS})$ è garantita per ogni λ se e solo se i due stimatori OLS e GLS sono incorrelati.

Alla luce di questo risultato si ha semplicemente che $\hat{q} = Var(\hat{\beta}_{OLS}) + Var(\hat{\beta}_{GLS})$. La distribuzione del test di Hausman è $H \sim \chi^2_k$ dove k è il numero delle colonne di X (numero di regressori).

3 Panel dinamici

Uno sviluppo naturale e recente della letteratura sui modelli di tipo panel è quella relativa ai panel dinamici caratterizzati dalla presenza della variabile dipendente ritardata all'interno della matrice dei regressori. In questo modo è possibile modellare, quindi distinguere tra due diversi tipi di correlazione:

1. "vera": autocorrelazione della variabile dipendente,

2. "spuria": correlazione dovuta ad eterogeneità non osservata.

Prendendo come riferimanto la singola osservazione e limitando per semplicità la trattazione ai modelli con un solo ritardo, l'equazione generale per un panel dinamico è

$$y_{it} = X'_{it}\beta + \phi y_{it-1} + u_{it}, \tag{37}$$

dove $u_{it} = \mu_i + \varepsilon_{it}$ e ϕ è il parametro relativo alla componente autoregressiva del modello.

Il problema principale di questo tipo di modelli è dato dal fatto che il termine di errore u_{it} non è incorrelato con y_{it-1} e ciò genera stime OLS e GLS inconsistenti. In particolare

$$\begin{aligned} E(u_{it}y_{it-1}) &= E[u_{it}(X'_{it-1}\beta + \phi y_{it-2} + u_{it-1})] \\ &= E[(\mu_i + \varepsilon_{it})(X'_{it-1}\beta + \phi y_{it-2} + \mu_i + \varepsilon_{it-1})] \\ &= E(\mu_i^2) \\ &= \sigma^2_\mu \neq 0, \end{aligned}$$

[11]È pertanto possibile dimostrare che lo stimatore GLS con effetti casuali correlati coi regressori si identifica nello stimatore within.

quindi i valori nel tempo della variabile dipendente dipendono da μ_i e non possono essere incorrlati col termine di errore. Gli stimatori applicabili nell'approccio statico sono perciò inconsistenti[12].

Applicando la trasformazione within all'equazione (37) implica l'ipotesi di una trattazione degli effetti individuali come fissi, ma tale strategia conduce ugualmente ad uno stimatore inconsistente; anche se si ha la seguente equazione che rimuove gli effetti fissi

$$y_{it} - \bar{y}_i = (X_{it} - \bar{x}_i)'\beta + \phi(y_{it-1} - \bar{y}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i),$$

tuttavia risulta

$$\begin{aligned} E[(y_{it-1} - \bar{y}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)] &= E[y_{it-1}\varepsilon_{it} - y_{it-1}\bar{\varepsilon}_i - \bar{y}_i\varepsilon_{it} + \bar{y}_i\bar{\varepsilon}_i] \\ &= E[-\bar{y}_i\varepsilon_{it}] \\ &= E\left[-\left(\frac{1}{T}\sum_{t=1}^{T} y_{it}\right)\varepsilon_{it}\right] \\ &= E\left[-\frac{1}{T}(y_{i1} + y_{i2} + \ldots + y_{it} \ldots + y_{iT})\varepsilon_{it}\right] \\ &= E\left[-\frac{1}{T}y_{it}\varepsilon_{it}\right] \\ &= -\frac{1}{T}E[\varepsilon_{it}^2] \\ &= -\frac{1}{T}\sigma_\varepsilon^2 \neq 0. \end{aligned}$$

Lo stimatore within è perciò anch'esso inconsistente per T finito, mentre diviene consistente per $T \to \infty$.

