

Econometria
Appunti del corso

Semestre Autunnale

Indice

I	Complementi al modello classico di regressione	3
1	La regressione scomposta	4
1.1	Il problema	4
1.2	Lo stimatore	4
1.3	Il modello ridotto	6
1.4	Interpretazione della trasformazione	8
II	La regressione generalizzata	11
2	Il modello generale	12
2.1	Introduzione	12
2.2	Il modello e le ipotesi	13
2.3	Quando V è conosciuta: i minimi quadrati generalizzati puri . .	14
2.4	Le conseguenze dell'applicazione dei <i>m.q.o.</i>	21
2.5	Quando V dipende da parametri sconosciuti	23
3	L'autocorrelazione	25
3.1	Il modello markoviano di primo ordine	25
3.2	Altre forme di autocorrelazione	32
III	La regressione stocastica	34
4	Elementi di analisi asintotica	35
4.1	La convergenza in probabilità	35
4.2	La convergenza in distribuzione	39
4.3	Analisi asintotica della regressione classica	41
5	Il modello di regressione stocastica	44
5.1	Il modello e le ipotesi	44
5.2	Quando i regressori e gli errori sono indipendenti	45
5.3	Quando sono correlati	45

IV	Modelli dinamici	50
6	Variabili endogene ritardate	51
6.1	Esempi economici	52
6.2	La stima dei modelli a variabile endogena ritardata	56
7	Modelli con ritardi distribuiti	63
7.1	Specificazione e proprietà. Moltiplicatori d’impatto, moltiplicatori dinamici, ritardo medio	63
7.2	Distribuzione dei ritardi finita	66
7.3	Distribuzione dei ritardi infinita	69
8	Serie macroeconomiche non stazionarie	77
8.1	Stazionarietà e non stazionarietà	77

Parte I

Complementi al modello classico di regressione

Capitolo 1

La regressione scomposta

1.1 Il problema

Nel quadro della regressione classica

$$y = X\beta + \varepsilon$$

con n osservazioni e K variabili esplicative scomponiamo le variabili esplicative ed i loro coefficienti in 2 gruppi:

- la matrice delle variabili esplicative

$$X = \left[\begin{array}{c|c} X_1 & X_2 \\ \hline (n \times K_1) & (n \times K_2) \end{array} \right]$$

- il vettore dei coefficienti

$$\beta = \left[\begin{array}{c} \beta_1 \\ \beta_2 \end{array} \right] \begin{array}{l} (K_1 \times 1) \\ (K_2 \times 1) \end{array}$$

Naturalmente, lo stimatore di β è lo stimatore abituale dei m.q.

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Supponiamo di essere interessati a β_2 (vettore $K_2 \times 1$). Evidentemente $\hat{\beta}_2$ è la seconda parte del vettore $\hat{\beta}$. Come esprimerlo analiticamente?

1.2 Lo stimatore

Sappiamo che $\hat{\beta}$ è la soluzione del sistema

$$X'X \hat{\beta} = X'y.$$

In forma scomposta possiamo riscrivere il precedente sistema come

$$\begin{aligned} \left[\begin{array}{c} X'_1 \\ X'_2 \end{array} \right] \left[\begin{array}{cc} X_1 & X_2 \end{array} \right] \left[\begin{array}{c} \hat{\beta}_1 \\ \hat{\beta}_2 \end{array} \right] &= \left[\begin{array}{c} X'_1 y \\ X'_2 y \end{array} \right] \\ \left[\begin{array}{cc} X'_1 X_1 & X'_1 X_2 \\ X'_2 X_1 & X'_2 X_2 \end{array} \right] \left[\begin{array}{c} \hat{\beta}_1 \\ \hat{\beta}_2 \end{array} \right] &= \left[\begin{array}{c} X'_1 y \\ X'_2 y \end{array} \right] \end{aligned}$$

$$X_1' X_1 \hat{\beta}_1 + X_1' X_2 \hat{\beta}_2 = X_1' y \quad (1.1)$$

$$X_2' X_1 \hat{\beta}_1 + X_2' X_2 \hat{\beta}_2 = X_2' y \quad (1.2)$$

Poiché per ipotesi $rg(X) = K \Leftrightarrow$ le K colonne di X sono linearmente indipendenti, anche le K_1 colonne di X_1 devono essere necessariamente linearmente indipendenti. Ciò significa che la matrice quadrata

$$\begin{matrix} X_1' X_1 \\ (K_1 \times K_1) \end{matrix}$$

è non singolare (invertibile). Dalla (1.1) risolvendo per $\hat{\beta}_1$ otteniamo:

$$\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' (y - X_2 \hat{\beta}_2) \quad (1.3)$$

Inserendo la (1.3) nella (1.2):

$$\begin{aligned} X_2' X_1 (X_1' X_1)^{-1} X_1' (y - X_2 \hat{\beta}_2) + X_2' X_2 \hat{\beta}_2 &= X_2' y \\ (X_2' X_2 - X_2' X_1 (X_1' X_1)^{-1} X_1' X_2) \hat{\beta}_2 &= X_2' y - X_2' X_1 (X_1' X_1)^{-1} X_1' y \\ X_2' (I - X_1 (X_1' X_1)^{-1} X_1') X_2 \hat{\beta}_2 &= X_2' (I - X_1 (X_1' X_1)^{-1} X_1') y \\ X_2' M_1 X_2 \hat{\beta}_2 &= X_2' M_1 y. \end{aligned} \quad (1.4)$$

Osservazioni sulla matrice M_1

- M_1 è della stessa forma e proprietà che la matrice

$$M = I - X (X' X)^{-1} X'$$

- soltanto che essa è formata a partire da X_1 e non da X . Come per M si dimostra facilmente che

- $M_1' = M_1$ e $M_1 M_1 = M_1$ (è simmetrica e idempotente)
- $rg(M_1) = tr(M_1) = n - K_1$
- $M_1 X_1 = 0$

Quando la matrice X è di rango K , la matrice

$$\begin{matrix} X_2' M_1 X_2 \\ (K_2 \times K_2) \end{matrix}$$

è non singolare¹ (lo deve essere necessariamente altrimenti il sistema non ammetterebbe una soluzione unica e ciò sarebbe in contraddizione con $\hat{\beta} = (X' X)^{-1} X' y$). Quindi, tornando all'equazione (1.4) abbiamo

$$\begin{aligned} X_2' M_1 X_2 \hat{\beta}_2 &= X_2' M_1 y \\ \hat{\beta}_2 &= (X_2' M_1 X_2)^{-1} X_2' M_1 y \end{aligned} \quad (1.5)$$

¹Dimostrazione: Supponiamo che $rg(X_2' M_1 X_2) < K_2$. Questo significa che il nucleo di $X_2' M_1 X_2$ non è vuoto e cioè che esiste un vettore $z \in \mathbb{R}^{K_2}$ diverso dal vettore nullo tale per cui

$$X_2' M_1 X_2 z = 0.$$

Inoltre:

$$\begin{aligned} V(\hat{\beta}_2) &= (X_2' M_1 X_2)^{-1} X_2' M_1 \underbrace{V(y)}_{\sigma^2 I} M_1 X_2 (X_2' M_1 X_2)^{-1} \\ &= \sigma^2 (X_2' M_1 X_2)^{-1} \end{aligned}$$

Si tratta quest'ultimo del blocco in posizione (2,2) della matrice $(X'X)^{-1}$. Abbiamo di fatto "solo" riscritto $V(\hat{\beta}_2)$ in funzione di X_2 e M_1

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{bmatrix} V(\hat{\beta}_1) & \dots \\ \dots & V(\hat{\beta}_2) \end{bmatrix}.$$

Punto sulla situazione:

- È possibile riscrivere il modello di regressione

$$y = X\beta + \varepsilon$$

a blocchi rispetto alle variabili esplicative di nostro interesse:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

- È possibile esprimere il coefficiente $\hat{\beta}_2$ dei m.q.o. utilizzando X_2 e la matrice M_1 come

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y$$

Osservazione: abbiamo solo esplicitato (riscritto) $\hat{\beta}_2$ in altro modo!

1.3 Il modello ridotto

Lo stimatore dei m.q. dei coefficienti di nostro interesse, $\hat{\beta}_2$, può essere ottenuto direttamente utilizzando i m.q. su un modello ridotto (o trasformato). Come?

Premoltiplicando la seguente uguaglianza per z' ed utilizzando la proprietà di idempotenza di M_1 , otteniamo la seguente forma quadratica

$$\begin{aligned} \underbrace{z' X_2' M_1}_{\tilde{z}'} \underbrace{M_1 X_2 z}_{\tilde{z}} &= 0 \\ \tilde{z}' \tilde{z} &= 0 \end{aligned}$$

Ma $\tilde{z}' \tilde{z} = 0 \Leftrightarrow \tilde{z} = 0$, da cui deduciamo che $M_1 X_2 z = 0$, ovvero

$$\begin{aligned} X_2 z - X_1 (X_1' X_1)^{-1} X_1' X_2 z &= 0 \\ \begin{bmatrix} X_1 & X_2 \end{bmatrix} \underbrace{\begin{bmatrix} - (X_1' X_1)^{-1} X_1' X_2 z \\ z \end{bmatrix}}_{y^* \neq 0} &= 0 \\ X y^* &= 0 \end{aligned}$$

$y^* \in \mathbb{R}^K$ è diverso dal vettore nullo poiché $z \neq 0$. Tuttavia $X y^* = 0$ è in contraddizione con l'ipotesi che $Rg(X) = K$.

1. Si modifica (trasforma) il modello di partenza

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

premultiplicando per M_1

$$\begin{aligned} M_1 y &= \underbrace{M_1 X_1 \beta_1}_{=0} + M_1 X_2 \beta_2 + M_1 \varepsilon \\ \underset{(n \times n)(n \times 1)}{M_1 y} &= \underset{(n \times n)(n \times K_2)}{M_1 X_2} \beta_2 + M_1 \varepsilon \\ \tilde{y} &= \tilde{X}_2 \beta_2 + \tilde{\varepsilon} \end{aligned} \quad (1.6)$$

2. Si applicano i minimi quadrati sul modello trasformato

$$\begin{aligned} \hat{\beta}_2 &= (\tilde{X}_2' \tilde{X}_2)^{-1} \tilde{X}_2' \tilde{y} \\ &= (X_2' M_1 M_1 X_2)^{-1} X_2' M_1 M_1 y \\ &= (X_2' M_1 X_2)^{-1} X_2' M_1 y \end{aligned}$$

ottenendo così esattamente la formula (1.5) precedente per $\hat{\beta}_2$.

Analizziamo i residui stimati di questa regressione:

$$\hat{\varepsilon} = \tilde{y} - \tilde{X}_2 \hat{\beta}_2.$$

Questi residui sono identici ai residui

$$\hat{\varepsilon} = y - X \hat{\beta}$$

della regressione del modello di partenza $y = X\beta + \varepsilon$. Infatti,

$$\begin{aligned} \hat{\varepsilon} &= y - X \hat{\beta} = y - X_1 \hat{\beta}_1 - X_2 \hat{\beta}_2 \\ &= y - \underbrace{X_1 (X_1' X_1)^{-1} X_1'}_{=\hat{\beta}_1, \text{ cf. formula (1.3)}} (y - X_2 \hat{\beta}_2) - X_2 \hat{\beta}_2 \\ &= (I - X_1 (X_1' X_1)^{-1} X_1') (y - X_2 \hat{\beta}_2) \\ &= M_1 (y - X_2 \hat{\beta}_2) = \tilde{y} - \tilde{X}_2 \hat{\beta}_2 = \hat{\varepsilon} \end{aligned}$$

Dall'uguaglianza degli errori stimati deduciamo che $\widetilde{SS} = \widetilde{\varepsilon}' \hat{\varepsilon}$ e $SS = \varepsilon' \hat{\varepsilon}$ sono pure identici.

Attenzione, consiglio pratico: se si stima direttamente il modello trasformato (1.6) al computer, utilizzando quindi \tilde{y} quale variabile dipendente e \tilde{X}_2 quale matrice delle variabili esplicative, occorre modificare la matrice stimata di varianza-covarianza di $\hat{\beta}_2$ in quanto il programma non può sapere che K_1 gradi di libertà sono stati persi a causa della trasformazione M_1 (che è di rango $n - K_1$). Il programma calcolerà

$$\widetilde{SS} = \widetilde{\varepsilon}' \hat{\varepsilon}$$

che come notato corrisponde al giusto SS ,

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y$$

che corrisponde al giusto $\hat{\beta}_2$. Ma per la stima di σ^2 il programma darà

$$\hat{\sigma}_{false}^2 = \frac{\widetilde{SS}}{n - K_2}$$

mentre il vero (giusto) stimatore è:

$$\hat{\sigma}_{vero}^2 = \frac{SS}{n - K} = \frac{\widetilde{SS}}{n - K_1 - K_2}.$$

Ne segue che occorre aggiustare l'output del programma in maniera tale da correggere i gradi di libertà

$$\hat{\sigma}^2 = \hat{\sigma}_{vero}^2 = \frac{\widetilde{SS}}{n - K_1 - K_2} = \frac{n - K_2}{n - K} \hat{\sigma}_{false}^2.$$

In maniera analoga,

$$\hat{V}_{false}(\hat{\beta}_2) = \hat{\sigma}_{false}^2 (X_2' M_1 X_2)^{-1}$$

andrà corretto con

$$\begin{aligned} \hat{V}(\hat{\beta}_2) &= \hat{\sigma}_{vero}^2 (X_2' M_1 X_2)^{-1} \\ &= \frac{n - K_2}{n - K} \hat{\sigma}_{false}^2 (X_2' M_1 X_2)^{-1} \\ &= \frac{n - K_2}{n - K} \hat{V}_{false}(\hat{\beta}_2) \end{aligned}$$

Regola: i risultati per \widetilde{SS} e $\hat{\beta}_2$ del modello trasformato sono Ok, quelli per $\hat{\sigma}^2$ e $\hat{V}(\hat{\beta}_2)$ vanno corretti!

1.4 Interpretazione della trasformazione

$$\tilde{y} = M_1 y = y - \underbrace{X_1 (X_1' X_1)^{-1} X_1' y}_{\hat{\beta}_1 \text{ della regressione } y = X_1 \beta_1 + \text{errore}}$$

Ora $(X_1' X_1)^{-1} X_1' y$ è lo stimatore dei minimi quadrati del modello parziale (incompleto) $y = X_1 \beta_1 + \text{errore}$.

- Si ha dunque che \tilde{y} è l'errore stimato del modello parziale, o in altre parole, y "epurato" dall'influenza parziale di X_1 :

$y -$ la parte di y che è spiegabile da X_1 .

- La stessa interpretazione vale per ogni colonna di X_2 :

$$M_1 X_2 = [M_1 (X_2)_{.,1} \mid M_1 (X_2)_{.,2} \mid \dots \mid M_1 (X_2)_{.,K_2}]$$

Ad ogni variabile esplicativa in X_2 (ad ogni sua colonna) tolgo la parte "spiegata" da X_1 . \Rightarrow nel modello trasformato, la variabile endogena \tilde{y} e i regressori \tilde{X}_2 sono ottenuti depurando le variabili originali y e X_2 dalla parte spiegata da X_1 .

Domanda: Quando lo stimatore di β_1 sul modello parziale (incompleto) ci dà il giusto stimatore $\hat{\beta}_1$?

Risposta:

- a) Abbiamo precedentemente osservato che il giusto stimatore può essere scritto (cf. (1.3))

$$\begin{aligned}\hat{\beta}_1 &= (X_1'X_1)^{-1} X_1' (y - X_2\hat{\beta}_2) \\ &= \underbrace{(X_1'X_1)^{-1} X_1'y}_{\text{Stimatore di } \hat{\beta}_1 \text{ sul modello parziale}} - \underbrace{(X_1'X_1)^{-1} X_1'X_2\hat{\beta}_2}_{\text{Fattore d'aggiustamento}}\end{aligned}$$

- b) Quando le colonne di X_1 sono ortogonali a quelle di X_2 (e quindi $X_1'X_2 = 0$) i due stimatori sono identici. Infatti

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1} X'y = \underbrace{\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}}_{\text{Matrice blocco diag. se } X_1'X_2=0} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} \\ &= \begin{bmatrix} (X_1'X_1)^{-1} X_1'y \\ (X_2'X_2)^{-1} X_2'y \end{bmatrix} \leftarrow \text{stimatori del modello parziale.}\end{aligned}$$

Esempio 1. Modello con costante

$$y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \varepsilon$$

dove $K_1 = 1$ e $X_1 = S$, il vettore somma d'ordine n .

$$\tilde{y} = \tilde{X}_2\hat{\beta}_2 + \tilde{\varepsilon}$$

con

$$\begin{aligned}\tilde{y} &= M_1 y \\ M_1 &= I - S(S'S)^{-1}S' = I - \frac{1}{n}SS' \\ M_1 y &= y - \frac{1}{n}SS'y = y - S\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \bar{y} \\ \bar{y} \\ \vdots \\ \bar{y} \end{bmatrix} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix}.\end{aligned}$$

Si tratta semplicemente del vettore degli scarti alla media. Il modello di partenza

$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_K x_{Ki} + \varepsilon_i$$

può essere stimato sul modello trasformato

$$y_i - \bar{y} = \beta_2 (x_{2i} - \bar{x}_2) + \beta_3 (x_{3i} - \bar{x}_3) + \dots + \beta_K (x_{Ki} - \bar{x}_K) + \tilde{\varepsilon}_i.$$

Inoltre

$$\begin{aligned}
 \hat{\beta}_1 &= (X_1' X_1)^{-1} X_1' \begin{pmatrix} y - X_2 \hat{\beta}_2 \\ \hat{\beta}_2 \end{pmatrix} \\
 &= \frac{1}{n} S'(y - X_2 \hat{\beta}_2) \\
 &= \bar{y} - \bar{x}' \hat{\beta}_2
 \end{aligned}$$

dove \bar{x} rappresenta il vettore $(K-1) \times 1$ delle medie delle variabili esplicative.

Nota bene: la regressione scomposta è utilizzata in vari ambiti dell'econometria e della statistica. Chi seguirà un corso di analisi di serie storiche vedrà che la regressione scomposta è utile per spiegare il concetto di "funzione di autocorrelazione parziale" utilizzato per l'identificazione dell'ordine (q, p) di un processo autoregressivo a media mobile (*ARMA*-process).

Parte II

La regressione generalizzata

Capitolo 2

Il modello generale

2.1 Introduzione

2.1.1 Breve ritorno sulla regressione classica

Durante il corso “Introduzione all’Econometria” abbiamo introdotto e studiato le proprietà del modello di regressione lineare. Per un singolo individuo (o istante temporale) i , la formulazione del modello lineare in forma vettoriale è data da

$$\underset{(1 \times 1)}{y_i} = \underset{(1 \times K)}{x'_i} \underset{(K \times 1)}{\beta} + \underset{(1 \times 1)}{\epsilon_i} \quad (2.1)$$

mentre per l’insieme degli n individui (periodi) abbiamo

$$\underset{(n \times 1)}{y} = \underset{(n \times K)}{X} \underset{(K \times 1)}{\beta} + \underset{(n \times 1)}{\epsilon} \quad (2.2)$$

Riprendiamo brevemente le ipotesi della regressione classica

1. sulle variabili esplicative: non stocastiche
2. sugli errori:
 - di valore atteso nullo
 - omoschedastici
 - non correlati (o indipendenti)
3. relazione d’indipendenza tra variabili esplicative ed errori.

Sotto queste ipotesi, abbiamo dimostrato che i *m.q.* forniscono degli stimatori efficienti (BLUE). Tuttavia, in molte applicazioni (sia in economia che in finanza) le ipotesi classiche non possono essere mantenute. In tal caso occorre porsi la domanda, se i minimi quadrati siano ancora efficienti e, qualora non fosse il caso, adottare nuovi metodi di stima. Questo è appunto il problema che affronteremo nella prima parte del semestre, analizzando in particolare due situazioni:

1. In primo luogo abbandoneremo l’ipotesi di omoschedasticità ed indipendenza degli errori in favore di una struttura di varianza-covarianza più generale \Rightarrow regressione generalizzata (prima parte).

2. In secondo luogo, ammetteremo che le variabili esplicative siano stocastiche, cioè esse stesse delle variabili aleatorie. Sappiamo già che quando le variabili esplicative sono aleatorie ma indipendenti dagli errori, i risultati della regressione classica restano validi. Invece, quando queste sono correlate con gli errori tutto crolla ed è necessario utilizzare metodi diversi \Rightarrow regressione stocastica (seconda parte).

2.1.2 Modelli econometrici che non soddisfano le ipotesi classiche

Quando e perché abbandonare l'ipotesi d'omoschedasticità e d'indipendenza degli errori?

1. con i dati trasversali (n individui allo stesso momento) si osservano frequentemente dei fenomeni di *eteroschedasticità*.
2. Con i dati temporali o più propriamente detto con le serie storiche si osserva che l'errore a un dato istante (al momento t) dipende dagli errori passati (in $t - 1, t - 2, \dots$): si parla di errori autocorrelati.

2.2 Il modello e le ipotesi

Ci interessiamo al modello di regressione

$$y = X\beta + \epsilon \quad (2.3)$$

con K variabili esplicative ed n osservazioni, munito delle seguenti ipotesi:

H1) Sulle variabili esplicative

H1a) X non stocastico.

H1b) $rg(X) = K$.

H2) Sugli errori

H2a) $E(\epsilon) = 0$.

H2b) $V(\epsilon) = \sigma^2 V$, dove V è una matrice qualsiasi definita positiva.

H3) X, ϵ mutualmente indipendenti.

Rispetto alla regressione classica la sola differenza è:

$$\begin{array}{ll} V(\epsilon) &= \sigma^2 I \quad \text{regressione classica} \\ V(\epsilon) &= \sigma^2 V \quad \text{regressione generalizzata} \end{array} \quad (2.4)$$

Esempio 2. (Eteroschedasticità pura)

Studiamo una funzione di risparmio del tipo:

$$S_i = a + bR_i + \epsilon_i, \quad i = 1, \dots, n \quad (2.5)$$

dove

- S_i è il risparmio della famiglia i .

- R_i è il reddito della famiglia.

Tipicamente il grafico degli n punti (S_i, R_i) è del tipo:

grafico

La varianza dell'errore non è costante: è una funzione crescente del reddito. Si può quindi scrivere:

$$V(\epsilon_i) = f(R_i) > 0 \text{ e } f'(R_i) > 0.$$

Una specificazione naturale potrebbe quindi essere

$$V(\epsilon_i) = \sigma^2 R_i^2,$$

dove σ^2 è un parametro di proporzionalità sconosciuto. Con un campione di n famiglie ed ammettendo $Cov(\epsilon_i, \epsilon_j) = 0$ per $i \neq j$ avremmo

$$V(\epsilon) = \sigma^2 \underbrace{\begin{bmatrix} R_1^2 & 0 & \dots & 0 \\ 0 & R_2^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R_n^2 \end{bmatrix}}_V \quad (2.6)$$

Da notare che in questo caso la matrice V è nota in quanto le variabili R_i sono osservate.

2.3 Quando V è conosciuta: i minimi quadrati generalizzati puri

Quando la matrice V è conosciuta (come nel precedente esempio) il metodo di stima prende il nome di minimi quadrati generalizzati puri, o brevemente *m.q.g. puri*. L'idea fondamentale è molto semplice. La presentiamo prendendo spunto dall'Esempio 2.

1. L'idea

Sappiamo risolvere questo problema quando la matrice $V = I$. Se fosse possibile, tramite una trasformazione appropriata, ricondurci a tale caso, il nostro problema sarebbe quindi risolto! Sul modello trasformato infatti, potremmo applicare i *m.q.* ordinari. Nell'esempio 2, la varianza per l'individuo i è uguale a $V(\epsilon_i) = \sigma^2 R_i^2$ e lo *s.q.m.* $\sqrt{V(\epsilon_i)} = \sigma R_i$. Dividiamo ogni osservazione per il suo *s.q.m.* (a meno di σ):

$$\begin{aligned} S_i &= a + bR_i + \epsilon_i \\ \frac{S_i}{R_i} &= a \frac{1}{R_i} + b + \frac{\epsilon_i}{R_i} \end{aligned}$$

(tasso di risparmio)

Questa è una regressione semplice dove:

- S_i/R_i è la variabile dipendente, spiegata
- $1/R_i$ è la variabile esplicativa
- b è la costante (che corrisponde alla propensione marginale al risparmio)
- a è il coefficiente della variabile esplicativa

e gli errori hanno le proprietà seguenti:

- $\tilde{\epsilon}_i = \epsilon_i/R_i$
- $E(\tilde{\epsilon}_i) = 1/R_i E(\epsilon_i) =$
- $V(\tilde{\epsilon}_i) = 1/R_i^2 V(\epsilon_i) = \quad \forall i \text{ costante!} \Rightarrow \text{omoschedasticità}$
- $Cov(\tilde{\epsilon}_i, \tilde{\epsilon}_j) = 1/(R_i R_j) Cov(\epsilon_i, \epsilon_j) = \quad i \neq j$

Dunque il modello così trasformato soddisfa appieno le ipotesi della regressione classica \Rightarrow applico su di esso lo stimatore dei *m.q.o.*!

2. L'esempio formalizzato

Vogliamo ora riscrivere quanto fatto nel punto precedente in forma matriciale per poi coglierne la struttura. Chiamiamo dunque

$$y = \begin{bmatrix} S_1 \\ \vdots \\ S_n \end{bmatrix}_{(n \times 1)} \quad X = \begin{bmatrix} 1 & R_1 \\ \vdots & \vdots \\ 1 & R_n \end{bmatrix}_{(n \times 2)} \quad \beta = \begin{bmatrix} a \\ b \end{bmatrix}_{(2 \times 1)} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}_{(n \times 1)}$$

$$\Rightarrow y = X\beta + \epsilon$$

Inoltre sappiamo che $E(\epsilon) = 0$ e $V(\epsilon) = \sigma^2 \begin{bmatrix} R_1^2 & & 0 \\ & \ddots & \\ 0 & & R_n^2 \end{bmatrix} = \sigma^2 V$.

Che trasformazione abbiamo utilizzato in precedenza al fine di applicare il metodo dei *m.q.* ordinari? Ogni componente del vettore y è stata divisa per il reddito corrispondente (o moltiplicata per $\frac{1}{R_i}$). Come possiamo

esprimere questa operazione utilizzando il calcolo matriciale? Basta *pre-moltiplicare* il vettore y per la matrice diagonale, notata P , avente $\frac{1}{R_i}$ in posizione (i, i) :

$$\underset{(n \times 1)}{\tilde{y}}, \underset{(n \times n)}{P} = \begin{bmatrix} \frac{1}{R_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{R_n} \end{bmatrix} \Rightarrow \tilde{y} = Py.$$

Lo stesso per la parte destra dell'equazione. Il modello trasformato è dunque:

$$\underset{(n \times n)(n \times 1)}{P} \underset{(n \times 1)}{y} = \underset{(n \times n)(n \times 2)}{P} \underset{(n \times 2)}{X\beta} + \underset{(n \times n)(n \times 1)}{P} \underset{(n \times 1)}{\epsilon}$$

La i -esima riga:

$$\frac{S_i}{R_i} = \left(\frac{1}{R_i}, \frac{R_i}{R_i} \right) \beta + \frac{\epsilon_i}{R_i}$$

Il modello trasformato con la matrice non-singolare P è:

$$\tilde{y} = \tilde{X}\beta + \tilde{\epsilon}$$

dove dalle proprietà di ϵ ricaviamo che

- $E(\tilde{\epsilon}) = E(P\epsilon) = 0$,
- $V(\tilde{\epsilon}) = V(P\epsilon) = PV(\epsilon)P' =$

$$\begin{aligned} &= \sigma^2 \begin{bmatrix} \frac{1}{R_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{R_n} \end{bmatrix} \begin{bmatrix} R_1^2 & & 0 \\ & \ddots & \\ 0 & & R_n^2 \end{bmatrix} \begin{bmatrix} \frac{1}{R_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{R_n} \end{bmatrix} \\ &= \sigma^2 I_n, \end{aligned}$$

che sono le proprietà della regressione classica.

2.3.1 Gli stimatori e le loro proprietà

1. Gli stimatori dei *m.q.g.* come stimatori dei *m.q.o.* sul modello trasformato

Il modello iniziale è il seguente

$$\underset{(n \times 1)}{y} = \underset{(n \times K)(K \times 1)}{X} \underset{(K \times 1)}{\beta} + \underset{(n \times 1)}{\epsilon},$$

con le ipotesi $H1 - H3$. In particolare abbiamo

- $E(\epsilon) = 0$,
- $V(\epsilon) = E(\epsilon\epsilon') = \underset{\text{def. pos.}}{\sigma^2 V}$.

Sappiamo che per ogni matrice def. positiva V esiste una matrice non-singolare P tale che

$$PVP' = I.$$

Trasformiamo dunque il modello *premultiplicando* per P

$$\begin{aligned} \underbrace{Py}_{\tilde{y}} &= \underbrace{PX}_{\tilde{X}} \beta + \underbrace{P\epsilon}_{\tilde{\epsilon}} \\ \tilde{y} &= \tilde{X}\beta + \tilde{\epsilon} \end{aligned} \quad (2.7)$$

Il modello trasformato (2.7) ha un vettore d'errori $\tilde{\epsilon}$ tale per cui

- $E(\tilde{\epsilon}) = 0$
- $V(\tilde{\epsilon}) = \sigma^2 I$.

Questo modello trasformato soddisfa tutte le ipotesi della regressione classica. Appliciamo dunque su di esso i *m.q.o.*! Sul modello trasformato i risultati della stima saranno pertanto quelli della regressione classica e cioè:

- lo stimatore di beta, chiamiamolo β^* :

$$\beta^* = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

- la sua vera varianza

$$V(\beta^*) = \sigma^2(\tilde{X}'\tilde{X})^{-1}$$

- il vettore d'errori stimati sul modello trasformato:

$$\hat{\tilde{\epsilon}} = \tilde{y} - \tilde{X}\beta^*$$

- la somma dei quadrati degli errori (del modello trasformato!)

$$\tilde{S}S = \tilde{\epsilon}'\hat{\tilde{\epsilon}} = \tilde{y}'\tilde{y} - \beta^{*'}\tilde{X}'\tilde{y} = \tilde{y}'\tilde{M}\tilde{y}$$

dove

$$\tilde{M} = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'.$$

- la stima corretta di σ^2

$$\hat{\sigma}^2 = \frac{\tilde{S}S}{n - K}$$

- la stima della matrice di varianza-covarianza di β^*

$$\hat{V}(\beta^*) = \hat{\sigma}^2(\tilde{X}'\tilde{X})^{-1}$$

2. Lo stimatore dei *m.q.g.*

Lo stimatore β^* può essere riscritto in termini di X , y e della matrice P nel seguente modo:

$$\begin{aligned} \beta^* &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \\ &= (X'P'PX)^{-1}X'P'Py \end{aligned}$$

Ora, sapendo che $PVP' = I$ otteniamo successivamente

$$\begin{aligned}VP' &= P^{-1} \\ P' &= V^{-1}P^{-1} \\ P'P &= V^{-1}\end{aligned}$$

Dunque per lo stimatore β^* otteniamo, in termini delle variabili di partenza:

- lo stimatore

$$\beta^* = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (2.8)$$

- La vera varianza di β^*

$$V(\beta^*) = \sigma^2(X'V^{-1}X)^{-1}$$

- Il vettore d'errori stimati sul modello di partenza

$$\hat{\epsilon} = y - X\beta^*$$

La relazione con gli errori stimati sul modello trasformato è la seguente:

$$\hat{\tilde{\epsilon}} = \tilde{y} - \tilde{X}\beta^* = Py - PX\beta^* = P(y - X\beta^*) = P\hat{\epsilon}$$

- sul modello trasformato avevamo:

$$\tilde{S}\tilde{S} = \hat{\tilde{\epsilon}}'\hat{\tilde{\epsilon}} = \hat{\epsilon}'P'P\hat{\epsilon} = \hat{\epsilon}'V^{-1}\hat{\epsilon}$$

Lo stimatore β^* definito dalla formula (2.8) è chiamato stimatore dei minimi quadrati generalizzati (*m.q.g.*). Esso è così chiamato perché ottenuto minimizzando la somma dei quadrati generalizzati

$$\min_{\beta} \underbrace{(y - X\beta)'V^{-1}(y - X\beta)}_{:=SSG} \quad (2.9)$$

la cui soluzione è appunto il nostro β^* . Infatti

$$\begin{aligned}SSG &= (y - X\beta)'V^{-1}(y - X\beta) \\ &= y'V^{-1}y - y'V^{-1}X\beta - \beta'X'V^{-1}y + \beta'X'V^{-1}X\beta \\ &= y'V^{-1}y - \underbrace{2y'V^{-1}X\beta}_{\text{forma lineare}} + \underbrace{\beta'X'V^{-1}X\beta}_{\text{forma quadratica}} \\ \frac{\partial SSG}{\partial \beta} &= -2X'V^{-1}y + 2X'V^{-1}X\beta \stackrel{!}{=} 0 \\ &\text{da cui risolvendo rispetto a } \beta \\ \beta^* &= (X'V^{-1}X)^{-1}X'V^{-1}y\end{aligned}$$

Noi lo abbiamo ricavato utilizzando un'argomentazione diversa. È utile esprimere la quantità $\tilde{S}S$ in termini delle variabili di partenza y , X :

$$\tilde{S}S = SSG = y'V^{-1}y - \beta^{*'}(X'V^{-1}y)$$

$$\begin{aligned}\tilde{M} &= I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}' \\ &= I - PX(X'V^{-1}X)^{-1}X'P'\end{aligned}$$

$$\begin{aligned}\tilde{S}S &= \tilde{y}'\tilde{M}\tilde{y} = y'P'\tilde{M}Py \\ &= y'P'Py - y'P'PX(X'V^{-1}X)^{-1}X'P'Py \\ &= y'V^{-1}y - y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y\end{aligned}$$

Tabella 2.1: Specchietto riassuntivo

$m.q.o.$ su $\tilde{y} = \tilde{X}\beta + \tilde{\epsilon}$	$m.q.g.$ su $y = X\beta + \varepsilon$
$\beta^* = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$	$\beta^* = (X'V^{-1}X)^{-1}X'V^{-1}y$
$V(\beta^*) = \sigma^2(\tilde{X}'\tilde{X})^{-1}$	$V(\beta^*) = \sigma^2(X'V^{-1}X)^{-1}$
$\begin{aligned}\tilde{S}S &= \tilde{\epsilon}'\tilde{\epsilon} = \tilde{y}'\tilde{M}\tilde{y} \\ &= \tilde{y}'(I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}')\tilde{y}\end{aligned}$	$\begin{aligned}SSG &= \tilde{\epsilon}'V^{-1}\tilde{\epsilon} \\ &= y'(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})y\end{aligned}$
$\hat{\sigma}^2 = \tilde{S}S/(n - K)$	$\hat{\sigma}^2 = SSG/(n - K)$
$\hat{V}(\beta^*) = \hat{\sigma}^2(\tilde{X}'\tilde{X})^{-1}$	$\hat{V}(\beta^*) = \hat{\sigma}^2(X'V^{-1}X)^{-1}$

3. Proprietà di β^*

- β^* è corretto! Infatti:

$$\beta^* = \underbrace{(X'V^{-1}X)^{-1}X'V^{-1}}_A y = Ay$$

e poiché $E(y) = X\beta$, $V(y) = \sigma^2V$, otteniamo $E(\beta^*) = AE(y) = \underbrace{AX}_I \beta = \beta$.

$$V(\beta^*) = AV(y)A' = \sigma^2AVA' = \sigma^2(X'V^{-1}X)^{-1}.$$

- È efficiente (BLUE)!

Dimostrazione: Condideriamo un qualsiasi altro stimatore lineare $\beta^\# = A^\#y$ e mostriamo che per $\beta^\#$ corretto

$$V(\beta^\#) > V(\beta^*)$$

Ora:

$$\begin{aligned} E(\beta^\#) &= A^\# E(y) = A^\# X \beta \\ V(\beta^\#) &= \sigma^2 A^\# V A^{\#'} \end{aligned}$$

La condizione di correttezza di $\beta^\#$ impone che $A^\# X = I$. Poniamo inoltre $A^\# = A + C$. Otteniamo:

$$\begin{aligned} V(\beta^\#) &= \sigma^2 (A + C) V (A' + C') \\ &= \sigma^2 \underbrace{A V A'}_{V(\beta^*)} + \underbrace{\sigma^2 A V C' + \sigma^2 C V A'}_{\text{nulli quando } \beta^\# \text{ è corretto}} + \underbrace{\sigma^2 C V C'}_{\text{def. non neg.}} \\ &\text{poiché} \\ A^\# X &= I \Rightarrow \underbrace{A X}_I + C X = I \Rightarrow C X = 0 \\ A V C' &= (X' V^{-1} X)^{-1} X' V^{-1} V C' = (X' V^{-1} X)^{-1} \underbrace{X' C'}_{=0} = 0 \end{aligned}$$

4. Stima corretta di σ^2

Definiamo $\hat{\sigma}^2$ come lo stimatore di σ^2 del modello trasformato, quindi

$$\hat{\sigma}^2 = \frac{SSG}{n - K} \text{ dove } SSG = \tilde{S} \tilde{S}$$

$$\begin{aligned} SSG &= y' \underbrace{[V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}]}_Q y \\ &= y' Q y \\ E(SSG) &= ? \end{aligned}$$

Sappiamo che $QX = 0$, da cui segue $Qy = Q(X\beta + \epsilon) = Q\epsilon$.
Quindi

$$SSG = y' Q y = \epsilon' Q \epsilon \text{ è una } FQ \text{ in } \epsilon \sim N(0, \sigma^2 V)$$

$$\begin{aligned} E(SSG) &= E(\epsilon' Q \epsilon) = \text{tr} Q V(\epsilon) = \sigma^2 \text{tr} Q V \\ &= \sigma^2 \text{tr}(I_n - V^{-1} X (X' V^{-1} X)^{-1} X') \\ &= \sigma^2 \text{tr}(I_n) - \sigma^2 \text{tr}(I_K) = \sigma^2 (n - K) \end{aligned}$$

Definendo $\hat{\sigma}^2 = \frac{SSG}{n - K}$ otteniamo uno stimatore corretto di σ^2 :

$$E(\hat{\sigma}^2) = \sigma^2.$$

Di conseguenza:

$$\hat{V}(\beta^*) = \hat{\sigma}^2 (\tilde{X}' \tilde{X})^{-1} = \hat{\sigma}^2 (X' V^{-1} X)^{-1}$$

è uno stimatore corretto della $V(\beta^*)$.