3.1 Stimatore di Anderson-Hsiao

Riscrivendo l'equazione (37) in termini di differenze prime si ottiene

$$\Delta y_{it} = \Delta X_{it}'\beta + \phi\Delta y_{it-1} + \Delta\varepsilon_{it}, \tag{38}$$

quindi gli effetti individuali vengono eliminati in quanto $\Delta u_{it} = \varepsilon_{it} - \varepsilon_{it-1}$; in particolare si ha $\Delta\varepsilon_{it} \sim MA(1)$, dove il coefficiente associato alla componente ritardata è ovviamente pari a 1.

Anche in questo caso però il problema della correlazione tra variabile dipendente ed errore ha il suo peso, infatti

$$\begin{aligned} E(\Delta y_{it-1}\Delta\varepsilon_{it}) &= E[(y_{it-1} - y_{it-2})(\varepsilon_{it} - \varepsilon_{it-1})] \\ &= E[y_{it-1}\varepsilon_{it} - y_{it-2}\varepsilon_{it} - y_{it-1}\varepsilon_{it-1} + y_{it-2}\varepsilon_{it-1}] \\ &= E[-y_{it-1}\varepsilon_{it-1}] \neq 0, \end{aligned}$$

in quanto y_{it-1} dipende da ε_{it-1}. Tale problema può essere superato ricorrendo allo stimatore a variabili strumentali (IV o 2SLS) utilizzando y_{it-2} come strumento per il quale vale

$$E(y_{it-2}\varepsilon_{it}) = 0.$$

Naturalmente la scelta dei ritardi della variabile dipendente da utilizzare come strumenti nella stima dipende strettamente dalla presenza di autocorrelazione negli errori. Tecnicamente, è perciò possibile spingersi molto indietro nel tempo per trovare uno strumento incorrelato coi regressori, ma ciò presenta il costo della perdita di osservazioni.

[12] Sostituendo ricorsivamente nella (37) si ottiene

$$y_{it} = \phi^t y_{i0} + \left(\sum_{j=0}^{t-1}\phi^j X_{it-j}'\right) + \sum_{j=0}^{t-1}\phi^j u_{it-j}.$$

La variabile dipendente è funzione dall'errore presente e passato, quindi è correlata con esso. Per la definizione di u_{it} emerge inoltre che essa dipende dagli effetti individuali μ_i. Se si considerano i ritardi di tale variabile il discorso non cambia.

3.2 Stimatore di Arellano-Bond

Lo stimatore di Arellano-Bond (1991) è uno stimatore a variabili strumentali che rappresenta lo strumento principe nella stima dei modelli di tipo panel dinamico.

3.2.1 Modello autoregressivo puro

Per semplicità, per la spiegazione del modello di Arellano-Bond si ricorre inizialmente al modello autoregressivo puro nel quale i regressori esogeni sono omessi ($\beta = 0$); si ha perciò l'equazione

$$y_{it} = \phi y_{it-1} + \mu_i + \varepsilon_{it}. \tag{39}$$

Le ipotesi alla base di questo metodo di stima sono:

- T è fisso,
- $N \to \infty$,
- $\varepsilon_{it} \sim i.i.d.(0, \sigma_\varepsilon^2)$.

Si considera pertanto il modello in differenze prime

$$\begin{aligned} \Delta y_{it} &= \phi \Delta y_{it-1} + \Delta\varepsilon_{it} \\ &= \phi(y_{it-1} - y_{it-2}) + \varepsilon_{it} - \varepsilon_{it-1} \end{aligned} \tag{40}$$

dove ovviamente $\Delta\varepsilon_{it} \sim MA(1)$, $i = 1, 2, \ldots, N$ e $t = 3, 4, \ldots, T$. L'equazione (40) equivale ad un sistema di equazioni simultanee con $(T-2)$ equazioni con N osservazioni ciascuna del tipo

$$\begin{cases} \Delta y_{i3} = \phi \Delta y_{i2} + \Delta\varepsilon_{i3} & \text{strumenti: } \Delta y_{i1} \\ \Delta y_{i4} = \phi \Delta y_{i3} + \Delta\varepsilon_{i4} & \text{strumenti: } \Delta y_{i1}, \Delta y_{i2} \\ \vdots & \\ \Delta y_{iT} = \phi \Delta y_{iT-1} + \Delta\varepsilon_{iT} & \text{strumenti: } \Delta y_{i1}, \Delta y_{i2}, \ldots, \Delta y_{iT-2}, \end{cases} \tag{41}$$

dove gli strumenti sono selezionati in base alla loro proprietà di essere incorrelati coi termini di errore. In questo modo è possibile ottenere una stima consistente del modello dinamico.