2.3.2 L'induzione statistica

Sotto l'ipotesi di normalità degli errori, tutte le procedure di test sul modello trasformato sono rigorosamente identiche a quelle della regressione classica. Le statistiche t per un'ipotesi semplice o F per un'ipotesi congiunta date in termini delle variabili trasformate si possono sempre esprimere in termini delle variabili di partenza. In particolare, dato $\epsilon \sim N(0, \sigma^2 V)$ segue:

- $\beta^* \sim N(\beta, \sigma^2 (X'V^{-1}X)^{-1})$.
- $\frac{SSG}{\sigma^2} \sim \chi_{n-K}^2$.
- β^* e SSG sono indipendenti.

Se chiamassimo \tilde{x}^{ij} l'elemento in posizione (i, j) della matrice inversa $(X'V^{-1}X)^{-1}$ allora

$$\frac{\beta_k^* - \beta_k}{\sqrt{\hat{\sigma}^2 \tilde{x}^{kk}}} \sim t_{n-K} \quad (2.10)$$

2.3.3 La previsione

Il problema:

- È dato un campione $y = X\beta + \epsilon$ di n osservazioni.
- Per un individuo s al di fuori del campione vogliamo prevedere y_s (il suo comportamento) date le sue caratteristiche contenute in x'_s .

Il miglior previsore lineare di y_s è tra tutti i previsori lineari e corretti¹ quello la cui varianza dell'errore di previsione è minima. Si ottiene:

1. Nella regressione classica β è stimato per mezzo dei *m.q.o.* e

$$y_s^p = x'_s \hat{\beta}$$

2. Nella regressione generalizzata, β è stimato per mezzo dei *m.q.g.* Quando l'errore dell'individuo s non è correlato con gli errori del campione,

$$y_s^p = x'_s \beta^*$$

3. Salvo nel caso di eteroschedasticità pura, nella regressione generalizzata l'errore dell'individuo s sarà correlato con gli errori del campione. Chiamiamo $\sigma^2 R'_{(1 \times n)} = Cov(\epsilon_s, \epsilon)$. Il miglior previsore lineare è dato da

$$y_s^p = x'_s \beta^* + Cov(\epsilon_s, \epsilon) V(\epsilon)^{-1} \hat{\epsilon}$$

2.4 Le conseguenze dell'applicazione dei *m.q.o.*

Vogliamo analizzare le conseguenze dell'applicazione dei minimi quadrati ordinari ad un modello con errori correlati. Poiché per ipotesi $rg(X) = K$ i *m.q.o.* sono calcolabili:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Che proprietà hanno?

¹Ricordiamo che un previsore y_s^p di y_s è corretto se $E(y_s^p) = E(y_s)$ o equivalentemente se l'errore di previsione $\epsilon_s^p = y_s - y_s^p$ è nullo in valore atteso.

2.4.1 Correttezza e perdita d'efficienza

- $\hat{\beta}$ è corretto.
- $\hat{\beta}$ non è efficiente: $V(\hat{\beta}) \geq V(\hat{\beta}^*)$

2.4.2 Stima convergente della varianza dello stimatore dei *m.q.o.*

La vera matrice delle varianze covarianze di $\hat{\beta}$ non è quella data dalla regressione classica $\sigma^2(X'X)^{-1}$ bensì

$$V(\hat{\beta}) = V((X'X)^{-1}X'y) = \sigma^2(X'X)^{-1}X'VX(X'X)^{-1} \quad (2.11)$$

Stimando la varianza di $\hat{\beta}$ con $V(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$ non si ottiene una stima corretta per 2 motivi:

1. $\hat{\sigma}^2 = \frac{SS}{n-K}$ non è uno stimatore corretto di σ^2 .
2. $(X'X)^{-1}$ dovrebbe essere rimpiazzato con $(X'X)^{-1}X'VX(X'X)^{-1}$.

⇒ Conseguenza: Utilizzare $\hat{\sigma}^2(X'X)^{-1}$ quale stimatore della varianza di $\hat{\beta}$ quando $V(\epsilon) \neq \sigma^2 I$ è sbagliato: i test sono falsi!

Osservazione 1. Nel caso della *regressione eteroschedastica*, cioè nella situazione in cui

$$V(\epsilon_i) = \sigma_i^2 \text{ e } Cov(\epsilon_i, \epsilon_j) = 0, \quad i \neq j$$

da cui

$$V(\epsilon) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix} = V \text{ matrice diagonale,}$$

è tuttavia possibile ottenere dei test approssimativamente esatti utilizzando quale stima della matrice delle varianze covarianze

$$\begin{aligned} \hat{V}(\hat{\beta}) &= (X'X)^{-1}X'\hat{V}X(X'X)^{-1}, \\ \text{con } \hat{V} &= \begin{bmatrix} \hat{\epsilon}_1^2 & & 0 \\ & \ddots & \\ 0 & & \hat{\epsilon}_n^2 \end{bmatrix}. \end{aligned}$$

Quest'ultima formula permette (nel caso della regressione eteroschedastica) di eseguire test asintoticamente corretti.

2.4.3 Equivalenza fra *m.q.o.* e *m.q.g.*

Esistono casi in cui $\hat{\beta} = \beta^*$ (per qualsiasi vettore y)

- Caso triviale $V = I$.
- Caso generale: Si può dimostrare che $\hat{\beta} = \beta^*$ se e solo se esiste una matrice non singolare A tale che

$$\begin{matrix} V & X \\ (n \times n) & (n \times K) \end{matrix} = \begin{matrix} X & A \\ (n \times K) & (K \times K) \end{matrix} \quad (2.12)$$

2.5 Quando V dipende da parametri sconosciuti

Quando V è completamente arbitraria, il problema non ha soluzioni. Infatti V , matrice d'ordine n , possiede $n(n+1)/2$ elementi. L'economista è interessato a strutture di varianze covarianze che dipendono da un numero fisso di parametri sconosciuti. Chiamiamo con m il numero di parametri sconosciuti contenuti nel vettore $\theta_{(m \times 1)}$. Ipotizziamo quindi che conoscendo θ si possa costruire V per ogni n . Designiamo quindi la matrice delle varianze covarianze di ϵ con $V(\epsilon) = \sigma^2 V(\theta)$.

Esempio 3. Struttura di equicorrelazione

- $E(\epsilon_i) = 0 \quad \forall i$.
- $V(\epsilon_i) = \sigma^2 \quad \forall i$.
- $Cov(\epsilon_i, \epsilon_j) = \theta \sigma^2, \quad i \neq j, \quad 0 \leq \theta < 1$
- Con un campione di n osservazioni, abbiamo dunque

$$V(\epsilon)_{(n \times n)} = \sigma^2 \underbrace{\begin{bmatrix} 1 & \theta & \dots & \theta \\ \theta & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \theta \\ \theta & \dots & \theta & 1 \end{bmatrix}}_{V(\theta)} \quad (2.13)$$

Strategia per la stima:

1. Prima possibilità

Siccome sappiamo risolvere il problema di stima quando θ è conosciuto, l'idea naturale è quella di stimare θ in una prima tappa per poi applicare lo stimatore dei *m.q.g.*. Questo procedimento è chiamato *minimi quadrati generalizzati a due tappe* (oppure *m.q.g. realizzabili*).

2. Seconda possibilità

Quale alternativa ai *m.q.g.* a due tappe si potrebbe immaginare di stimare *simultaneamente* tutti i parametri sconosciuti del modello che sono

$$\beta_{(K \times 1)}, \quad \sigma^2_{(1 \times 1)}, \quad \theta_{(m \times 1)}$$

utilizzando il metodo di massima verosimiglianza ($MLE = \text{Maximum Likelihood Estimation}$).

2.5.1 I *m.q.g.* a due tappe

Prima tappa: Si cerca uno stimatore convergente di θ . Non c'è una ricetta universale: in numerosi casi la stima del modello con i minimi quadrati ordinari permette di stimare θ come funzione degli errori stimati.

Seconda tappa: Si utilizzano i minimi quadrati generalizzati con $V(\hat{\theta})$ al posto di $V(\theta)$.

Proprietà: Asintoticamente (con $n \rightarrow \infty$) lo stimatore a due tappe possiede la medesima efficienza dello stimatore dei *m.q.g.* puri (cioè come se conoscessimo il vero θ).

2.5.2 Massimo di Verosimiglianza

Assumendo la normalità degli errori $\epsilon \sim N(0, \sigma^2 V(\theta))$ otteniamo grazie alla proprietà della legge normale che $y \sim N(X\beta, \sigma^2 V(\theta))$. La verosimiglianza, per un campione di n osservazioni dipende dai parametri sconosciuti β, σ^2, θ e si scrive

$$L(y \mid \beta, \sigma^2, \theta) = (2\pi\sigma^2)^{-n/2} \det[V(\theta)]^{-1/2} \exp\left(-\frac{1}{2}(y - X\beta)' \frac{1}{\sigma^2} V(\theta)^{-1} (y - X\beta)\right)$$

Gli stimatori di massima verosimiglianza sono soluzioni del programma

$$\max_{\{\beta, \sigma^2, \theta\}} L(y \mid \beta, \sigma^2, \theta)$$

dal quale si ricavano le condizioni di primo ordine

$$\begin{cases} \frac{\partial L}{\partial \sigma^2} = 0 \\ \frac{\partial L}{\partial \beta} = 0 \\ \frac{\partial L}{\partial \theta} = 0 \end{cases}$$

Questo sistema non possiede una soluzione analitica e quindi è necessario utilizzare un algoritmo numerico di risoluzione. Un'importante proprietà dello stimatore di massima verosimiglianza consiste nella sua efficienza asintotica.

Capitolo 3

L'autocorrelazione

Questo capitolo è presentato nell'ambito delle serie storiche (o cronologiche). L'autocorrelazione è un fenomeno proprio delle serie storiche. Essa esprime l'idea che l'errore a un dato istante dipende in parte dagli errori passati. Ci sono svariati modelli per descrivere errori autocorrelati. In questo capitolo tratteremo un modello molto diffuso in econometria, il modello autoregressivo di primo ordine: l'errore al tempo t dipende dall'errore al tempo $t - 1$.

3.1 Il modello markoviano di primo ordine

Il modello markoviano di primo ordine è anche detto modello *autoregressivo* di primo ordine, abbreviato $AR(1)$. Vediamo ora le ipotesi di questo modello.

3.1.1 Ipotesi e proprietà

La regressione abituale

$$\left\{ \begin{array}{l} y_t = x_t' \beta + \epsilon_t \quad t = 1, \dots, n \\ \epsilon_t = \underbrace{\rho \epsilon_{t-1}}_{\text{parte regressiva}} + \underbrace{u_t}_{\text{parte residua}} \\ |\rho| < 1 \text{ e } u_t \sim iid(0, \sigma^2) \end{array} \right.$$

dove u_t rappresenta un residuo aleatorio anch'esso non osservabile, $iid(0, \sigma^2)$. Vale dunque

$$\begin{aligned} E(u_t) &= 0 \quad \forall t \\ E(u_t^2) &= \sigma^2 \quad \forall t \\ Cov(u_t, u_s) &= 0 \quad \forall t \neq s \end{aligned}$$

Interpretazione ed osservazioni

1. Il termine "modello markoviano" non è riferito all'equazione

$$y_t = x_t' \beta + \epsilon_t$$

ma bensì alla struttura autoregressiva del suo termine d'errore

$$\epsilon_t = \rho \epsilon_{t-1} + u_t.$$

Nella seconda parte del corso, quando tratteremo i modelli dinamici, studieremo le proprietà dello stimatore dei minimi quadrati sul modello a variabile endogena ritardata

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t$$

con $\varepsilon_t \sim iid(0, \sigma^2)$. Come potete vedere, questo non è altro che il modello markoviano di ordine 1 appena definito applicato alla variabile endogena y .

2. Il processo $\epsilon_t = \rho\epsilon_{t-1} + u_t$ è senza memoria, nel senso che tutta l'informazione disponibile in $t-1$ al fine di spiegare ϵ_t è riassunta in ϵ_{t-1} . In altre parole, tutta la storia antecedente a ϵ_{t-1} non serve a nulla per spiegare ϵ_t : quello che capita nel periodo t dipende unicamente¹ dallo stato attuale di ϵ_{t-1} non dalla storia antecedente $\epsilon_{t-2}, \epsilon_{t-3}, \dots$.

grafico

3. Per ρ positivo (caso tipico in economia) il processo $\{\epsilon_t\}_{t \in \mathbb{Z}}$ presenta una persistenza di valori dello stesso segno: $P(\epsilon_t > 0 \mid \epsilon_{t-1} > 0) > 0.5$.
4. Sostituzione per ricorrenza

Sostituendo recursivamente otteniamo la rappresentazione di ϵ_t in funzione degli errori u_t :

$$\begin{aligned} \epsilon_t &= u_t + \rho \epsilon_{t-1} \\ &\quad \uparrow \\ &\quad \epsilon_{t-1} = u_{t-1} + \rho \epsilon_{t-2} \\ &= u_t + \rho u_{t-1} + \rho^2 \epsilon_{t-2} \\ &\quad \uparrow \\ &\quad \epsilon_{t-2} = u_{t-2} + \rho \epsilon_{t-3} \\ &\vdots \\ &= \sum_{j=0}^{\infty} \rho^j u_{t-j} . \end{aligned} \tag{3.1}$$

ϵ_t è una somma ponderata di variabili non correlate di media 0 e varianza σ^2 . Dalla rappresentazione (3.1) ricaviamo che

¹Salvo il termine u_t che per altro è indipendente sia da ϵ_{t-1} che da $\epsilon_{t-2}, \epsilon_{t-3}, \dots$.

- Proprietà 1: $E(\epsilon_t) = 0$
- Proprietà 2: $V(\epsilon_t) = \sigma^2 \sum_{j=0}^{\infty} (\rho^2)^j = \frac{1}{1-\rho^2} \sigma^2$.
- Proprietà 3: La covarianza fra ϵ_t e ϵ_{t-1}

$$\begin{aligned}
Cov(\epsilon_t, \epsilon_{t-1}) &= E(\epsilon_t \epsilon_{t-1}) \\
&= E([u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \dots][u_{t-1} + \rho u_{t-2} + \dots]) \\
&= \rho E(u_{t-1}^2) + \rho^3 E(u_{t-2}^2) + \rho^5 E(u_{t-3}^2) + \dots \\
&= \sigma^2 \rho (1 + \rho^2 + \rho^4 + \dots) \\
&= \frac{\rho}{1-\rho^2} \sigma^2.
\end{aligned}$$

Facilmente si ottiene che $Cov(\epsilon_t, \epsilon_{t-k}) = \frac{\rho^k}{1-\rho^2} \sigma^2$. Notiamo che la covarianza non dipende da t ma solo dalla distanza k tra i due errori.

Osservazione 2. Un processo che soddisfa le proprietà 1-3 è chiamato debolmente stazionario.

Esempio 4. Per un campione di n osservazioni costruiamo la matrice delle varianze covarianze di ϵ . Per $n = 3$

$$V(\epsilon)_{(3 \times 3)} = \sigma^2 \frac{1}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

In generale

$$V(\epsilon) = \sigma^2 \frac{1}{1-\rho^2} \underbrace{\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \ddots & \vdots \\ \rho^2 & \rho & 1 & \ddots & \rho^2 \\ \vdots & \ddots & \ddots & \ddots & \rho \\ \rho^{n-1} & \dots & \rho^2 & \rho & 1 \end{bmatrix}}$$

è la matrice $V(\theta)$ che qui è $V(\rho)$ e che dipende da un solo parametro ρ .

Sappiamo tutto sulla matrice $V(\rho)$:

$$\begin{aligned}
|V(\rho)| &= \frac{1}{1-\rho^2} \\
V(\rho)^{-1} &= \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \ddots & \vdots \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix}
\end{aligned}$$

matrice tridiagonale.

Cerchiamo ora la matrice $P(\rho)$ che trasforma il modello di regressione generalizzata in un modello di regressione classica e cioè

$$P(\rho)V(\rho)P(\rho)' = I$$

Se io considerassi la trasformazione $\tilde{\epsilon}_t = \epsilon_t - \rho\epsilon_{t-1}$ otterrei u_t . Il modello di partenza è come sappiamo

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \epsilon_t \\ \underbrace{y_t - \rho y_{t-1}}_{\tilde{y}_t} &= \underbrace{\beta_0(1 - \rho)}_{\text{costante}} + \beta_1 \underbrace{(x_t - \rho x_{t-1})}_{\tilde{x}_t} + u_t \\ \tilde{y}_t &= \tilde{\beta}_0 + \beta_1 \tilde{x}_t + u_t \end{aligned}$$

Quest'ultima equazione rappresenta il modello trasformato i cui parametri possono essere stimati per mezzo dei *m.q.o.*

Problema: per l'osservazione $t = 1$ non abbiamo l'osservazione precedente. Si perde così un'osservazione (grado di libertà). La soluzione esatta consiste nel trasformare la prima osservazione nel seguente modo:

$$\sqrt{1 - \rho^2} y_1 = \beta_0 \sqrt{1 - \rho^2} + \beta_1 \sqrt{1 - \rho^2} x_1 + \text{errore}$$

Come sarà quindi la matrice $P(\rho)$ con un campione di $n = 5$ osservazioni? Riempite lo spazio a disposizione

$$P(\rho) = \begin{bmatrix} & \\ & \\ & \\ & \\ & \end{bmatrix}$$

3.1.2 Test di autocorrelazione

In un problema di regressione con osservazioni temporali come possiamo verificare la presenza di errori autocorrelati? Sappiamo che quando $\rho = 0$ non c'è autocorrelazione. Questo parametro ρ è una caratteristica degli errori ϵ_t che tuttavia non sono osservabili. La prima tappa dunque è quella di stimare questi errori applicando i *m.q.o.* sul modello, avendo cura di sempre includere la costante. Dopodiché deriviamo una stima degli ϵ_t :

$$\hat{\epsilon}_t = y_t - x_t' \hat{\beta}$$

Il test d'ipotesi $H_0 : \rho = 0$ si basa sull'esame di questi residui.

1. Test informale

Si fa il grafico degli errori. Se, come nel seguente esempio

Grafico

si osserva una persistenza di errori dello stesso segno, allora c'è evidenza di autocorrelazione.

2. Test formale: Durbin-Watson

(a) Si calcola la quantità

$$d = \frac{\sum_{t=2}^n (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\epsilon}_t^2}.$$

Per n sufficientemente grande $d \underset{\text{circa}}{\sim} 2(1 - \rho)$.

Interpretazione di d :

$$d = \frac{\sum_{t=2}^n \hat{\epsilon}_t^2 + \sum_{t=2}^n \hat{\epsilon}_{t-1}^2 - 2 \sum_{t=2}^n \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^n \hat{\epsilon}_t^2} \simeq 2 - 2 \frac{\frac{1}{n} \sum_{t=2}^n \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\frac{1}{n} \sum_{t=1}^n \hat{\epsilon}_t^2}$$

Il *denominator* è la varianza empirica di ϵ_t e tende verso la vera varianza che, come calcolato precedentemente, è uguale a $\sigma^2 / (1 - \rho^2)$.

Il *numeratore* è una somma di prodotti incrociati, è la covarianza empirica fra l'errore ϵ_t e l'errore ritardato di un periodo, ϵ_{t-1} . Questa covarianza empirica converge verso la vera covarianza $Cov(\epsilon_t, \epsilon_{t-1}) = \frac{\rho}{1 - \rho^2} \sigma^2$.

Ecco spiegato perché al crescere di n la statistica d converge verso $2(1 - \rho)$. Avremo quindi per n sufficientemente grande

$$d \underset{\text{circa}}{\simeq} 2(1 - \rho). \quad (3.2)$$

Abbiamo allora la seguente situazione:

- Esiste una forte autocorrelazione positiva quando $\rho \simeq 1$ che per la (3.2) è equivalente a $d \simeq 0$.

- Se gli ϵ_t sono indipendenti segue $\rho = 0$. Per la (3.2) l'indipendenza implica $d \simeq 2$.
- Esiste una forte autocorrelazione negativa quando $\rho \simeq -1$ che per la (3.2) è equivalente a $d \simeq 4$.

Per testare l'autocorrelazione positiva (che è il caso tipicamente incontrato in economia) eseguirò il test dell'ipotesi nulla $H_0 : \rho = 0 \Leftrightarrow d = 2$ contro l'ipotesi alternativa $H_A : \rho > 0 \Leftrightarrow d < 2$. Si tratta quindi di un test unilaterale sinistro.

- Quindi, quando $d \sim 2 \Rightarrow$ assenza di autocorrelazione.

grafico

Idealmente vorremmo trovare la soglia d^* tale che se $d < d^* \Rightarrow$ autocorrelazione, mentre se $d > d^* \Rightarrow$ no. Sfortunatamente la vera distribuzione della statistica d è sconosciuta. Essa dipende dai valori presenti nella matrice X o, in altre parole, dal campione osservato. Gli autori Durbin-Watson hanno potuto trovare due limiti che delimitano il livello critico d^* , notati rispettivamente d_L e d_U . Vale la seguente regola di decisione:

- se $d < d_L$ sarà sicuramente $d < d^* \Rightarrow$ autocorrelazione
- se $d > d_U$ sarà sicuramente $d > d^* \Rightarrow$ assenza di autocorr.
- se $d_L < d < d_U \Rightarrow$ non possiamo concludere nulla

Questi limiti d_L e d_U dipendono da:

- Livello di significatività
- n : il numero d'osservazioni
- K' : il numero di variabili esplicative senza la costante (quindi $K' = K - 1$).

Esempio 5. Abbiamo stimato una regressione semplice con $n = 30$, $K' = 1$. Se d fosse uguale a 1.30, all'1% di significatività non potrei rifiutare l'ipotesi nulla $H_0 : \rho = 0$. Se invece avessimo $K' = 2$, con un $d = 1.30$ non si potrebbe concludere.

Osservazione 3. Attenzione: il test non è applicabile quando nel modello appare una variabile endogena ritardata. Ad esempio:

$$c_t = a + bR_t + cC_{t-1} + \varepsilon_t$$

Se desiderate testare l'autocorrelazione negativa si prende $4 - d$ e si procede come per l'autocorrelazione positiva.

3.1.3 I metodi di stima

Abbiamo stimato il modello con i *m.q.o.* ed il test *DW* ci indica autocorrelazione. Cosa fare? Il *DW* è un allarme che indica la presenza di una sistematicità nel fenomeno studiato che il modello non ha saputo incorporare. Non è necessariamente frutto dell'autocorrelazione ma potrebbe essere causato da una cattiva specificazione del modello, ad esempio

1.
 - sono state trascurate delle variabili esplicative
 - la forma funzionale che lega le variabili esplicative alla variabile dipendente (quadratica anziché lineare)
 - forma dinamica invece che statica:
vero modello (DGP)

$$C_t = a + bR_t + cC_{t-1} + \varepsilon_t,$$

modello stimato

$$C_t = a + bR_t + \varepsilon_t \Leftarrow \text{autocorrelazione}$$

2. Dopo aver considerato tutte le osservazioni [1], se l'autocorrelazione persiste allora utilizziamo i *m.q.g.*

2.1 *m.q.g.* a due tappe

- i. Prima tappa: stimiamo utilizzando i *m.q.o.* e ricaviamo $\hat{\varepsilon}_t$. In seguito costruiamo la regressione ausiliaria

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + u_t \quad (3.3)$$

ed utilizzando nuovamente i *m.q.o.* sul modello (3.3) otteniamo

$$\hat{\rho} = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2} \xrightarrow{n \rightarrow \infty} \rho.$$

- ii. Seconda tappa: *m.q.g.* calcolabili

- Sia con la formula

$$\beta^* = (X'V(\hat{\rho})^{-1}X)^{-1}X'V(\hat{\rho})^{-1}y$$

- oppure con i *m.q.o.* sul modello trasformato con $P(\hat{\rho})$.

- 2.2 Quale alternativa al 2.1 eseguire una stima di massima verosimiglianza.

3.1.4 La previsione

Abbiamo delle osservazioni $t = 1, \dots, n$ e desideriamo prevedere y per il periodo successivo conoscendo x_{n+1} . Il vero y_{n+1} è dato dal medesimo modello:

$$y_{n+1} = x'_{n+1}\beta + \varepsilon_{n+1}$$

con

$$\varepsilon_{n+1} = \rho\varepsilon_n + u_{n+1}.$$

Sappiamo che il previsore lineare ottimale è dato da:

$$y_{n+1}^p = x'_{n+1}\beta^* + R'V^{-1}\hat{\varepsilon}$$

dove

$$\hat{\varepsilon} = y - X\beta^*.$$

Nel nostro caso:

$$V^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \ddots & \vdots \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 + \rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{bmatrix}$$

e dobbiamo calcolare

$$\begin{aligned} cov(\varepsilon_{n+1}, \varepsilon) &= E(\varepsilon_{n+1}\varepsilon') \\ &= E(\varepsilon_{n+1}\varepsilon_1, \dots, \varepsilon_{n+1}\varepsilon_n) \\ &= \frac{\sigma^2}{1 - \rho^2} [\rho^n, \rho^{n-1}, \dots, \rho] \end{aligned}$$

Un semplice calcolo mostra che

$$\begin{matrix} R' & V^{-1} \\ (1 \times n) & (n \times n) \end{matrix} = \begin{matrix} [0, 0, \dots, \rho] \\ (1 \times n) \end{matrix}$$

e quindi $R'V^{-1}\hat{\varepsilon} = \rho\hat{\varepsilon}_n$ da cui segue la formula

$$y_{n+1}^p = x'_{n+1}\beta^* + \rho\hat{\varepsilon}_n$$

Poiché ρ è sconosciuto lo sostituiamo con la stima $\hat{\rho}$ ottenendo in questo modo la formula finale

$$y_{n+1}^p = x'_{n+1}\beta^* + \hat{\rho}\hat{\varepsilon}_n$$

3.2 Altre forme di autocorrelazione

Il modello statistico considerato in questo capitolo si chiama autoregressivo di primo ordine: $AR(1)$

$$\epsilon_t = \rho\epsilon_{t-1} + u_t, \quad u_t \sim iid(0, \sigma^2)$$

È possibile generalizzare i modelli autoregressivi ad ordini superiori. In tal caso si parla di $AR(p)$, dove il numero intero p indica appunto l'ordine del modello autoregressivo

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \dots + \rho_p \epsilon_{t-p} + u_t$$

I modelli a media mobile (Moving Average) definiscono un'altra classe di modelli molto utilizzati nell'analisi delle serie storiche. Un modello a media mobile di ordine 1 $MA(1)$ è definito da

$$\epsilon_t = u_t + c u_{t-1} .$$

Anche in questo caso è immediata la sua generalizzazione a ordini superiori:

$$MA(q) : \epsilon_t = u_t + c_1 u_{t-1} + c_2 u_{t-2} + \dots + c_q u_{t-q} .$$

Infine un processo stocastico generale frequentemente incontrato in economia e statistica è l' $ARMA(p, q)$:

$$\epsilon_t = \underbrace{\rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \dots + \rho_p \epsilon_{t-p}}_{\text{parte autoregressiva}} + \underbrace{u_t + c_1 u_{t-1} + c_2 u_{t-2} + \dots + c_q u_{t-q}}_{\text{parte media mobile}}$$

Parte III

La regressione stocastica

Capitolo 4

Elementi di analisi asintotica

Nel contesto di una stima del modello econometrico, con l'analisi asintotica si intende lo studio delle proprietà del campione quando la sua numerosità cresce infinitamente. Perché un tale studio?

1. Un motivo teorico.

Quando la numerosità del campione aumenta, l'informazione disponibile diventa sempre più importante. Col crescere dell'informazione la precisione dello stimatore dovrebbe aumentare. Al limite per $n = \infty$ l'informazione è totale e la precisione dello stimatore dovrebbe essere assoluta. Il metodo di stima proposto dovrebbe quindi condurre asintoticamente al vero valore del parametro sconosciuto con assoluta certezza (con probabilità 1). Questo è il problema della convergenza dello stimatore verso il vero valore del parametro.

2. Un motivo pratico

Fin qui ci siamo occupati quasi esclusivamente di stimatori lineari. Per questi è possibile calcolare il valore atteso che ci permette di verificare la loro correttezza. È pure possibile calcolare la varianza dello stimatore lineare. Essa ci indica la precisione con la quale stimiamo il parametro e ci permette di condurre dei test.

Al di là della regressione classica saremo confrontati con stimatori non lineari. Per questi è praticamente impossibile calcolare il valore atteso (e quindi parlare di correttezza) e a fortiori calcolare la varianza ed eseguire i test.

L'analisi asintotica dà una risposta a questo tipo di problemi. I risultati saranno da interpretare come un'approssimazione molto buona per grandi campioni.

4.1 La convergenza in probabilità

Prima di dare la definizione formale di convergenza in probabilità consideriamo il seguente esempio.

Esempio 6. Estraiamo un campione aleatorio da una popolazione di variabili aleatorie (μ, σ^2) , μ sconosciuto. Per stimare μ prendiamo la media del campione. Quando nel campione c'è una sola osservazione

$$\bar{X}_1 = X_1,$$

due osservazioni

$$\bar{X}_2 = \frac{X_1 + X_2}{2}$$

e così via \dots

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Le quantità $\bar{X}_1, \bar{X}_2, \dots$ formano una successione che noteremo \bar{X}_n . Qual è il limite di questa successione quando $n \rightarrow \infty$? La legge dei grandi numeri ci assicura che questo limite è proprio μ . Possiamo dunque concludere che la successione \bar{X}_n converge verso μ . È questo un primo esempio di convergenza in probabilità.

4.1.1 Definizione e proprietà

Uno stimatore è detto convergente se la probabilità che ci si discosti anche solo minimamente dal suo vero valore tende a 0 con $n \rightarrow \infty$. Sia θ il vero parametro, $\hat{\theta}_n$ uno stimatore di θ basato sul campione di numerosità n . Formalmente:

Definizione 1. $\hat{\theta}_n$ è uno stimatore convergente di θ se, $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} Prob \left\{ \left| \hat{\theta}_n - \theta \right| > \varepsilon \right\} \rightarrow 0$$

Grafico

Al limite dobbiamo arrivare ad una situazione per cui tutta la massa di probabilità è concentrata in un solo punto, il vero valore θ . Per indicare che il vero valore converge verso θ , scriveremo $p \lim \hat{\theta}_n = \theta$. In certi casi è possibile valutare esplicitamente questo limite in probabilità, come ad esempio nel caso della media del campione se la popolazione è normale. In altri casi invece è necessario ricorrere ad alcune proprietà che il limite in probabilità $p \lim$ possiede e che enunciamo adesso.

Consideriamo le due successioni $\hat{\theta}_n$ e $\hat{\beta}_n$ con

$$p \lim \hat{\theta}_n = \theta < \infty \quad \text{e} \quad p \lim \hat{\beta}_n = \beta < \infty.$$

Il limite in probabilità possiede le seguenti 4 proprietà:

$$P1 \quad p \lim \left(\widehat{\theta}_n \pm \widehat{\beta}_n \right) = p \lim \widehat{\theta}_n \pm p \lim \widehat{\beta}_n = \theta \pm \beta$$

$$P2 \quad p \lim \left(\widehat{\theta}_n * \widehat{\beta}_n \right) = p \lim \widehat{\theta}_n * p \lim \widehat{\beta}_n = \theta * \beta$$

$$P3 \quad \beta \neq 0,$$

$$p \lim \left(\frac{\widehat{\theta}_n}{\widehat{\beta}_n} \right) = \frac{p \lim \widehat{\theta}_n}{p \lim \widehat{\beta}_n} = \frac{\theta}{\beta}$$

P4 Invarianza. Se $g(\cdot)$ è una funzione continua:

$$p \lim g \left(\widehat{\theta}_n \right) = g \left(p \lim \widehat{\theta}_n \right) = g(\theta)$$

4.1.2 La convergenza in media quadratica

Sia $\widehat{\theta}_n$ uno stimatore di θ . Se per ogni n è possibile

- Calcolare il valore atteso $E(\widehat{\theta}_n)$ e quindi anche il termine di distorsione o *Bias* definito come

$$B(\widehat{\theta}_n) := E(\widehat{\theta}_n) - \theta = E(\widehat{\theta}_n - \theta)$$

- Calcolare $V(\widehat{\theta}_n) = E \left(\widehat{\theta}_n - E(\widehat{\theta}_n) \right)^2$

è chiaro allora che per la convergenza in probabilità è sufficiente che

1. Il Bias tenda a zero con $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} B(\widehat{\theta}_n) = 0 \quad (4.1)$$

2. La varianza di $\widehat{\theta}_n$ tenda anch'essa a 0:

$$\lim_{n \rightarrow \infty} V(\widehat{\theta}_n) = 0 \quad (4.2)$$

Questo criterio è chiamato *convergenza in media quadratica*. Come andremo a dimostrare le condizioni (4.1) e (4.2) sono equivalenti a

$$\lim_{n \rightarrow \infty} E(\widehat{\theta}_n - \theta)^2 = 0 \quad (4.3)$$

Dimostriamo che $\lim_{n \rightarrow \infty} E(\widehat{\theta}_n - \theta)^2 = 0 \Leftrightarrow B(\widehat{\theta}_n) \rightarrow 0$ e $V(\widehat{\theta}_n) \rightarrow 0$. Abbiamo:

$$\begin{aligned} E \left(\widehat{\theta}_n - \theta \right)^2 &= E \left[\left(\widehat{\theta}_n - E(\widehat{\theta}_n) \right) + \left(E(\widehat{\theta}_n) - \theta \right) \right]^2 = \\ &= E \left(\widehat{\theta}_n - E(\widehat{\theta}_n) \right)^2 + B(\widehat{\theta}_n)^2 + 2B(\widehat{\theta}_n)E \left(\widehat{\theta}_n - E(\widehat{\theta}_n) \right) = \\ &= E \left(\widehat{\theta}_n - E(\widehat{\theta}_n) \right)^2 + B(\widehat{\theta}_n)^2 = \\ &= V \left(\widehat{\theta}_n \right) + B(\widehat{\theta}_n)^2 \end{aligned}$$

Nell'Esempio 6 anche senza la normalità di X

$$\begin{cases} E(\overline{X}_n) = \mu \Rightarrow B(\overline{X}_n) = 0 \\ V(\overline{X}_n) = \frac{\sigma^2}{n} \rightarrow 0 \text{ quando } n \rightarrow \infty \end{cases}$$

4.1.3 Convergenza dei momenti del campione

Consideriamo una popolazione caratterizzata dalla variabile aleatoria X .

- Il primo momento della popolazione è: $E(X) = \mu$
- Il secondo momento è: $E(X^2) = \mu_2$
- Il secondo momento centrato: $E(X - \mu)^2 = V(X) = \sigma^2$

Prendiamo un campione aleatorio di numerosità n : X_1, \dots, X_n . I momenti del campione sono:

- Primo momento: $\bar{X} = \frac{1}{n} \sum X_i$.
- Secondo momento: $\frac{1}{n} \sum X_i^2$.
- Secondo momento centrato: $\frac{1}{n} \sum (X_i - \bar{X})^2$

Proprietà fondamentale: I momenti del campione convergono in probabilità verso i momenti corrispondenti della popolazione. In particolare

$$\begin{aligned} p \lim \frac{1}{n} \sum X_i &= E(X) = \mu \\ p \lim \frac{1}{n} \sum X_i^2 &= E(X^2) = \mu_2 \\ p \lim \frac{1}{n} \sum (X_i - \bar{X})^2 &= V(X) = \sigma^2 \end{aligned}$$

Se la popolazione è caratterizzata da due variabili aleatorie X e Z , oltre ai momenti individuali ci sono i momenti incrociati.

Popolazione:

- Il secondo momento incrociato: $E(XZ)$
- Il secondo momento centrato incrociato:

$$E(X - \mu_X)(Z - \mu_Z) = Cov(X, Z)$$

dove $\mu_X = E(X)$ e $\mu_Z = E(Z)$

Campione: X_1, \dots, X_n e Z_1, \dots, Z_n

- Secondo momento incrociato: $\frac{1}{n} \sum X_i Z_i$
- Centrato: $\frac{1}{n} \sum (X_i - \bar{X})(Z_i - \bar{Z})$

Anche in questo caso vale la proprietà fondamentale,

$$\begin{aligned} p \lim \frac{1}{n} \sum X_i Z_i &= E(XZ) \\ p \lim \frac{1}{n} \sum (X_i - \bar{X})(Z_i - \bar{Z}) &= Cov(X, Z) \end{aligned}$$

Esempio 7. La popolazione è composta da X, Z indipendenti. Calcolare:

$$p \lim \frac{1}{n} \sum \underbrace{(X_i - \bar{X})(Z_i - \bar{Z})}_{\text{Covarianza empirica}} = 0 = \underbrace{Cov(X, Z)}_{\text{vera covarianza}}$$

4.2 La convergenza in distribuzione

4.2.1 Un esempio

Ritornando alle medie del campione abbiamo "dimostrato" che,

$$\begin{aligned}\bar{X}_n &\rightarrow \mu \\ (\bar{X}_n - \mu) &\rightarrow 0\end{aligned}$$

Ciò significa che la distribuzione limite di $\bar{X}_n - \mu$ è degenerata (tutta la massa di probabilità è centrata su 0). Se ammettiamo la normalità della popolazione

$$\begin{aligned}(\bar{X}_n - \mu) &\sim N\left(0, \frac{\sigma^2}{n}\right) \\ \sqrt{n}(\bar{X}_n - \mu) &\sim N(0, \sigma^2)\end{aligned}$$

Notiamo che la distribuzione di $\sqrt{n}(\bar{X}_n - \mu)$ è sempre la stessa al variare di n . Anche al limite sarà sempre una distribuzione normale, $N(0, \sigma^2)$. In questo caso la distribuzione di $(\bar{X}_n - \mu)$ è degenerata, ma quella di $(\bar{X}_n - \mu)$ dilatata per \sqrt{n} è una normale di varianza σ^2 e valore atteso 0.