A queso punto è importante costruire la matrice delle varianze e delle covarianze di $\Delta\varepsilon_{it}$ che risulta essere composta da

- $Var(\Delta\varepsilon_{it}) = Var(\varepsilon_{it} - \varepsilon_{it-1}) = 2\sigma_\varepsilon^2$,
- $Cov(\Delta\varepsilon_{it}\Delta\varepsilon_{it-1}) = E(\varepsilon_{it}\varepsilon_{it-1} - \varepsilon_{it-1}^2 - \varepsilon_{it}\varepsilon_{it-2} + \varepsilon_{it-1}\varepsilon_{it-2}) = -\sigma_\varepsilon^2$
- $Cov(\Delta\varepsilon_{it}\Delta\varepsilon_{it-k}) = E(\varepsilon_{it}\varepsilon_{it-k} - \varepsilon_{it-1}\varepsilon_{it-k} - \varepsilon_{it}\varepsilon_{it-k-1} + \varepsilon_{it-1}\varepsilon_{it-k-1}) = 0$ per $k > 1$.

Utlizzando la forma matriciale, per l'individuo i-esimo si ha perciò una matrice quadrata e simmetrica di dimensione $(T-2) \times (T-2)$ così composta

$$V_i = E(\Delta\varepsilon_i \Delta\varepsilon_i') = \sigma_\varepsilon^2 \begin{bmatrix} 2 & -1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & -1 & 2 \end{bmatrix}. \tag{42}$$

Naturalmente, considerando il modello nella forma generale la matrice delle varianze e delle covarianze[13] è data da

$$V = I_N \otimes V_i. \tag{43}$$

Allo stesso modo si definisce la matrice $(T-2) \times C$ degli strumenti, dove $C = \sum_{j=1}^{T-2} j$

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i1} & y_{i2} & \dots & y_{iT-2} \end{bmatrix}, \tag{44}$$

dove ogni riga contiene gli strumenti validi per ciascun istante nel tempo $t = 3, 4, \dots, T$. Considerando tutte le osservazioni del modello, tale matrice è definita come

$$Z = \iota_N \otimes Z_i \tag{45}$$

ed ha dimensione $N(T-2) \times C$. Naturalmente, se gli strumenti sono validi, deve risultare $E(Z'\Delta\varepsilon) = 0$. Riscrivendo la (40) nella forma compatta si ha

$$\underset{N(T-2)\times 1}{\Delta Y_t} = \phi \underset{N(T-2)\times 1}{\Delta Y_{t-1}} + \underset{N(T-2)\times 1}{\Delta \varepsilon_t}, \tag{46}$$

dove ϕ è un parametro scalare. Il modello (46) è caratterizzato dalla presenza di correlazione tra l'errore ed i regressori, nonché dalla presenza di eteroschedasticità; Arellano e Bond (1991) risolvono il primo inconveniente strumentando l'equazione come segue

$$\underset{C\times 1}{Z'\Delta Y_t} = \phi \underset{C\times 1}{Z'\Delta Y_{t-1}} + \underset{C\times 1}{Z'\Delta \varepsilon_t}. \tag{47}$$

Per quanto riguarda l'eteroschedasticità, la matrice delle varianze e delle covarianze dipende strettamente dalla presenza di N individui e risulta essere

$$\begin{aligned} \Omega &= Var(Z'\Delta\varepsilon) \\ &= E(Z'\Delta\varepsilon\Delta\varepsilon' Z) \\ &= \sigma_\varepsilon^2 Z'VZ \\ &= \sigma_\varepsilon^2 Z'(I_N \otimes V_i)Z. \end{aligned} \tag{48}$$

Lo stimatore di Arellano-Bond è perciò uno stimatore GLS del tipo

$$\begin{aligned} \hat{\phi} &= (\Delta Y'_{t-1} Z \Omega^{-1} Z' \Delta Y_{t-1})^{-1} \Delta Y'_{t-1} Z \Omega^{-1} Z' \Delta Y_t \\ &= \{\Delta Y'_{t-1} Z[Z'(I_N \otimes V_i)Z]^{-1} Z' \Delta Y_{t-1}\}^{-1} \Delta Y'_{t-1} Z[Z'(I_N \otimes V_i)Z]^{-1} Z' \Delta Y_t. \end{aligned} \tag{49}$$

Tale stimatore è noto col nome STIMATORE DI ARELLANO-BOND ONE STEP CONSISTENT. Lo stimatore TWO STEP CONSISTENT invece è ottenibile sostituendo la matrice dei momenti secondi della popolazione $V_i = E(\Delta\varepsilon\Delta\varepsilon')$ con quella dei corrispondenti momenti secondi campionari data da $W_i = E(\Delta\hat{\varepsilon}\Delta\hat{\varepsilon}')$, dove $\hat{varepsilon}$ è ottenuto come residuo del modello (40) stimato attraverso lo stimatore (49). I due stimatori sono asintoticamente equivalenti per $N \to \infty$.

[13]Tale matrice ha dimensione $N(T-2) \times N(T-2)$.

3.2.2 Regressori esogeni

Inserendo nella trattazione anche i regressori esogeni l'equazione (39) si modifica nella seguente espressione

$$y_{it} = \phi y_{it-1} + X'_{it}\beta + \mu_i + \varepsilon_{it}, \tag{50}$$

dove X'_{it} ha $K-1$ colonne; in questo modo il numero totale dei parametri da stimare sia pari a K (tutte le componenti di β più lo scalare ϕ).

Anche in questo contesto si esprime il modello utilizzando le differenze prime in modo da determinare quali siano gli strumenti validi. Analiticamente si ottiene perciò

$$\Delta y_{it} = \phi \Delta y_{it-1} + \Delta X'_{it}\beta + \Delta\varepsilon_{it}, \tag{51}$$

dove gli effetti fissi sono rimossi. A questo punto occorre distinguere due casi:

1. Regressori predeterminati $\Rightarrow E(X'_{it}\varepsilon_{is}) \neq 0$ solo quando $t > s$. La matrice degli strumenti è analoga alla (44) con l'aggiunta di altri strumenti ottenibili dalla matrice dei regressori esogeni, infatti

$$Z_i = \begin{bmatrix} y_{i1} & X_{i1} & X_{i2} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & X_{i1} & X_{i2} & X_{i3} & \dots & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i1} & y_{i2} & \dots & y_{iT-2} & X_{i1} & X_{i2} & \dots & X_{T-1} \end{bmatrix}. \tag{52}$$

2. Regressori esogeni in senso stretto $\Rightarrow E(X'_{it}\varepsilon_{is}) = 0$ per ogni $t, s = 1, 2, \dots, T-2$. In questo caso le variabili $X_{i1}, X_{i2}, \dots, X_{iT-1}$ sono sempre tutti strumenti validi e vanno inseriti nelle righe della matrice Z_i.

Una volta determinate le matrici degli strumenti validi la procedura illustrata nella sezione 3.2.1 resta valida anche in quest'ambito.