Se la popolazione non è normale, il teorema del limite centrale ci assicura che la distribuzione limite di $\sqrt{n}(\bar{X}_n - \mu)$ è normale $(0, \sigma^2)$. Scriveremo:

$$\sqrt{n}(\bar{X}_n - \mu) \stackrel{D}{\sim} N(0, \sigma^2)$$

4.2.2 Teoremi del limite centrale

Prendendo $X_i - \mu$ dall'esempio precedente, chiamiamo $\varepsilon_i = X_i - \mu$, con $E(\varepsilon_i) = 0$ e $V(\varepsilon_i) = \sigma^2$. Dato il carattere aleatorio del campione si ha che:

$$\varepsilon_i \sim i.i.d.(0, \sigma^2)$$

D'altra parte

$$\begin{aligned}\sqrt{n}(\bar{X}_n - \mu) &= \sqrt{n}\left(\frac{1}{n} \sum X_i - \mu\right) = \sqrt{n}\left(\frac{1}{n} \sum (X_i - \mu)\right) \\ &= \sqrt{n}\left(\frac{1}{n} \sum \varepsilon_i\right) = \frac{1}{\sqrt{n}} \sum \varepsilon_i\end{aligned}$$

Teorema 1. Limite centrale semplice. Condizione:

$$\varepsilon_i \sim i.i.d.(0, \sigma^2).$$

Ne segue che

$$\frac{1}{\sqrt{n}} \sum \varepsilon_i \stackrel{D}{\sim} N(0, \sigma^2).$$

Consideriamo ora i vettori

$$\begin{matrix} v_i & \varepsilon_i \\ (K \times 1) & (1 \times 1) \end{matrix}$$

dove i v_i sono composti da numeri (non aleatori). Abbiamo quindi:

$$\begin{aligned}E[v_i \varepsilon_i] &= v_i E[\varepsilon_i] = v_i \cdot 0 = 0 \\ V[v_i \varepsilon_i] &= v_i V[\varepsilon_i] v_i' = \sigma^2 v_i v_i' \\ &\quad (K \times K)\end{aligned}$$

Teorema 2. Limite centrale generalizzato. Condizioni:

1. $\varepsilon_i \sim i.i.d.(0, \sigma^2)$
2. v_i sono dei vettori non stocastici $(K \times 1)$ tali che

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n V(v_i \varepsilon_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma^2 v_i v_i' = Q, \text{ matrice d.p.}$$

Ne segue che

$$\Rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^n v_i \varepsilon_i \stackrel{D}{\sim} N(0, Q)$$

Teorema 3. Prodotto. Dati:

1. Una successione di vettori aleatori $\{Y_n\}$ d'ordine K che tende in distribuzione verso la variabile $Y \sim N(0, Q)$, dove Q è una matrice definita positiva, $Y_n \stackrel{D}{\sim} Y$ ed $Y \sim N(0, Q)$.
2. Una successione di matrici $\{A_n\}$ d'ordine $(m \times K)$ di elementi aleatori o meno, tale che $\text{plim} A_n = A$, A matrice non stocastica di rango m .

Ne segue che la successione dei prodotti $\{A_n Y_n\}$ tende in distribuzione verso la variabile aleatoria $AY \sim N(0, AQA')$, e cioè

$$A_n Y_n \stackrel{D}{\sim} N(0, AQA')$$

4.2.3 Consiglio pratico

In pratica, per uno stimatore $\hat{\theta}_n$ del parametro θ , si deve verificare dapprima la convergenza:

$$p \lim \hat{\theta}_n = \theta$$

o equivalentemente

$$p \lim (\hat{\theta}_n - \theta) = 0$$

Poi si eseguono le seguenti tappe:

- [1] La distribuzione asintotica di $\hat{\theta}_n$ è la distribuzione limite della successione $\left\{ \sqrt{n}(\hat{\theta}_n - \theta) \right\}$ che in moltissimi casi sarà normale. Per trovarla si applicano i teoremi 1, 2 e 3.
- [2] La varianza asintotica dello stimatore $\hat{\theta}_n$, che noteremo $\text{VarAs}(\hat{\theta}_n)$ è la varianza della distribuzione asintotica divisa per n .

Osservazione 4. Lo stimatore $\hat{\theta}_n$ è asintoticamente efficiente se è convergente e asintoticamente normale a varianza asintotica minima nella classe di tutti gli stimatori convergenti e asintoticamente normali (BAN). Lo stimatore di massima verosimiglianza è asintoticamente efficiente.

4.3 Analisi asintotica della regressione classica

1. Il modello

$$\begin{aligned}y_i &= x_i' \beta + \varepsilon_i \\ y &= X \beta + \varepsilon\end{aligned}$$

2. Le ipotesi

H1: Sulle variabili esplicative

- (a) Non stocastiche
- (b) $rg(X) = K$
- (c) Nuovo: comportamento asintotico

$$\lim_{n \rightarrow \infty} \frac{1}{n} X' X = Q, \quad d.p.$$

H2: Sugli errori

- (a) $E[\varepsilon] = 0$
- (b) $V[\varepsilon] = \sigma^2 I$
- (c) Nuovo: $\varepsilon_i \sim i.i.d.(0, \sigma^2)$

H3: X, ε mutuamente indipendenti.

3. Lo stimatore dei *m.q.o.*

$$\begin{aligned}\hat{\beta} &= (X' X)^{-1} X' Y \\ E[\hat{\beta}] &= \beta \\ V[\hat{\beta}] &= \sigma^2 (X' X)^{-1}\end{aligned}$$

4. La convergenza quadratica

- La distorsione: $B(\hat{\beta}) = 0$
- La varianza: $V(\hat{\beta}) = \sigma^2 (X' X)^{-1} = \frac{\sigma^2}{n} \left(\frac{X' X}{n} \right)^{-1}$
Segue che :

$$\lim \frac{\sigma^2}{n} \lim \left(\frac{X' X}{n} \right)^{-1} = \lim \frac{\sigma^2}{n} Q^{-1} = 0$$

5. Prova diretta della convergenza di $\hat{\beta}$

$$\begin{aligned}\hat{\beta} &= (X' X)^{-1} X' Y \\ &= \beta + (X' X)^{-1} X' \varepsilon \\ p \lim \hat{\beta} &= p \lim \beta + p \lim (X' X)^{-1} X' \varepsilon \\ &= \beta + p \lim \left(\frac{X' X}{n} \right)^{-1} p \lim \left(\frac{X' \varepsilon}{n} \right) \\ &= \beta + Q^{-1} p \lim \left(\frac{X' \varepsilon}{n} \right)\end{aligned}$$

Ora:

$$\frac{1}{n} X' \varepsilon = \begin{bmatrix} \frac{1}{n} \sum X_{i1} \varepsilon_i \\ \frac{1}{n} \sum X_{i2} \varepsilon_i \\ \vdots \\ \frac{1}{n} \sum X_{iK} \varepsilon_i \end{bmatrix}$$

ogni termine è una somma di prodotti incrociati divisi per n . Siccome un termine del prodotto è nullo in valore atteso, questa somma rappresenta la covarianza empirica nel campione tra una variabile esplicativa (k -esima) e gli errori. Questa tende verso la vera covarianza che è 0 poiché le variabili esplicative sono indipendenti dagli errori.

Segue che:

$$\begin{aligned} p \lim \frac{1}{n} X' \varepsilon &= 0 \\ p \lim \hat{\beta} &= \beta, \text{ convergente} \end{aligned}$$

6. Normalità asintotica

Si cerca la distribuzione limite di $\sqrt{n}(\hat{\beta} - \beta)$. Ora:

$$\hat{\beta} - \beta = (X' X)^{-1} X' \varepsilon = \left(\frac{1}{n} X' X \right)^{-1} \frac{1}{n} X' \varepsilon$$

Ne consegue che,

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} X' X \right)^{-1} \frac{1}{\sqrt{n}} X' \varepsilon.$$

Si tratta di un prodotto di due termini.

- Il primo termine, $\left(\frac{1}{n} X' X \right)^{-1}$, tende verso la matrice di costanti, Q^{-1} .
- Il secondo termine può scriversi:

$$\frac{1}{\sqrt{n}} X' \varepsilon = \frac{1}{\sqrt{n}} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \varepsilon_i$$

Quest'espressione è proprio quella del teorema 2: verifichiamo le due condizioni.

(a) $\varepsilon_i \sim i.i.d.(0, \sigma^2)$

(b) x_i non stocastici, tali che

$$\begin{aligned} \lim \frac{1}{n} \sum V(x_i \varepsilon_i) &= \lim \frac{1}{n} \sum x_i \sigma^2 x_i' = \\ &= \sigma^2 \lim \frac{1}{n} \sum x_i x_i' = \\ &= \sigma^2 \lim \frac{1}{n} X' X = \sigma^2 Q, d.p. \end{aligned}$$

Per il teorema 2: $\frac{1}{\sqrt{n}}X'\varepsilon \stackrel{D}{\sim} N(0, \sigma^2 Q)$, e per il teorema del prodotto:

$$\sqrt{n}(\hat{\beta} - \beta) \stackrel{D}{\sim} N(0, \sigma^2 Q^{-1}).$$

Perché questa varianza? Siccome $\sqrt{n}(\hat{\beta} - \beta) \stackrel{D}{\sim} Q^{-1} \cdot N(0, \sigma^2 Q)$, $Var = Q^{-1}(\sigma^2 Q)(Q^{-1})' = \sigma^2 Q^{-1}$.

La *varianza asintotica* come definita in precedenza è la varianza della distribuzione asintotica divisa per n :

$$VarAs(\hat{\beta}) = \frac{\sigma^2}{n} Q^{-1}$$

Capitolo 5

Il modello di regressione stocastica

5.1 Il modello e le ipotesi

5.1.1 Il modello

Sin qui abbiamo ammesso che le variabili esplicative sono non stocastiche, dei numeri. Quest'ipotesi è valida nelle scienze sperimentali dove colui che conduce l'analisi può scegliere debitamente le variabili esplicative. In economia questa ipotesi non è sempre soddisfatta. Quando le variabili esplicative sono (almeno in parte) aleatorie i metodi di stima e le loro proprietà dipendono non soltanto dalla distribuzione degli errori, ma dalla *distribuzione congiunta* delle variabili esplicative e degli errori. In particolare, tali le proprietà dipenderanno dall'indipendenza o meno tra le variabili esplicative e l'errore.

In certi casi l'indipendenza tra variabili esplicative ed errori è una conseguenza logica del modello studiato.

Esempio 8. Il modello keynesiano semplice

Funzione di consumo: $C_t = a + bR_t + \varepsilon_t$

La definizione del reddito: $R_t = C_t + I_t$

R_t è la variabile esplicativa. Tuttavia R_t e ε_t non possono essere indipendenti in quanto, dalla definizione di reddito, R_t contiene C_t il quale contiene ε_t .

5.1.2 Le ipotesi

Partiamo come al solito dal modello di regressione

$$\begin{aligned} y_i &= x_i' \beta + \varepsilon_i \quad i = 1, \dots, n \\ &\text{oppure} \\ y &= X\beta + \varepsilon \end{aligned}$$

H1: Sulle variabili esplicative

1. Stocastiche (almeno in parte)

2. $rg(X) = K$, con probabilità 1
3. $p \lim \frac{1}{n} X'X = Q_X$, definita positiva

H2: Sugli errori

$$\varepsilon \sim i.i.d.(0, \sigma^2)$$

H3: Sulla relazione tra X e ε . Due casi:

1. Indipendenza: $p \lim \frac{1}{n} X'\varepsilon = 0$
2. Dipendenza: $p \lim \frac{1}{n} X'\varepsilon = C \neq 0$

5.2 Quando i regressori e gli errori sono indipendenti

Tutto funziona come nella regressione classica.

5.3 Quando sono correlati

5.3.1 La non convergenza dei *m.q.o.*

È immediato verificare che i *m.q.o.* non sono convergenti quando le variabili esplicative sono correlate con gli errori:

$$\begin{aligned}\hat{\beta} &= \beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + \left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'\varepsilon \\ p \lim \hat{\beta} &= \beta + Q^{-1}C \neq 0.\end{aligned}$$

In tal caso come bisogna procedere? Prima di proporre un nuovo stimatore consideriamo una reinterpretazione dello stimatore dei minimi quadrati. Pre-moltiplichiamo il modello

$$y = X\beta + \varepsilon$$

per la matrice X' e dividiamo per n

$$\frac{1}{n}X'y = \frac{1}{n}X'X\beta + \frac{1}{n}X'\varepsilon \quad (5.1)$$

Quando X e ε sono indipendenti $\frac{1}{n}X'\varepsilon \rightarrow 0$ e lo si può tralasciare. Risolvendo (5.1) rispetto a β otteniamo così lo stimatore dei minimi quadrati ordinari:

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Quando, invece, X ed ε sono correlati, il termine $\frac{1}{n}X'\varepsilon$ non tende a zero, ed è questa la ragione della non convergenza al vero valore. Viene però l'idea di

premultiplicare il modello non per X' , ma per una matrice Z' (dello stesso ordine $K \times n$ che X') non correlata con ε ,

$$\frac{1}{n}Z'y = \frac{1}{n}Z'X\beta + \frac{1}{n}Z'\varepsilon$$

Se allora la matrice $Z'X$ è non singolare (ciò che si deve ammettere), risolvendo per β si ottiene,

$$\hat{\beta}_{vs} = (Z'X)^{-1}Z'y \quad (5.2)$$

Per concludere, la matrice Z è chiamata matrice di strumenti e lo stimatore definito dalla (5.2) è chiamato stimatore delle variabili strumentali. È questo il caso in cui il numero di strumenti (il numero di colonne della matrice Z) è uguale al numero di variabili esplicative e cioè K .

Esempio 9. Ritornando all'esempio per la funzione di consumo

$$y = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & R_1 \\ \vdots & \vdots \\ 1 & R_n \end{bmatrix}, \quad p \lim \frac{1}{n} \sum R_i \varepsilon_i \neq 0$$

I *m.q.o.* non sono convergenti! Devo trovare due strumenti:

- Il termine costante non è stocastico: è il suo proprio strumento.
- Dobbiamo trovare invece uno strumento per R_t . Ammettiamo che l'investimento sia esogeno, dunque I_t indipendente da ε_t . R_t è sicuramente correlato con I_t , dato che $R_t = C_t + I_t$. Dunque la matrice Z dei due strumenti è data da:

$$Z = \begin{bmatrix} 1 & I_1 \\ \vdots & \vdots \\ 1 & I_n \end{bmatrix}$$

Sciogliamo l'espressione

$$\hat{\beta}_{vs} = (Z'X)^{-1}Z'y \iff (Z'X)\beta = Z'y$$

Introduciamo i valori nell'espressione:

$$\begin{aligned} \begin{bmatrix} 1 & \cdots & 1 \\ I_1 & \cdots & I_n \end{bmatrix} \begin{bmatrix} 1 & R_1 \\ \vdots & \vdots \\ 1 & R_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 1 & \cdots & 1 \\ I_1 & \cdots & I_n \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix} \\ \begin{bmatrix} n & \sum R_t \\ \sum I_t & \sum R_t I_t \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} \sum C_t \\ \sum I_t C_t \end{bmatrix} \\ \begin{cases} na + b \sum R_t = \sum C_t \\ a \sum I_t + b \sum R_t I_t = \sum I_t C_t \end{cases} \end{aligned}$$

Risolvi il sistema e troviamo lo stimatore di a .

$$\begin{aligned} \hat{a}_{vs} &= \frac{1}{n} \sum C_t - \hat{b}_{vs} \frac{1}{n} \sum R_t \\ \hat{a}_{vs} &= \bar{C} - \hat{b}_{vs} \bar{R} \end{aligned}$$

Troviamo ora lo stimatore per b .

$$\begin{aligned}\sum I_t(\bar{C} - \hat{b}_{vs}\bar{R}) + \sum I_t R_t \hat{b}_{vs} &= \sum C_t I_t \\ \underbrace{\left[\sum I_t R_t - \bar{R} \sum I_t \right]}_{m_{IR}} \hat{b}_{vs} &= \underbrace{\sum C_t I_t - \bar{C} \sum I_t}_{m_{IC}} \\ \hat{b}_{vs} = \frac{m_{IC}}{m_{IR}} &\neq \hat{b} = \frac{m_{CR}}{m_{RR}}\end{aligned}$$

5.3.2 Il metodo delle variabili strumentali

Abbiamo discusso precedentemente il caso particolare dove il numero di strumenti utilizzati (il numero di colonne della matrice Z) era uguale a K , ovvero uguale al numero di variabili esplicative. Consideriamo ora la procedura generale nella quale il numero di strumenti può essere superiore a K .

5.3.2.1 Gli strumenti

Si ammette l'esistenza di una matrice Z d'ordine $n \times m$ di strumenti tali che

1. Gli strumenti non sono correlati con gli errori:

$$p \lim \frac{1}{n} \frac{Z' \varepsilon}{(m \times 1)} = \frac{0}{(m \times 1)}$$

2. Gli strumenti sono correlati con le variabili esplicative:

$$p \lim \frac{1}{n} \frac{Z' X}{(m \times K)} = \frac{Q_{ZX}}{(m \times K)}$$

dove Q_{ZX} è di rango K . Ovviamente deve essere $m \geq K$. Questa condizione essenziale è detta condizione d'identificazione.

3. Gli strumenti posseggono dei momenti finiti di ordine 2.

$$p \lim \frac{1}{n} \sum Z' Z = Q_Z, \quad d.p.$$

Nell'esempio:

1. La costante è indipendente da ε_t , e I_t è indipendente da ε_t .

- 2.

$$\frac{1}{n} \sum Z' X = \begin{bmatrix} \frac{1}{I} & \frac{\bar{R}}{\frac{1}{n} \sum R_t I_t} \end{bmatrix}, \text{ di rango } 2 \iff m_{IR} \neq 0$$

- 3.

$$\frac{1}{n} \sum Z' Z = \begin{bmatrix} \frac{1}{I} & \frac{\bar{I}}{\frac{1}{n} \sum I_t^2} \end{bmatrix}, \text{ di rango } 2 \text{ (d.p.)} \iff m_{ZZ} \neq 0$$

5.3.2.2 La procedura

Si premoltiplica il modello $y = X\beta + \varepsilon$ per Z' per $(m \times n)$:

$$\begin{aligned} Z'y &= Z'X\beta + Z'\varepsilon \\ \tilde{y}_{(m \times 1)} &= \tilde{X}_{(m \times K)}\beta + \tilde{\varepsilon}_{(m \times 1)} \end{aligned}$$

Questo ha tutti gli aspetti di un modello di regressione con m osservazioni e K variabili esplicative $m \geq K$. In questa regressione gli errori $\tilde{\varepsilon}$ non hanno le proprietà dei $m.q.o.$. Infatti

$$\begin{aligned} \tilde{\varepsilon} &= Z'\varepsilon \\ E[\tilde{\varepsilon} | Z] &= 0 \\ V[\tilde{\varepsilon} | Z] &= Z'V(\varepsilon)Z = \sigma^2 Z'Z \end{aligned}$$

Questa varianza è conosciuta. Siamo dunque in una situazione in cui i $m.q.g.$ puri sono applicabili. L'applicazione dei $m.q.g.$ puri ci dà:

$$\begin{aligned} \hat{\beta}_{vs} &= [\tilde{X}'V^{-1}\tilde{X}]^{-1} \tilde{X}'V^{-1}\tilde{y} \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'y \\ &= (X'P_ZX)^{-1} X'P_Zy. \end{aligned}$$

Quando $m = K$, $X'Z$ è quadrata e invertibile per ipotesi:

$$\begin{aligned} \hat{\beta}_{vs} &= (Z'X)^{-1}Z'Z(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y \\ &= (Z'X)^{-1}Z'y \end{aligned}$$

5.3.3 La scelta degli strumenti

Dal punto di vista teorico: gli strumenti devono essere

- non correlati con gli errori e
- il più correlati possibile con le variabili esplicative.

Dal punto di vista pratico:

- Le variabili esplicative non stocastiche (come la costante, il trend, le dummy) come pure quelle non correlate con gli errori, sono i loro propri strumenti.
- Nei modelli macroeconomici, le variabili *esogene del sistema* (nel nostro caso l'investimento) costituiscono gli strumenti naturali.
- In certi casi in economia si può ammettere che il valore ritardato delle variabili esplicative non sia correlato con l'errore presente. In tal caso il valore ritardato è uno strumento valido.
- L'ordine con cui gli strumenti sono posizionati nelle colonne di Z non ha alcuna conseguenza.