Appendice

A-1 Proprietà delle matrici P e M

Matrice P

La matrice di proiezione P è definita come $P = (I_N \otimes P_\iota)$ con $P_\iota = \iota_T(\iota'_T\iota_T)^{-1}\iota'_T$. Essa risulta essere

quadrata: dato che $P_\iota = \iota_T(\iota'_T\iota_T)^{-1}\iota'_T$ è quadrata di dimensione $(T \times T)$

$$P_\iota = \frac{1}{T}\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix},$$

il prodotto $P = (I_N \otimes P_\iota)$ è esso stesso una matrice quadrata di dimensione $(NT \times NT)$.

diagonale a blocchi: in tutto ci sono N blocchi composti dalla matrice P_ι

$$P = \begin{bmatrix} P_\iota & 0 & \dots & 0 \\ 0 & P_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_\iota \end{bmatrix}.$$

simmetrica: poiché tutti i blocchi sono simmetrici, naturalmente risulta anche $P = P'$.

idempotente: dato che $P_\iota P_\iota = \iota_T(\iota_T'\iota_T)^{-1}\iota_T'\iota_T(\iota_T'\iota_T)^{-1}\iota_T' = \iota_T(\iota_T'\iota_T)^{-1}\iota_T' = P_\iota$ risulta

$$\begin{aligned} PP &= \begin{bmatrix} P_\iota & 0 & \dots & 0 \\ 0 & P_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_\iota \end{bmatrix} \begin{bmatrix} P_\iota & 0 & \dots & 0 \\ 0 & P_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_\iota \end{bmatrix} \\ &= \begin{bmatrix} P_\iota P_\iota & 0 & \dots & 0 \\ 0 & P_\iota P_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_\iota P_\iota \end{bmatrix} \\ &= \begin{bmatrix} P_\iota & 0 & \dots & 0 \\ 0 & P_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_\iota \end{bmatrix} = P \end{aligned}$$

Quando moltiplica una matrice in formato panel X di dimensione $(NT \times k)$, P ritorna la matrice $\bar{X}$ avente le stesse dimensioni della matrice data e contenente le sue **medie individuali** calcolate sulle colonne.

$$\begin{aligned} PX &= \underset{(NT\times NT)}{(I_N \otimes P_\iota)} \underset{(NT\times k)}{X} \\ &= \begin{bmatrix} \iota_T(\iota_T'\iota_T)^{-1}\iota_T' & 0 & \dots & 0 \\ 0 & \iota_T(\iota_T'\iota_T)^{-1}\iota_T' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \iota_T(\iota_T'\iota_T)^{-1}\iota_T' \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \\ &= \begin{bmatrix} \iota_T(\iota_T'\iota_T)^{-1}\iota_T' x_1 \\ \iota_T(\iota_T'\iota_T)^{-1}\iota_T' x_2 \\ \vdots \\ \iota_T(\iota_T'\iota_T)^{-1}\iota_T' x_N \end{bmatrix} \end{aligned}$$

Dato che $(\iota_T'\iota_T)^{-1}\iota_T' x_i = \dfrac{1}{T}\sum_{j=1}^{T} x_{ij}' = \bar{x}_i'$ (vettore riga k-dimensionale contenente le medie aritmetiche temporali relative all'i-esimo individuo), si ottiene

$$PX = \begin{bmatrix} \iota_T\bar{x}_1' \\ \iota_T\bar{x}_2' \\ \vdots \\ \iota_T\bar{x}_N' \end{bmatrix} = \bar{X}.$$

Dal punto di vista geometrico P si configura come la matrice delle proiezioni ortogonali sullo spazio generato da ι_T di tutte le variabili individuali $\underset{(N\times 1)}{y_i}$ e $\underset{(N\times k)}{x_i}$.