5.3.3.1 Le proprietà di $\hat{\beta}_{vs}$

Scriviamo, sostituendo y con $X\beta + \varepsilon$, il nostro stimatore

$$\begin{aligned}\hat{\beta}_{vs} &= (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'(X\beta + \varepsilon) \\ &= \beta + (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'\varepsilon\end{aligned}$$

Non essendo lineare in ε e X non si può calcolare né il valore atteso né la varianza.

- La convergenza in probabilità

$$\begin{aligned}\hat{\beta}_{vs} &= \beta + \left[\frac{1}{n} X'Z \left(\frac{1}{n} Z'Z \right)^{-1} \frac{1}{n} Z'X \right]^{-1} \frac{1}{n} X'Z \left(\frac{1}{n} Z'Z \right)^{-1} \frac{1}{n} Z'\varepsilon \\ &= \beta + A_n \frac{1}{n} Z'\varepsilon\end{aligned}$$

Ora:

$$\begin{aligned}- p \lim \frac{1}{n} Z'\varepsilon &= p \lim \frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i = \dots\dots\dots = 0 \\ - p \lim A_n &= [Q_{XZ} Q_Z^{-1} Q_{ZX}]^{-1} Q_{XZ} Q_Z^{-1} = A, \text{ finito}\end{aligned}$$

Ne segue che:

$$p \lim \hat{\beta} = \beta$$

- La normalità asintotica

$$\sqrt{n}(\hat{\beta} - \beta) = \underbrace{A_n}_{p \lim A_n = A} \underbrace{\frac{1}{\sqrt{n}} Z'\varepsilon}_{N(0, \sigma^2 Q_Z)}$$

Per il teorema 3:

$$\sqrt{n}(\hat{\beta} - \beta) \stackrel{D}{\sim} N(0, \sigma^2 A Q_Z A')$$

Ora

$$\begin{aligned}A Q_Z A' &= [Q_{XZ} Q_Z^{-1} Q_{ZX}]^{-1} Q_{XZ} Q_Z^{-1} (Q_Z) Q_Z^{-1} \dots \\ &= [Q_{XZ} Q_Z^{-1} Q_{ZX}]^{-1}\end{aligned}$$

- La varianza asintotica

$$VarAs(\hat{\beta}_{vs}) = \frac{\sigma^2}{n} [Q_{XZ} Q_Z^{-1} Q_{ZX}]^{-1}$$

Sarà stimata in maniera convergente con:

$$\widehat{VarAs}(\hat{\beta}_{vs}) = \hat{\sigma}^2 [X'Z(Z'Z)^{-1}Z'X]^{-1}$$

dove:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n} \hat{\varepsilon}' \hat{\varepsilon} \\ \hat{\varepsilon} &= y - X \hat{\beta}_{vs}\end{aligned}$$

Parte IV

Modelli dinamici

Capitolo 6

Variabili endogene ritardate

I modelli di regressione considerati fino a questo momento sono di *natura statica*, nel senso che sia la variabile dipendente che le variabili esplicative (di cui essa ne è funzione) sono osservate al medesimo istante temporale. Per contro, un modello *dinamico* implica una relazione non contemporanea fra la variabile dipendente ed una o più variabili esplicative. Un semplice esempio potrebbe essere il seguente:

$$y_t = bx_{t-3} + \varepsilon_t. \quad (6.1)$$

In questo caso l'effetto di una variazione in x è trasmesso ad y solo dopo tre periodi. Un modello economico è detto dinamico se è capace di evolvere nel tempo di sua propria iniziativa. Una semplice definizione di modello dinamico è la seguente:

Definizione 2. Un modello economico è detto dinamico se nell'equazione che lo definisce intervengono variabili datate a momenti diversi nel tempo.

Sebbene il modello (6.1) sia per definizione dinamico, esso non ne cattura pienamente lo spirito in quanto l'effetto che x esercita su y è realizzato in un'unica tappa (e per questo motivo non è di grande interesse). Nel Capitolo 7 vedremo modelli più ricchi dal punto di vista dinamico in cui la variabile dipendente è influenzata da due o più ritardi della stessa variabile esplicativa. Ad esempio il modello precedente potrebbe essere esteso nel seguente modo

$$y_t = b_1x_{t-1} + b_2x_{t-2} + b_3x_{t-3} + \varepsilon_t.$$

Un caso particolare e molto interessante di modello dinamico è quello in cui nell'insieme delle variabili esplicative sono presenti uno o più ritardi della variabile dipendente y . Si parla in questo caso di modelli dinamici a variabili endogene ritardate. In economia capita sovente di imbattersi in modelli a variabili endogene ritardate ed è quindi giusto analizzare quali sono le problematiche che questa classe di modelli pone dal punto di vista della stima e dei test.

6.1 Esempi economici

6.1.1 La formazioni d'abitudini

Prendiamo una funzione di consumo. Per Keynes il consumo è una funzione del reddito: $C = f(R)$

$$C_t = a + bR_t + \varepsilon_t \quad (6.2)$$

con $\varepsilon_t \sim iid(0, \sigma^2)$ ed il coefficiente b a rappresentare la propensione marginale al consumo: $0 < b < 1$. È questa una relazione statistica fra consumo e reddito. Il consumo dipende dal reddito in modo istantaneo. L'equazione (6.2) rappresenta piuttosto una relazione d'equilibrio di lungo termine. Questa relazione statica non è sufficiente per rappresentare il comportamento dei consumatori nel corto periodo. Supponiamo infatti, che il reddito diminuisca. Il consumo tenderà anch'esso a diminuire, ma non immediatamente a causa della formazione di abitudini: si è abituati ad un certo tenore di vita e anche se il reddito cambia le abitudini al consumo non reagiscono altrettanto velocemente. Le formazioni di abitudini sono rappresentate dal consumo passato.

$$C_t = a + cR_t + dC_{t-1} + \varepsilon_t .$$

C_{t-1} rappresenta la variabile endogena ritardata. Essa è aleatoria e quindi siamo in presenza di un modello di regressione stocastica! Come interpretiamo i parametri? Per rispondere a questa domanda poniamoci in $t = 0$ e supponiamo che da lungo tempo ($t < 0$) il sistema si trovi in equilibrio con reddito e consumo costanti. Ciò significa che prima di $t = 0$, $\Delta R_t = 0$ e $\Delta C_t = 0$. Scriviamo il nostro modello in termini di variazioni, trascurando momentaneamente il termine d'errore:

$$\Delta C_t = c\Delta R_t + d\Delta C_{t-1} .$$

Supponiamo che al tempo $t = 0$ il reddito aumenti di una unità (una volta per sempre), cioè

$$\begin{aligned} \Delta R_t &= 0 \text{ per } t < 0 \\ \Delta R_t &= 1 \text{ per } t = 0 \\ \Delta R_t &= 0 \text{ per } t > 0 \end{aligned}$$

Studiamo la variazione indotta sul consumo al tempo $t = 0$ e nei momenti successivi.

$$t = 0 : \quad \Delta C_0 = \underbrace{c\Delta R_0}_1 + \underbrace{d\Delta C_{-1}}_0 = c$$

c è la propensione marginale al consumo di corto periodo, l'effetto immediato.

$$\begin{aligned} t &= 1 : \quad \Delta C_1 = \underbrace{c\Delta R_1}_0 + \underbrace{d\Delta C_0}_c = dc \\ t &= 2 : \quad \Delta C_2 = \underbrace{c\Delta R_2}_0 + \underbrace{d\Delta C_1}_{dc} = d^2c \\ &\vdots \\ t &= n : \quad \Delta C_n = \underbrace{c\Delta R_n}_0 + \underbrace{d\Delta C_{n-1}}_{d^{n-1}c} = d^n c \end{aligned}$$

L'effetto cumulato dopo n periodi:

$$\begin{aligned}\Delta C_{TOT(n)} &= \Delta C_0 + \Delta C_1 + \dots + \Delta C_n \\ &= c(1 + d + d^2 + \dots + d^n)\end{aligned}$$

Quando $n \rightarrow \infty$, l'effetto totale ΔC_{TOT} è finito quando $|d| < 1$. In tal caso si ha: $\Delta C_{TOT} = \frac{c}{1-d} = b$ che chiameremo la propensione marginale di lungo periodo. Possiamo dedurla anche direttamente dall'equazione di partenza. Infatti supponiamo che tutte le variabili del modello siano in valore uguale al loro livello d'equilibrio di lungo termine:

$$\begin{aligned}R_t &= R^* \quad \forall t \\ C_t &= C^* \quad \forall t\end{aligned}$$

da cui

$$C^* = a + cR^* + dC^* \Rightarrow (1-d)C^* = a + cR^*$$

e per finire

$$\begin{aligned}C^* &= \frac{a}{1-d} + \frac{c}{1-d}R^* \\ &= a^* + c^*R^*\end{aligned}$$

Nell'equazione $C_t = a + cR_t + dC_{t-1} + \varepsilon_t$

- C_t è la variabile spiegata, detta anche endogena,
- R_t è una variabile esplicativa esogena,
- C_{t-1} è pure una variabile esplicativa: si tratta di una variabile endogena ritardata.

In contrasto, la formulazione

$$C_t = a + b_1R_t + b_2R_{t-1} + \varepsilon_t$$

è pure un modello dinamico ma senza variabile endogena ritardata. La variabile ritardata è la variabile esogena

$$\Delta C_t = b_1\Delta R_t + b_2\Delta R_{t-1}$$

$$\begin{aligned}\Delta R_0 = 1 \Rightarrow \Delta C_0 &= b_1 \underbrace{\Delta R_0}_1 + b_2 \underbrace{\Delta R_{-1}}_0 = b_1 \\ \Delta C_1 &= b_1 \underbrace{\Delta R_1}_0 + b_2 \underbrace{\Delta R_0}_1 = b_2 \\ \Delta C_2 &= b_1 \underbrace{\Delta R_2}_0 + b_2 \underbrace{\Delta R_1}_0 = 0\end{aligned}$$

da cui deduciamo

$$\Delta C_{TOT} = \Delta C_0 + \Delta C_1 = b_1 + b_2$$

6.1.2 Il modello d'aggiustamento parziale

In una data industria il livello desiderato degli stock, notato S_t^* , è dato da

$$S_t^* = a + bV_t + \varepsilon_t \quad (6.3)$$

Gli imprenditori hanno un comportamento d'aggiustamento parziale: colmano soltanto una parte della differenza tra il livello attuale e quello desiderato:

$$\underbrace{S_t - S_{t-1}}_{\substack{\text{variazione} \\ \text{(o investimento)} \\ \text{negli stock}}} = \lambda \underbrace{(S_t^* - S_{t-1})}_{\substack{\text{distanza che ci} \\ \text{separa dall'obiettivo}}} \quad \text{dove } 0 < \lambda < 1 \quad (6.4)$$

S^* non è osservabile. Le sole variabili osservabili sono S_t e V_t con i loro valori passati. In questo tipo di problemi bisogna fare 3 cose:

1. Derivare un'equazione stimabile, cioè un'equazione nella quale intervengono esclusivamente variabili effettivamente osservate.
2. Studiare l'identificazione dei parametri del modello di partenza (detto modello strutturale) a partire dai coefficienti stimati nell'equazione stimabile.
3. Studiare le proprietà statistiche dell'equazione stimabile e proporre un metodo di stima appropriato.

Consideriamo questi tre punti in dettaglio.

1. L'equazione stimata

Dalla (6.4):

$$\begin{aligned} \lambda S_t^* &= S_t - (1 - \lambda)S_{t-1} \\ \lambda a + \lambda bV_t + \lambda \varepsilon_t &= S_t - (1 - \lambda)S_{t-1} \\ S_t &= \lambda a + \lambda bV_t + (1 - \lambda)S_{t-1} + \lambda \varepsilon_t \\ S_t &= \beta_1 + \beta_2 V_t + \beta_3 S_{t-1} + u_t \end{aligned}$$

2. L'identificazione dei parametri strutturali

Conoscendo β_1 , β_2 e β_3 (dopo averli stimati) è possibile ottenere in maniera *unica* i parametri strutturali: a , b e λ .

$$\begin{aligned} \beta_3 &= (1 - \lambda) \Rightarrow \lambda = 1 - \beta_3 \\ \beta_2 &= \lambda b \Rightarrow b = \frac{\beta_2}{1 - \beta_3} \\ \beta_1 &= \lambda a \Rightarrow a = \frac{\beta_1}{1 - \beta_3} \end{aligned}$$

3. Le proprietà statistiche dell'equazione stimabile

- (a) L'equazione stimabile è un'equazione con *variabile endogena ritardata*.
- (b) Proprietà degli errori: $u_t = \lambda \varepsilon_t$.
Se $\varepsilon_t \sim iid(0, \sigma^2) \Rightarrow u_t \sim iid(0, \lambda^2 \sigma^2)$.

Si tratta quindi di un modello a variabile endogena ritardata ed errori indipendenti: i *m.q.o.* sono convergenti (e asintoticamente efficienti).

6.1.3 Modello ad aspettative adattive (adaptive expectation)

In una certa industria il volume delle scorte, S_t , è legato al volume previsto delle vendite, V_{t+1}^* , secondo la relazione

$$\begin{aligned} S_t &= a + bV_{t+1}^* + \varepsilon_t \\ \varepsilon_t &\sim iid(0, \sigma^2) \end{aligned} \quad (6.5)$$

Le imprese aggiornano le loro previsioni in funzione degli errori di previsione passati

$$\underbrace{V_{t+1}^* - V_t^*}_{\text{aggiornamento previsione}} = \gamma \underbrace{(V_t - V_t^*)}_{\text{errore passato}} \quad |\gamma| < 1 \quad (6.6)$$

1. L'equazione stimata

Dalla (6.6)

$$V_{t+1}^* - (1 - \gamma)V_t^* = \gamma V_t$$

Dalla (6.5)

$$\begin{cases} S_t = a + bV_{t+1}^* + \varepsilon_t & (a) \\ S_{t-1} = a + bV_t^* + \varepsilon_{t-1} & (b) \end{cases}$$

(a) - (1 - \gamma)(b):

$$\begin{aligned} S_t - (1 - \gamma)S_{t-1} &= \gamma a + b(V_{t+1}^* - (1 - \gamma)V_t^*) + \varepsilon_t - (1 - \gamma)\varepsilon_{t-1} \\ &= \gamma a + b\gamma V_t + \varepsilon_t - (1 - \gamma)\varepsilon_{t-1} \\ S_t &= \gamma a + b\gamma V_t + (1 - \gamma)S_{t-1} + \varepsilon_t - (1 - \gamma)\varepsilon_{t-1} \\ S_t &= \beta_1 + \beta_2 V_t + \beta_3 S_{t-1} + u_t \end{aligned}$$

dove $u_t = \varepsilon_t - (1 - \gamma)\varepsilon_{t-1}$.

2. L'identificazione dei parametri strutturali

$$\begin{aligned} \beta_3 &= 1 - \gamma \Rightarrow \gamma = 1 - \beta_3 \\ \beta_2 &= b\gamma \Rightarrow b = \frac{\beta_2}{1 - \beta_3} \\ \beta_1 &= \gamma a \Rightarrow a = \frac{\beta_1}{1 - \beta_3} \end{aligned}$$

3. Le proprietà statistiche dell'equazione stimabile

L'equazione stimabile è un modello a variabile endogena ritardata con errori che non sono indipendenti. Infatti $u_t = \varepsilon_t - (1 - \gamma)\varepsilon_{t-1}$: poiché $\varepsilon_t \sim iid(0, \sigma^2)$ otteniamo

$$\begin{aligned} E(u_t) &= \\ V(u_t) &= \\ Cov(u_t, u_{t-1}) &= \\ &= \end{aligned}$$

$$Cov(u_t, u_{t-s}) = 0 \quad \forall s > 1$$

Mostriamo che in un modello a variabile endogena ritardata ed errori autocorrelati i *m.q.o.* non sono convergenti. Sarà quindi necessario definire una nuova classe di stimatori.

6.2 La stima dei modelli a variabile endogena ritardata

La formulazione generale di questa classe di modelli è la seguente:

$$y_t = \underbrace{\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p}}_{p \text{ ritardi della variabile endogena}} + \underbrace{\beta_1 x_{1t} + \dots + \beta_K x_{Kt}}_{K \text{ variabili esogene}} + \varepsilon_t$$

ε_t è il solito errore non osservabile, $\varepsilon_t \sim (0, \sigma^2)$.

I problemi di stima di un tale modello si possono esaminare studiando il modello semplice

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t$$

che come abbiamo visto nella prima parte del corso è detto modello $AR(1)$ (modello autoregressivo di ordine 1).

Prima costatazione: la variabile endogena ritardata (variabile stocastica) è necessariamente correlata con gli *errori passati*. Infatti

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t \quad (6.7)$$

per il periodo precedente

$$\begin{aligned} y_{t-1} &= \alpha_1 y_{t-2} + \varepsilon_{t-1} \\ y_{t-2} &= \alpha_1 y_{t-3} + \varepsilon_{t-2} \\ &\vdots \end{aligned}$$

Quindi la variabile y_{t-1} della (6.7) dipende da y_{t-2} e da ε_{t-1} . y_{t-2} a sua volta dipende da ε_{t-2} e così via: y_{t-1} è dunque correlato con $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$

Seconda costatazione: siccome y_{t-1} dipende da $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ e quindi non direttamente da ε_t , y_{t-1} non è correlata con ε_t (nella (6.7)) quando gli errori ε_t sono indipendenti (non correlati).

Invece, quando gli errori ε_t sono correlati, la situazione è la seguente

$$\left. \begin{array}{ll} \varepsilon_t & \text{dipende da } \varepsilon_{t-1}, \varepsilon_{t-2}, \dots \\ y_{t-1} & \text{dipende da } \varepsilon_{t-1}, \varepsilon_{t-2}, \dots \end{array} \right\} \Rightarrow y_{t-1} \text{ e } \varepsilon_t \text{ nella (6.7) sono correlati!}$$

In conclusione: il modello a variabile endogena ritardata è un modello stocastico non correlato con gli errori quando quest'ultimi sono indipendenti (in tal caso i *m.q.o.* sono ok) ed invece si tratta di un modello di regressione stocastica dipendente quando quest'ultimi sono correlati (i *m.q.o.* non sono convergenti).

6.2.1 Quando gli errori sono indipendenti

$$y_t = \alpha y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma^2). \quad (6.8)$$

y_{t-1} e ε_t sono non correlati: $Cov(y_{t-1}, \varepsilon_t) = E(y_{t-1} \varepsilon_t) = 0$

Lo stimatore dei *m.q.o.* applicato al modello (6.8) è semplicemente

$$\hat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}.$$

$\hat{\alpha}$ non è uno stimatore lineare. Non è possibile calcolare il suo valore atteso e la sua varianza. Le sole proprietà verificabili sono le proprietà asintotiche.

1. Convergenza

$$\hat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \alpha + \frac{\frac{1}{n} \sum \varepsilon_t y_{t-1}}{\frac{1}{n} \sum y_{t-1}^2}$$

Il *p lim* del numeratore $\frac{1}{n} \sum \varepsilon_t y_{t-1}$ corrisponde alla covarianza empirica nel campione tra ε_t e y_{t-1} : esso converge alla covarianza della popolazione $cov(\varepsilon_t, y_{t-1})$.

Il denominatore $\frac{1}{n} \sum y_{t-1}^2$ corrisponde alla varianza empirica del campione. Il suo limite in probabilità è $E(y_t^2) = \frac{\sigma^2}{1-\alpha^2}$ che corrisponde anche alla varianza di y_t (perché?).