Matrice M

La matrice M è definita come $M = (I_N \otimes M_\iota)$ con $M_\iota = I_T - P_\iota = I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T'$. Anche la matrice M è

quadrata: dato che $M_\iota = I_T - P_\iota$ è quadrata di dimensione $(T \times T)$

$$M_\iota = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} - \frac{1}{T} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix},$$

il prodotto $P = (I_N \otimes M_\iota)$ è esso stesso una matrice quadrata di dimensione $(NT \times NT)$.

diagonale a blocchi: analogamente a quanto accadeva per la matrice P, anche in questo caso ci sono in tutto N blocchi composti dalla matrice M_ι

$$P = \begin{bmatrix} M_\iota & 0 & \dots & 0 \\ 0 & M_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_\iota \end{bmatrix}.$$

simmetrica: poiché sia I_t sia P_ι, sono simmetriche, tutti i blocchi di M sono simmetrici, quindi $M = M'$.

idempotente: dato che

$$\begin{aligned} M_\iota M_\iota &= [I_T - P_\iota][I_T - P_\iota] \\ &= I_T - P_\iota - P_\iota - P_\iota P_\iota \\ &= I_T - P_\iota \\ &= M_\iota, \end{aligned}$$

risulta

$$\begin{aligned} MM &= \begin{bmatrix} M_\iota & 0 & \dots & 0 \\ 0 & M_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_\iota \end{bmatrix} \begin{bmatrix} M_\iota & 0 & \dots & 0 \\ 0 & M_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_\iota \end{bmatrix} \\ &= \begin{bmatrix} M_\iota M_\iota & 0 & \dots & 0 \\ 0 & M_\iota M_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_\iota M_\iota \end{bmatrix} \\ &= \begin{bmatrix} M_\iota & 0 & \dots & 0 \\ 0 & M_\iota & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_\iota \end{bmatrix} = M \end{aligned}$$

Quando moltiplica una matrice in formato panel X di dimensione $(NT \times k)$, M ritorna la matrice $X - \bar{X}$ avente le stesse dimensioni della matrice data e, per ciascun individuo, contenente gli **scarti delle colonne dalle loro medie individuali**. Questo risultato è facilmente dimostrabile come segue considerando la matrice X di dimensione $NT \times k$:

$$\begin{aligned}
MX &= \underset{(NT\times NT)}{(I_N \otimes M_\iota)} \underset{(NT\times k)}{X} \\
&= \begin{bmatrix} I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' & 0 & \dots & 0 \\ 0 & I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \\
&= \begin{bmatrix} x_1 - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' x_1 \\ x_2 - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' x_2 \\ \vdots \\ x_N - \iota_T(\iota_T'\iota_T)^{-1}\iota_T' x_N \end{bmatrix} \\
&= \begin{bmatrix} x_1 - \iota_T \bar{x}_1 \\ x_2 - \iota_T \bar{x}_2 \\ \vdots \\ x_T - \iota_T \bar{x}_N \end{bmatrix} = X - \bar{X}.
\end{aligned}$$

Dal punto di vista della singola osservazione si ha perciò $x_{it} - \bar{x}_i$ che rappresenta lo scarto dalla media aritmetica individuale calcolata attraverso le diverse osservazioni nel tempo.

Dal punto di vista geometrico M si configura come la matrice della distanza tra i vettori colonna delle variabili individuali $\underset{(N\times 1)}{y_i}$ e $\underset{(N\times k)}{x_i}$ e le loro proiezioni ortogonali sullo spazio generato da ι_T.

Relazioni tra P e M

Date le proprietà delle matrici P ed M risulta:

$P + M = I_{NT} \Rightarrow M$ infatti equivale a $I_{NT} - P$,

$PM = 0 \Rightarrow PM = P(I_{NT} - P) = P - PP = P - P = 0.$

Naturalmente, per i singoli blocchi, vale $P_\iota + M_\iota = I_T$ e $P_\iota M_\iota = 0$.
Inoltre, valgono le seguenti relazioni

- $\iota_T' M_\iota = M_\iota \iota_T = 0$,
- $\iota_T' P_\iota = P_\iota \iota_T = \iota_T$.

Dati due numeri scalari c_1 e c_2 risulta

$$(c_1 P + c_2 M)^s = c_1^s P + c_2^s M,$$

quindi risulta facile ad esempio determinare

- la matrice inversa
$$(c_1 P + c_2 M)^{-1} = \frac{1}{c_1} P + \frac{1}{c_2} M.$$

La dimostrazione si basa sulle proprietà di idempotenza, somma e prodotto delle matrici P ed M, infatti

$$\begin{aligned}
(c_1 P + c_2 M)\left(\frac{1}{c_1} P + \frac{1}{c_2} M\right) &= \frac{c_1}{c_1} PP + \frac{c_2}{c_1} MP + \frac{c_1}{c_2} PM + \frac{c_2}{c_2} MM \\
&= P + M \\
&= I_{NT}
\end{aligned}$$

- la forma quadratica
$$(c_1P + c_2M)'(c_1P + c_2M) = c_1^2P + c_2^2M.$$
Anche in questo caso, sfruttando le proprietà di idempotenza, somma e prodotto delle matrici P ed M, si ottiene
$$\begin{aligned}(c_1P + c_2M)'(c_1P + c_2M) &= (c_1P + c_2M)^2 \\ &= c_1^2PP + c_2c_1MP + c_1c_2PM + c_2^2MM \\ &= c_1^2P + c_2^2M\end{aligned}$$

A-2 Determinazione dello stimatore ad effetti fissi

Data l'espressione dell'inversa di una matrice partizionata (di veda la nota 3), il blocco di Sud-Est si ottiene attraverso i seguenti passaggi
$$\begin{aligned}S_2 &= [X'X - X'(I_N \otimes \iota_T)\frac{1}{T}I_N(I_N \otimes \iota_T)'X]^{-1} \\ &= \{X'[I_{NT} - \frac{1}{T}(I_N \otimes \iota_T)(I_N \otimes \iota_T)']X\}^{-1} \\ &= \{X'[(I_N \otimes I_T) - \frac{1}{T}(I_N \otimes \iota_T\iota_T')]X\}^{-1} \\ &= \{X'[I_N \otimes (I_T - \frac{1}{T}\iota_T\iota_T')]X\}^{-1} \\ &= \{X'[I_N \otimes (I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T')]X\}^{-1} \\ &= [X'(I_N \otimes M_\iota)X]^{-1} \\ &= (X'MX)^{-1}.\end{aligned}$$
Una volta ottenuta questa quantità, l'equazione dello stimatore diventa
$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} S_1 & -\frac{1}{T}(I_N \otimes \iota_T)'XS_2 \\ -\frac{1}{T}S_2X'(I_N \otimes \iota_T) & S_2 \end{bmatrix} \begin{bmatrix} (I_N \otimes \iota_T)'Y \\ X'Y \end{bmatrix}$$
dove $S_1 = \frac{1}{T}I_N + \frac{1}{T}I_N(I_N \otimes \iota_T)'XS_2X'(I_N \otimes \iota_T)\frac{1}{T}I_N$. Svolgendo i prodotti
$$\begin{aligned}\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} &= \begin{bmatrix} \frac{1}{T}(I_N \otimes \iota_T)'Y + \frac{1}{T}(I_N \otimes \iota_T)'XS_2X'(I_N \otimes \iota_T)\frac{1}{T}(I_N \otimes \iota_T)'Y - \frac{1}{T}(I_N \otimes \iota_T)'XS_2X'Y \\ S_2X'Y - \frac{1}{T}S_2X'(I_N \otimes \iota_T)(I_N \otimes \iota_T)'Y \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T}(I_N \otimes \iota_T)'\left\{I_{NT} - XS_2X'\left[I_{NT} - \frac{1}{T}(I_N \otimes \iota_T\iota_T')\right]\right\}Y \\ S_2X'\left[I_{NT} - \frac{1}{T}(I_N \otimes \iota_T\iota_T')\right]Y \end{bmatrix}\end{aligned}$$
Poiché
$$\begin{aligned}I_{NT} - \frac{1}{T}(I_N \otimes \iota_T\iota_T') &= (I_N \otimes I_T) - [I_N \otimes \iota_T(\iota_T'\iota_T)^{-1}\iota_T] \\ &= I_N \otimes [I_T - \iota_T(\iota_T'\iota_T)^{-1}\iota_T] \\ &= I_N \otimes (I_T - P_\iota) \\ &= I_N \otimes M_\iota \\ &= M,\end{aligned}$$