2. Distribuzione asintotica

$$\sqrt{n}(\hat{\alpha} - \alpha) = \underbrace{\left(\frac{1}{n} \sum y_{t-1}^2\right)^{-1}}_{p \lim \dots = \frac{\sigma^2}{1-\alpha^2}} \underbrace{\frac{1}{\sqrt{n}} \sum \varepsilon_t y_{t-1}}_{\xrightarrow{D} N(0, \sigma^4/(1-\alpha^2))} \xrightarrow{D} N(0, 1-\alpha^2)$$

La varianza asintotica è dunque

$$VarAs(\hat{\alpha}) = \frac{1-\alpha^2}{n}$$

che potrebbe essere stimato con

$$\widehat{VarAs}(\hat{\alpha}) = \frac{1-\hat{\alpha}^2}{n}$$

Se a torto utilizzassimo la formula per la varianza di $\hat{\alpha}$ data dai *m.q.o.*:

$$\begin{aligned} \hat{Var}(\hat{\alpha}) &= \frac{\hat{\sigma}^2}{\sum y_{t-1}^2}, \\ \hat{\sigma}^2 &= \frac{1}{n-1} \sum \hat{\varepsilon}_t^2 = \frac{1}{n-1} \sum (y_t - \hat{\alpha} y_{t-1})^2 \\ \hat{Var}(\hat{\alpha}) &\simeq \frac{\frac{1}{n} \sum (y_t^2 - 2\hat{\alpha} y_t y_{t-1} + \hat{\alpha}^2 y_{t-1}^2)}{\sum y_{t-1}^2} \\ &= \frac{\frac{1}{n} \sum y_t^2 - \hat{\alpha} \sum y_t y_{t-1}}{\sum y_{t-1}^2} \simeq \frac{1}{n} (1 - \hat{\alpha}^2) \end{aligned}$$

In pratica dunque, quando gli errori sono indipendenti, si applicano tutte le formule dei *m.q.o.*, anche per le varianze. Il metodo dei *m.q.* è la migliore cosa che si possa fare. Sotto l'ipotesi di normalità lo stimatore dei minimi quadrati è essenzialmente uguale allo stimatore *ML* (Massima Verosimiglianza) e quindi asintoticamente efficiente.

6.2.2 Quando gli errori sono autocorrelati

Studiamo il comportamento dello stimatore dei *m.q.o.* nel caso in cui gli errori siano autocorrelati, ed in particolare nel caso in cui l'errore ε_t segua un processo di tipo *AR*(1).

$$y_t = \alpha y_{t-1} + \varepsilon_t \quad (6.9)$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (6.10)$$

dove $|\rho| < 1$, $u_t \sim iid(0, \sigma^2)$.

Combinando la (6.9) e (6.10)

$$\begin{cases} y_t = \alpha y_{t-1} + \varepsilon_t & (a) \\ y_{t-1} = \alpha y_{t-2} + \varepsilon_{t-1} & (b) \end{cases}$$

(a) - ρ (b):

$$\begin{aligned} y_t - \rho y_{t-1} &= \alpha y_{t-1} - \alpha \rho y_{t-2} + \varepsilon_t - \rho \varepsilon_{t-1} \\ y_t &= \underbrace{(\alpha + \rho)}_{\beta_1} y_{t-1} + \underbrace{-\alpha \rho}_{\beta_2} y_{t-2} + u_t \end{aligned} \quad (6.11)$$

Questo modello (6.11) è un modello con variabile endogena ritardata ed errori indipendenti. I *m.q.o.* sull'equazione (6.11) produrrebbero delle stime convergenti di β_1 e β_2 . Sfortunatamente conoscendo i coefficienti β_1 e β_2 non si ottiene una soluzione unica di α e ρ : α e ρ non sono identificabili! Ad esempio

$$\left. \begin{aligned} \beta_1 &= \alpha + \rho = \frac{3}{4} \\ -\beta_2 &= \alpha \rho = \frac{1}{8} \end{aligned} \right\} \Rightarrow \text{due soluzioni} \quad \begin{aligned} \alpha &= 0.5, \quad \rho = 0.25 \\ \alpha &= 0.25, \quad \rho = 0.5 \end{aligned}$$

Se lo scopo è la previsione, il modello (6.11) è la miglior cosa che possiate fare. Per la stima dei parametri α e ρ il modello (6.11) non è di alcun aiuto.

6.2.2.1 La stima di α con i *m.q.o.* nella (6.9)

Partiamo dallo stimatore dei *m.q.o.* di α

$$\hat{\alpha} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$$

e sostituiamo y_t con l'espressione data dalla (6.11), ottenendo in tal modo

$$\begin{aligned} \hat{\alpha} &= \frac{\sum ((\alpha + \rho) y_{t-1} - \alpha \rho y_{t-2} + u_t) y_{t-1}}{\sum y_{t-1}^2} \\ &= \alpha + \rho - \alpha \rho \underbrace{\frac{\sum y_{t-2} y_{t-1}}{\sum y_{t-1}^2}}_{\approx \hat{\alpha}} + \frac{\frac{1}{n} \sum u_t y_{t-1}}{\frac{1}{n} \sum y_{t-1}^2} \end{aligned}$$

Poiché $E(y_t) = 0$, $V(y_t) = E(y_t^2)$. Inoltre, sappiamo che

$$\begin{aligned} p \lim \frac{1}{n} \sum u_t y_{t-1} &= 0 \\ p \lim \frac{1}{n} \sum y_{t-1}^2 &= V(y_t) \end{aligned}$$

Quindi

$$\begin{aligned} p \lim \hat{\alpha} &= \alpha + \rho - \alpha \rho p \lim \hat{\alpha} + \frac{p \lim \frac{1}{n} \sum u_t y_{t-1}}{p \lim \frac{1}{n} \sum y_{t-1}^2} \\ &= \alpha + \rho - \alpha \rho p \lim \hat{\alpha} \end{aligned}$$

da cui concludiamo che

$$p \lim \hat{\alpha} = \frac{\alpha + \rho}{1 + \alpha \rho} = \alpha + \underbrace{\frac{\rho(1 - \alpha^2)}{1 + \alpha \rho}}_{\text{distorsione asintotica}} \neq \alpha, \quad \text{quando } \rho \neq 0.$$

Osserviamo quindi che

- nel caso in cui $\rho \neq 0$, lo stimatore dei minimi quadrati sul modello (6.11) è distorto: la distorsione asintotica non scompare all'aumentare dell'informazione all'infinito.
- $\hat{\alpha}$ è convergente solo quando $\rho = 0$, cioè quando gli errori ε_t sono indipendenti.
- Dato che ρ e α sono positivi e compresi tra 0 e 1, la distorsione è positiva.

Conclusion: Nel caso di variabili endogene ritardate ed errori autocorrelati i *m.q.o.* non sono convergenti!

6.2.2.2 Come stimare α

Notiamo che se il termine ρ fosse conosciuto la (6.9) potrebbe scriversi

$$y_t - \rho y_{t-1} = \alpha(y_{t-1} - \rho y_{t-2}) + u_t \quad (6.12)$$

$$\tilde{y}_t = \alpha \tilde{y}_{t-1} + u_t \quad (6.13)$$

Il nuovo modello così trasformato sarebbe un modello a variabile endogena ritardata con errori indipendenti e, come appena dimostrato nel paragrafo (6.2.1), α sarebbe stimabile tramite i *m.q.o.*. Come possiamo stimare ρ ? Ci troviamo in una situazione simile a quella incontrata nel paragrafo (3.1.3) quando abbiamo utilizzato i *m.q.g.* a due tappe stimando dapprima il parametro ρ tramite la regressione ausiliaria (3.3). Si potrebbe quindi pensare di fare la stessa cosa e cioè:

- [1] Stimare con i *m.q.o.* l'equazione (6.9)

$$y_t = \alpha y_{t-1} + \varepsilon_t.$$

- [2] Prendere i residui stimati $\hat{\varepsilon}$ dell'equazione (6.9) e stimare ρ applicando i *m.q.o.* all'equazione

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + u_t.$$

[3] Stimare α applicando i *m.q.o.* al modello trasformato (6.13).

Tuttavia abbiamo dimostrato che i *m.q.o.* sul modello (6.9) (passo [1]) non sono convergenti. Questo implica una distorsione nella stima degli $\hat{\varepsilon}_t$ e a sua volta anche nella stima di ρ nel passo [2]:

$$\hat{\rho} = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2}.$$

Come vedremo in seguito la soluzione sarà quella di modificare il passo [1], rimpiazzando i *m.q.o.* con uno stimatore convergente quale ad esempio lo stimatore delle variabili strumentali.

6.2.3 Il test di autocorrelazione

Consideriamo il modello generale

$$\begin{aligned} y_t &= \underbrace{x_t' \beta}_{K \text{ variabili esogene}} + \underbrace{\alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p}}_{p \text{ ritardi nella variabile endogena}} + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + u_t \quad u_t \sim iid(0, \sigma^2) \end{aligned} \quad (6.14)$$

È dunque un modello a variabile endogena ritardata. Desideriamo testare l'ipotesi nulla di assenza di autocorrelazione negli errori ε_t . Formalmente $H_0 : \rho = 0$.

Osservazione 5. In un modello a variabile endogena ritardata il test *DW* non è appropriato poiché necessariamente distorto in favore dell'indipendenza. Non va assolutamente applicato.

Al suo posto si utilizza un test proposto da Durbin e chiamato test "h". La procedura del test è la seguente.

- Si stima il modello sotto l'ipotesi nulla $\rho = 0$, vale a dire con i *m.q.o.*, ottenendo $\hat{\beta}, \hat{\alpha}_1, \dots, \hat{\alpha}_p, \hat{V}(\hat{\alpha}_1)$ e costruendo

$$\hat{\varepsilon}_t = y_t - x_t' \hat{\beta} - \hat{\alpha}_1 y_{t-1} - \dots - \hat{\alpha}_p y_{t-p}.$$

- Si calcola la quantità

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n \hat{V}(\hat{\alpha}_1)}}$$

dove $\hat{\rho}$ è uno stimatore convergente di ρ sotto H_0 quale ad esempio

–

$$\hat{\rho} = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_{t-1}^2}$$

ottenuto dalla regressione $\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \text{errore}$,

– oppure $\hat{\rho} = 1 - \frac{1}{2}d$, dove d è la quantità del test *DW*.

Si dimostra che asintoticamente $h \sim N(0, 1)$ e quindi, al 5% di significatività:

$$\begin{aligned} |h| &> 1.96 \Rightarrow \text{rifiuto } H_0 \\ |h| &< 1.96 \Rightarrow \text{accetto } H_0 \end{aligned}$$

6.2.4 Stima convergente ed asintoticamente efficiente

Riprendendo il modello (6.14) consideriamo un modello con un solo ritardo e con K variabili esplicative esogene

$$y_t = x_t' \beta + \alpha y_{t-1} + \varepsilon_t \quad (6.15)$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad u_t \sim iid(0, \sigma^2) \quad (6.16)$$

Discutiamo una stima convergente dei parametri dell'equazione (6.15) ottenuta con il metodo delle variabili strumentali.

- Poiché il numero di variabili esplicative è $K + 1$ dobbiamo trovare almeno $K + 1$ strumenti.
- Le K variabili x_t' sono esogene e conseguentemente indipendenti dagli errori e quindi i loro stessi strumenti.
- Il compito si riconduce a trovare strumenti per y_{t-1} !

Ammettiamo che il modello contenga la costante ed almeno una variabile esogena. Quali strumenti per y_{t-1} si possono prendere i valori ritardati in $t - 1$ delle variabili esogene (escludendo la costante!), che indichiamo con \tilde{x}_t' . Abbiamo in totale $K + (K - 1)$ strumenti raccolti nella variabile $z_t' := [x_t' \tilde{x}_t']$. Con questi strumenti otteniamo una stima convergente di β, α ed inoltre anche di ρ tramite la regressione ausiliaria

$$\hat{\rho} = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_{t-1}^2} \quad (6.17)$$

dell'equazione

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \text{errore}$$

dove $\hat{\varepsilon}_t = y_t - x_t' \hat{\beta}_{vs} - \hat{\alpha}_{vs} y_{t-1}$.

6.2.4.1 Stima efficiente

La tappa precedente non fornisce degli stimatori efficienti in quanto traslascia la struttura di varianza/covarianza degli errori.

Idea: avendo una stima convergente di ρ si può pensare ad una stima coi *m.q.g.*, vale a dire ad una stima coi *m.q.o.* del modello trasformato (confronta 6.2.2.2)

$$y_t - \rho y_{t-1} = (x_t' - \rho x_{t-1}') \beta + \alpha (y_{t-1} - \rho y_{t-2}) + u_t \quad (6.18)$$

rimpiazzando ρ con $\hat{\rho}$ dalla (6.17).

Nel suo insieme questa procedura è simile ai *m.q.g.* realizzabili (o a due tappe) descritta in precedenza ma con un'importante differenza: anziché utilizzare i *m.q.o.* nel primo passo abbiamo utilizzato il metodo delle variabili strumentali.

Osservazione 6. Quando nel modello ci sono delle variabili endogene ritardate i *m.q.g.* realizzabili *non assicurano* l'efficienza asintotica. L'efficienza asintotica è ottenuta stimando sempre con i *m.q.o.* il modello trasformato (6.18) ma con l'aggiunta di una variabile esplicative: $\hat{\varepsilon}_{t-1}$. In altre parole il modello trasformato da utilizzare è

$$y_t - \rho y_{t-1} = (x_t' - \rho x_{t-1}') \beta + \alpha (y_{t-1} - \rho y_{t-2}) + d \hat{\varepsilon}_{t-1} + u_t. \quad (6.19)$$

Il passaggio da (6.18) a (6.19) si interpreta nel modo seguente. Definiamo

$$\rho = \widehat{\rho} + \underbrace{(\rho - \widehat{\rho})}_d = \widehat{\rho} + d.$$

e sostituiamo nella (6.18)

$$\begin{aligned} y_t - (\widehat{\rho} + d) y_{t-1} &= [x'_t - (\widehat{\rho} + d) x'_{t-1}] \beta + [y_{t-1} - (\widehat{\rho} + d) y_{t-2}] \alpha + u_t \\ y_t - \widehat{\rho} y_{t-1} &= (x'_t - \widehat{\rho} x'_{t-1}) \beta + \alpha (y_{t-1} - \widehat{\rho} y_{t-2}) + \underbrace{d(y_{t-1} - x'_{t-1} \beta - \alpha y_{t-2})}_{\varepsilon_{t-1}} + u_t. \end{aligned}$$

dove ε_{t-1} è sostituito da $\widehat{\varepsilon}_{t-1} = y_{t-1} - x'_{t-1} \widehat{\beta}_{vs} - \widehat{\alpha}_{vs} y_{t-2}$.

La stima con i *m.q.o.* di questo modello garantisce una stima efficiente di β , α e $\rho : \widehat{\rho}_{effic.} = \widehat{\rho} + \widehat{d}$ e di tutte le varianze covarianze asintotiche¹.

¹La varianza asintotica di $\widehat{\rho}_{effic.}$ è data solo dalla varianza di \widehat{d} . $\widehat{\rho}$ è da trascurare nel calcolo della varianza di $\widehat{\rho}_{effic.}$.

Capitolo 7

Modelli con ritardi distribuiti

7.1 Specificazione e proprietà. Moltiplicatori d'impatto, moltiplicatori dinamici, ritardo medio

7.1.1 Specificazione

7.1.1.1 Idea di fondo

In un modello statico l'influenza esercitata da una variabile è istantanea e immediata. Si esaurisce completamente nel momento in cui appare: $x_t \longrightarrow y_t$. In economia questa maniera d'interpretare una relazione non è molto attrattiva. Molto spesso infatti l'influenza di una variabile sull'altra non si esaurisce in un solo periodo, ma si realizza progressivamente nel tempo. In un modello dinamico

Grafico

Ci possono essere svariati motivi che giustificano una simile dinamica, fra cui possiamo citare

- Abitudini: i comportamenti abitudinari sono un altro motivo per cui è necessario introdurre modelli dinamici in economia. I consumatori hanno certi standard e si adeguano progressivamente.
- Costi inerenti ad ogni cambiamento.
- Ritardi nell'esecuzione di svariati compiti.

- Rigidità istituzionali.
- Presenza di scorte.
- Incertezza: spesso prima di reagire si attende per vedere se l'evento che ci sorprende è temporaneo o permanente.
- Tempo necessario per pianificare e realizzare un certo investimento: ogni decisione in economia richiede un certo tempo di reazione e di realizzazione. La reazione non è quasi mai immediata.

7.1.1.2 Formulazione generale

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \dots + \varepsilon_t$$

Le due ipotesi della formulazione generale sono:

1. i b_i sono tutti dello stesso segno
2. $\sum_{i=1}^{\infty} b_i = b$ finito. È questa una *condizione di stabilità*: c'è un equilibrio di lungo termine.

Idea

- Partire da una situazione iniziale di equilibrio (variabili costanti)
- darsi nel periodo successivo un incremento unitario di x e studiarne gli effetti su y (presenti e futuri).

All'inizio, in $t = 0$, ci troviamo in una situazione di equilibrio. Tutti gli incrementi di y e di x sono nulli per tutti i periodi anteriori a $t = 0$ incluso:

$$\Delta y_t = \Delta x_t = 0 \quad t = 0, -1, \dots$$

Nel periodo $t = 1$ consideriamo un aumento unitario di x , vale a dire

$$\Delta x_1 = 1.$$

In termine di incrementi, l'equazione del modello si scrive

$$\Delta y_t = b_0 \Delta x_t + b_1 \Delta x_{t-1} + b_2 \Delta x_{t-2} + \dots$$

da cui segue che

$$\begin{aligned} \Delta y_0 &= 0 \\ \Delta y_1 &= b_0 \underbrace{\Delta x_1}_{=1} + b_1 \underbrace{\Delta x_0}_{=0} + b_2 \underbrace{\Delta x_{-1}}_{=0} + \dots = b_0 \\ \Delta y_2 &= b_0 \underbrace{\Delta x_2}_{=0} + b_1 \underbrace{\Delta x_1}_{=1} + b_2 \underbrace{\Delta x_0}_{=0} + \dots = b_1 \\ \Delta y_3 &= b_0 \underbrace{\Delta x_3}_{=0} + b_1 \underbrace{\Delta x_2}_{=0} + b_2 \underbrace{\Delta x_1}_{=1} + \dots = b_2 \end{aligned}$$

b_0 è l'effetto immediato.

b_1 è l'effetto che si manifesta nel primo periodo successivo.

b_2 è l'effetto che si manifesta nel secondo periodo successivo.

Dopo s periodi, l'effetto cumulato è $\Delta y_1 + \Delta y_2 + \Delta y_3 + \dots \Delta y_s = \sum_{i=0}^{s-1} b_i$ e al limite, l'effetto totale di lungo termine è

$$\Delta_{y_{tot}} = \sum_{i=0}^{\infty} b_i = b$$

b è quindi l'effetto di lungo termine o *moltiplicatore dinamico*.