lo stimatore diventa quello di cui all'equazione (15)

$$\begin{aligned}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} &= \begin{bmatrix}\frac{1}{T}(I_N\otimes\iota_T)'\left\{I_{NT}-XS_2X'M\right\}Y\\ S_2X'MY\end{bmatrix}\\ &= \begin{bmatrix}\frac{1}{T}(I_N\otimes\iota_T)'[Y-X(X'MX)^{-1}X'MY]\\ (X'MX)^{-1}X'MY\end{bmatrix}\\ &= \begin{bmatrix}\frac{1}{T}(I_N\otimes\iota_T)'(Y-X\hat{\beta})\\ (X'MX)^{-1}X'MY\end{bmatrix}.\end{aligned}$$

A-3 Test BP

Considerando la regressione ausiliaria $\hat{\varepsilon}^2 = Z\gamma+\eta$, l'indice di determinazione corrispondente è

$$R^2 = \frac{\hat{\varepsilon}^{2\prime}Z(Z'Z)^{-1}Z'\hat{\varepsilon}^2}{\hat{\varepsilon}^{2\prime}\hat{\varepsilon}^2}.$$

Sostituendo $\hat{\varepsilon}^2$ con $(\hat{\varepsilon}^2/\hat{\gamma}_0-1)$ in pratica si sottrae e successivamente si divide per la quantità costante $\hat{\gamma}_0$, quindi l'indice R^2 non cambia e si ha

$$\begin{aligned}R^2 &= \frac{(\hat{\varepsilon}^2/\hat{\gamma}_0-1)'Z(Z'Z)^{-1}Z'(\hat{\varepsilon}^2/\hat{\gamma}_0-1)}{(\hat{\varepsilon}^2/\hat{\gamma}_0-1)'(\hat{\varepsilon}^2/\hat{\gamma}_0-1)}\\ &= \frac{\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)'Z(Z'Z)^{-1}Z'\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)}{\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)'\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)}.\end{aligned}$$

Poiché $\hat{\varepsilon}\sim N(0,\sigma\Omega)$, sotto l'ipotesi nulla risulta $\hat{\varepsilon}/\sqrt{\hat{\gamma}_0}\sim N(0,I_n)$ per il TCL di Lindeberg-Lévy, quindi il denominatore della statistica test converge al valore 2 una volta diviso per l'ampiezza campionaria n, infatti

$$\begin{aligned}\frac{1}{n}\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)'\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right) &= \frac{1}{n}\frac{(\hat{\varepsilon}^2-\hat{\gamma}_0)'(\hat{\varepsilon}^2-\hat{\gamma}_0)}{\hat{\gamma}_0^2}\\ &= \frac{1}{n}\frac{\hat{\varepsilon}^{2\prime}\hat{\varepsilon}^2-n\hat{\gamma}_0^2}{\hat{\gamma}_0^2}\\ &= \frac{\frac{\hat{\varepsilon}^{2\prime}\hat{\varepsilon}^2}{n}}{\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}\right)^2}-1\end{aligned}$$

Questa espressione si configura come il rapporto tra il momento 4° ed il quadrato della varianza di $\hat{\varepsilon}$; nel caso di distribuzione normale del residuo tale rapporto converge al valore 3 quindi

$$\frac{1}{n}\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)'\left(\frac{\hat{\varepsilon}^2-\hat{\gamma}_0}{\hat{\gamma}_0}\right)\xrightarrow{\text{p}}3-1=2$$

Alla luce di questo risultato si ottiene

$$nR^2 = \frac{1}{2\hat{\gamma}_0{}^2}(\hat{\varepsilon}^2-\hat{\gamma}_0)'Z(Z'Z)^{-1}Z'(\hat{\varepsilon}^2-\hat{\gamma}_0).$$