7.1.1.3 La relazione d'equilibrio di lungo termine

Per esprimere la relazione d'equilibrio di lungo termine, basta utilizzare nella relazione i valori d'equilibrio delle variabili:

$$\begin{aligned} y_t &= y^* \quad \forall t \text{ in equilibrio di lungo termine} \\ x_t &= x^* \quad \forall t \text{ in equilibrio di lungo termine} \end{aligned}$$

da cui ricaviamo

$$y^* = a + b_0 x^* + b_1 x^* + b_2 x^* + \dots = a + \underbrace{(b_0 + b_1 + \dots)}_b x^* = a + b x^*$$

e quindi vale pure

$$\frac{dy^*}{dx^*} = b.$$

Definiamo per ogni i la proporzione

$$p_i := \frac{b_i}{b}$$

Risulta che:

1. $p_i \geq 0$
2. $\sum_{i=1}^{\infty} p_i = 1$ I pesi p_i sono le probabilità di una distribuzione discreta

Questo ci permette di definire il profilo dei ritardi, cioè il grafico dei coefficienti "standardizzati".

Grafico

È possibile definire il ritardo medio

$$\bar{R} = \sum_{i=0}^{\infty} i p_i = \sum_{i=0}^{\infty} i \frac{b_i}{b} = \frac{\sum_{i=0}^{\infty} i b_i}{\sum_{i=0}^{\infty} b_i}$$

7.1.1.4 Problemi di stima

La formulazione generale del modello non è stimabile in quanto comporta un'infinità di parametri b_i . È quindi necessario *approssimare* la dinamica fondamentale con un modello che faccia intervenire un numero fisso e piccolo di parametri da stimare. Due sono le approssimazioni proposte nella letteratura.

1. Limitare il numero di ritardi, ammettendo che al di là di un certo ritardo i coefficienti siano nulli. Si parla in questo caso di *distribuzione finita dei ritardi*.
2. Ammettere un'infinità di ritardi i cui coefficienti sono generati da un meccanismo che dipende da un piccolo numero fisso di parametri (distribuzione infinita).

7.2 Distribuzione dei ritardi finita

7.2.1 Il caso generale

Come detto ipotizziamo che a partire da un certo ritardo tutti i coefficienti siano uguali a zero. Formalmente ciò equivale a dire che

$$b_i = 0 \quad i > r$$

dove r è un intero positivo che denota il ritardo massimo diverso da 0

$$y_t = a + b_0 x_t + \underbrace{b_1 x_{t-1} + b_2 x_{t-2} + \dots + b_r x_{t-r}}_{r - \text{ritardi}} + \varepsilon_t$$

Ammettiamo che il campione comporti n osservazioni (ad esempio osservazioni annuali dal 1961 al 2005 per un totale di $2005 - 1961 + 1 = n = 45$ osservazioni). Il numero effettivo di osservazioni, chiamiamolo n^* , disponibile per stimare il modello con r ritardi è $n - r$! Il numero di parametri da stimare è:

$$k = 1 + 1 + r = r + 2$$

ed i gradi di libertà saranno quindi pari a $g.l. = n - 2r - 2$. Il modello è una regressione di tipo abituale che verrà stimata con i *m.q.o.* (oppure con i *m.q.g.* se necessario). L'implementazione di questo modello comporta alcuni problemi e specificatamente

- La scelta arbitraria di r .
- La perdita di gradi di libertà (due per ogni ritardo supplementare) \rightarrow ciò comporta una stima meno precisa.
- Il rischio di collinearità che aumenta all'aumentare del numero di ritardi: $rg(X)$ non è pieno \Rightarrow poca precisione nella stima individuale dei parametri.

Per risolvere questi problemi:

- Per il primo: darsi un criterio di selezione del numero di ritardi.
- Per gli altri due Almon propone un metodo che prende il suo nome (vedi punto 7.2.2).

Un criterio utilizzato frequentemente per la selezione di r è il criterio di informazione di Akaike

$$AIC := n^* \ln \hat{\sigma}^2 + 2K$$

Si stima il modello per diversi valori di r e dunque di K . Per esempio $r = 1, 2, 3, \dots$ e si sceglie quello con l' AIC minimo.

7.2.2 Distribuzione di Almon (una donna!)

Come in precedenza si ammette un numero r fisso di ritardi. Un numero appropriato dipenderà dai dati (annuali, trimestrali, ...) dal numero di osservazioni e da quanto suggerito dalla teoria economica. Supponiamo che l'effetto sia positivo ($b > 0$). In certe applicazioni economiche si può ammettere che i coefficienti b_i decrescano linearmente, seguendo il profilo

Grafico

$$\begin{aligned} b_0 &= h_0 \\ b_1 &= h_0 + h_1 \\ b_2 &= h_0 + 2h_1 \\ &\vdots \\ b_r &= h_0 + rh_1 \end{aligned}$$

Conoscendo i due valori di h_0 e h_1 possiamo generare tutti i coefficienti b_i ! Introduciamo quindi questi b_i nel modello di partenza:

$$\begin{aligned} y_t &= a + b_0 x_t + b_1 x_{t-1} + \dots + b_r x_{t-r} + \varepsilon_t \\ &= a + h_0 x_t + (h_0 + h_1) x_{t-1} + \dots + (h_0 + rh_1) x_{t-r} + \varepsilon_t \\ &= a + h_0 \underbrace{(x_t + x_{t-1} + \dots + x_{t-r})}_{Z_{1t}} + h_1 \underbrace{(x_{t-1} + 2x_{t-2} + \dots + rx_{t-r})}_{Z_{2t}} + \varepsilon_t \end{aligned}$$

La regressione da stimare è dunque

$$y_t = a + h_0 Z_{1t} + h_1 Z_{2t} + \varepsilon_t$$

Se imponessimo la restrizione

$$\begin{aligned} b_{r+1} &= 0 \\ \Updownarrow \\ h_0 + (r+1)h_1 &= 0 \\ h_1 &= \frac{-h_0}{r+1} \end{aligned}$$

otterremmo

$$\begin{aligned} y_t &= a + h_0 Z_{1t} - \frac{h_0}{r+1} Z_{2t} + \varepsilon_t \\ &= a + h_0 \underbrace{\left(Z_{1t} - \frac{1}{r+1} Z_{2t} \right)}_{w_t} + \varepsilon_t \\ &= a + h_0 w_t + \varepsilon_t \end{aligned}$$

Nel caso specifico abbiamo generato i coefficienti b_i con un polinomio di primo grado: $b_i = h_0 + h_1 i$. Questo profilo è troppo restrittivo. Un profilo che all'inizio aumenta e poi diminuisce

Grafico

può essere approssimato con un polinomio di secondo grado:

$$b_i = h_0 + h_1 i + h_2 i^2$$

mentre un profilo

Grafico

con un polinomio di terzo grado

$$b_i = h_0 + h_1 i + h_2 i^2 + h_3 i^3$$

In genere Almon propone di utilizzare un polinomio di grado $s < r$:

$$b_i = h_0 + h_1 i + \dots + h_s i^s$$

7.3 Distribuzione dei ritardi infinita

7.3.1 Il modello

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \dots + \varepsilon_t$$

del quale analizziamo l'aspetto dinamico e non di stima (tralasciamo l'errore ε_t).

Esempio 10. La progressione geometrica decrescente

Grafico

$$\begin{aligned} b_0 &= b_0 \\ b_1 &= b_0 c & 0 < c < 1 \\ b_2 &= b_0 c^2 \\ &\vdots \\ b_s &= b_0 c^s \end{aligned}$$

- Effetto immediato: b_0 .
- Effetto di lungo periodo: $b = \sum b_i = b_0 (1 + c + c^2 + \dots) = b_0 \frac{1}{1-c}$.
- Ritardo medio: $\bar{R} = \frac{c}{1-c}$

Introduciamo $b_i = b_0 c^i$ nel modello:

$$y_t = a + b_0 x_t + b_0 c x_{t-1} + b_0 c^2 x_{t-2} + \dots$$

dipende da soli 3 parametri: a , b_0 , c .

$$\begin{array}{rcl} y_{t-1} & = & a + b_0 x_{t-1} + b_0 c x_{t-2} + b_0 c^2 x_{t-3} + \dots \\ & | \cdot c & \\ y_t - c y_{t-1} & = & a(1-c) + b_0 c \end{array}$$

da cui ricaviamo la trasformazione denominata *trasformazione di Koyck*

$$y_t = a^* + b_0 x_t + c y_{t-1}$$

In quest'esempio la distribuzione infinita di ritardi in x ha un'espressione equivalente in termini di due distribuzioni finite di ritardi:

- una in y_t di grado 1 e
- l'altra in x_t di grado 0.

Esempio 11. Estensione dell'Esempio 10

È immaginabile che la progressione geometrica decrescente si manifesti non già all'inizio ma soltanto a partire da un certo ritardo.

Grafico

$$\begin{array}{rcl} b_0 & = & \text{libero} \\ b_1 & = & \text{libero} \\ b_2 & = & \text{libero} \\ b_3 & = & b_2 c \\ b_4 & = & b_2 c^2 \\ & \vdots & \end{array}$$

In una situazione del genere, applicando la tecnica utilizzata precedentemente, si ottiene

$$\begin{array}{rcl} y_t & = & a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + b_2 c x_{t-3} + b_2 c^2 x_{t-4} + \dots \\ y_{t-1} & = & a + \quad b_0 x_{t-1} + b_1 x_{t-2} + b_2 x_{t-3} + b_2 c x_{t-4} + \dots \quad | \cdot c \\ y_t - c y_{t-1} & = & a^* + b_0 x_t + (b_1 - c b_0) x_{t-1} + (b_2 - c b_1) x_{t-2} + 0 + \dots \end{array}$$

Ciò che è importante notare in quest'ultimo esempio è che la distribuzione infinita in x ha un'espressione finita equivalente in termini di due distribuzioni finite, una in y di grado 1 ed una in x di grado 2.

7.3.2 Distribuzione razionale dei ritardi: idee introduttive

Nei due esempi precedenti la distribuzione dei ritardi infinita in x ha un'espressione equivalente in termini di due distribuzioni finite, una in x , l'altra in y . L'idea fondamentale di Jorgenson è che qualsiasi distribuzione infinita in x può essere approssimata a piacere tramite due distribuzioni finite di ritardi, una in x , l'altra in y .

7.3.3 L'operatore ritardo

Definizione 3. L'operatore di ritardo, notato L , fa corrispondere a x_t il valore precedente:

$$Lx_t = x_{t-1}$$

È un operatore lineare, poiché

1. $L(x_t + z_t) \underset{def}{=} x_{t-1} + z_{t-1} = Lx_t + Lz_t$
2. λ scalare, $L(\lambda x_t) = \lambda x_{t-1} = \lambda L(x_t)$

Definiamo ora l'elevamento a potenza:

$$L^2 x_t = L(Lx_t) = L(x_{t-1}) = x_{t-2}$$

Con questo operatore, il modello generale si scrive:

$$\begin{aligned} y_t &= a + b_0 x_t + b_1 Lx_t + b_2 L^2 x_t + \dots \\ &= a + \underbrace{(b_0 + b_1 L + b_2 L^2 + \dots)}_{B(L)_\infty} x_t \end{aligned}$$

$B(L)_\infty$ rappresenta il polinomio di grado infinito in L . Si verifichi quanto segue:

- $B(0) = b_0$, l'effetto immediato.
- $B(1) = b_0 + b_1 + b_2 + \dots = b$, l'effetto di lungo periodo.
- $B'(L) = b_1 + 2b_2 L + 3b_3 L^2 + \dots$
- $B'(1) = b_1 + 2b_2 + 3b_3 + \dots$

$$\Rightarrow \bar{R} = \frac{B'(1)}{B(1)}$$

7.3.4 Il modello dinamico di Jorgenson (distribuzione razionale dei ritardi)

Riprendiamo il modello generale

$$\begin{aligned}
 y_t &= a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \dots \\
 &= a + b_0 x_t + b_1 L x_t + b_2 L^2 x_t + \dots \\
 &= a + (b_0 + b_1 L + b_2 L^2 + \dots) x_t \\
 &= a + B(L)_\infty x_t
 \end{aligned} \tag{7.1}$$

Nei due esempi trattati il polinomio infinito $B(L)_\infty$ aveva un'espressione esatta equivalente in termini di due distribuzioni finite di ritardi in x e y . Jorgenson avvalendosi di un teorema fondamentale dell'algebra afferma che il polinomio infinito $B(L)_\infty$ può essere approssimato con precisione a piacere da un rapporto di due polinomi finiti.

$$B(L)_\infty \simeq \frac{\beta(L)_s}{\alpha(L)_r}, \quad s \text{ e } r \text{ finiti}$$

con

$$\begin{aligned}
 \beta(L)_s &= \beta_0 + \beta_1 L + \beta_2 L^2 + \dots + \beta_s L^s \\
 \alpha(L)_r &= 1 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_r L^r
 \end{aligned}$$

($\alpha_0 = 1$ è la cosiddetta normalizzazione).

Introduciamo l'approssimazione nel modello:

$$\begin{aligned}
 y_t &= a + \frac{\beta(L)_s}{\alpha(L)_r} x_t \\
 \alpha(L)_r y_t &= \alpha(L)_r a + \beta(L)_s x_t
 \end{aligned} \tag{7.2}$$

$$(1 + \alpha_1 L + \dots + \alpha_r L^r) y_t = a^* + (\beta_0 + \beta_1 L + \dots + \beta_s L^s) x_t \tag{7.3}$$

L'equazione (7.3) è anche chiamata *la forma canonica* del modello. a^* rappresenta l'espressione $\alpha(L)_r a$ ed è uguale ad $\alpha(1)_r a$ in quanto

$$\begin{aligned}
 \alpha(L)_r a &= (1 + \alpha_1 L + \dots + \alpha_r L^r) a \\
 \text{con } La &= a \\
 &= (1 + \alpha_1 + \dots + \alpha_r) a \\
 &= \alpha(1)_r a .
 \end{aligned}$$

L'equazione stimabile di questo modello:

$$\begin{aligned}
 y_t &= a^* + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_s x_{t-s} - \alpha_1 y_{t-1} - \dots - \alpha_r y_{t-r} + \varepsilon_t \\
 y_t &= a^* + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_s x_{t-s} + \alpha_1^* y_{t-1} + \dots + \alpha_r^* y_{t-r} + \varepsilon_t
 \end{aligned} \tag{7.4}$$

In pratica s e r non sono conosciuti, quindi si stima il modello per svariate coppie (s, r) , ad esempio la coppia $(1, 1); (1, 2); \dots$. Si sceglierà la coppia che dà migliori risultati in termini di significatività dei coefficienti e di valori del criterio d'informazione di Akaike.

Avendo stimato i polinomi $\alpha(L)$ e $\beta(L)$ (tralasciamo i rispettivi indici r ed s) si possono ottenere:

1. L'effetto immediato
2. L'effetto dinamico di lungo periodo
3. Il ritardo medio
4. Il profilo dei ritardi (ovvero i coefficienti del polinomio $B(L)_\infty$)

Per le prime 3 domande utilizziamo la relazione

$$B(L)_\infty = \frac{\beta(L)}{\alpha(L)}$$

per la quarta sfruttiamo la ricorrenza.

1. L'effetto immediato

$$b_0 = B(0)_\infty = \frac{\beta(0)}{\alpha(0)} = \frac{\beta_0}{1} = \beta_0$$

2. L'effetto di lungo periodo

$$b = b_0 + b_1 + b_2 + \dots = B(1)_\infty = \frac{\beta(1)}{\alpha(1)} = \frac{\beta_0 + \beta_1 + \dots + \beta_s}{1 + \alpha_1 + \dots + \alpha_r} = \frac{\beta_0 + \beta_1 + \dots + \beta_s}{1 - \alpha_1^* - \dots - \alpha_r^*}$$

3. Il ritardo medio. Ricordiamo che

$$\bar{R} = \frac{B'(1)}{B(1)}$$

Ora,

$$\begin{aligned} B'(L) &= \frac{\beta'(L)\alpha(L) - \beta(L)\alpha'(L)}{\alpha(L)^2} \\ &= \frac{\beta'(L)}{\alpha(L)} - \frac{\beta(L)}{\alpha(L)^2}\alpha'(L) \\ &= \frac{\beta(L)}{\alpha(L)} \left[\frac{\beta'(L)}{\beta(L)} - \frac{\alpha'(L)}{\alpha(L)} \right] \end{aligned}$$

e quindi essendo

$$B'(1) = \frac{\beta(1)}{\alpha(1)} \left[\frac{\beta'(1)}{\beta(1)} - \frac{\alpha'(1)}{\alpha(1)} \right]$$

otteniamo che

$$\bar{R} = \frac{B'(1)}{B(1)} = \frac{\beta'(1)}{\beta(1)} - \frac{\alpha'(1)}{\alpha(1)}$$

Esempio 12. Con dati trimestrali abbiamo stimato la seguente funzione di consumo:

$$C_t = 160 + 0.30R_t + 0.20R_{t-1} + 0.14R_{t-2} + 0.2C_{t-1}$$

- La forma canonica

$$\begin{aligned} C_t - 0.2C_{t-1} &= 160 + 0.30R_t + 0.20R_{t-1} + 0.14R_{t-2} \\ \underbrace{(1 - 0.2L)C_t}_{\alpha(L)_1} &= 160 + \underbrace{(0.30 + 0.20L + 0.14L^2)R_t}_{\beta(L)_2} \end{aligned}$$

Calcoliamo:

- $\beta(1) = 0.3 + 0.2 + 0.14 = 0.64$
- $\beta'(L) = 0.2 + 0.28L$
- $\beta'(1) = 0.48$
- $\alpha(0) = 1$ (sempre!), $\alpha(1) = 0.8$
- $\alpha'(L) = -0.2$
- $\alpha'(1) = -0.2$

- L'effetto immediato: $b_0 = \beta(0) = 0.3$
- L'effetto dinamico di lungo periodo:

$$b = B(1)_\infty = \frac{\beta(1)}{\alpha(1)} = \frac{0.48}{0.64} = 0.8 \text{ (propensione marginale di L.P.)}$$

- Il ritardo medio

$$\begin{aligned} \bar{R} &= \frac{B'(1)}{B(1)} = \frac{\beta'(1)}{\beta(1)} - \frac{\alpha'(1)}{\alpha(1)} \\ &= \frac{0.48}{0.64} + \frac{0.2}{0.8} = 1 \end{aligned}$$

- Il profilo dei ritardi

Ammettiamo un punto di partenza d'equilibrio di lungo periodo con tutti gli incrementi passati uguali a 0 per ogni $t < 0$.

$$\Delta C_t = \Delta R_t = 0 \quad \forall t < 0$$

Scriviamo la nostra equazione in termini di incrementi:

$$\Delta C_t = 0.3\Delta R_t + 0.2\Delta R_{t-1} + 0.14\Delta R_{t-2} + 0.2\Delta C_{t-1}$$

In $t = 0$, consideriamo un aumento unitario del reddito

$$\Delta R_0 = 1$$

e studiamo il suo influsso sul consumo presente e futuro

$$\begin{aligned} \Delta C_0 &= \underbrace{0.3\Delta R_0}_1 + \underbrace{0.2\Delta R_{-1}}_0 + \underbrace{0.14\Delta R_{-2}}_0 + \underbrace{0.2\Delta C_{-1}}_0 = 0.3 = b_0 \\ \Delta C_1 &= \underbrace{0.3\Delta R_1}_0 + \underbrace{0.2\Delta R_0}_1 + 0.14\Delta R_{-1} + \underbrace{0.2\Delta C_0}_{0.3} = 0.26 = b_1 \\ \Delta C_2 &= \underbrace{0.3\Delta R_2}_0 + \underbrace{0.2\Delta R_1}_0 + \underbrace{0.14\Delta R_0}_1 + \underbrace{0.2\Delta C_1}_{0.26} = 0.192 = b_2 \\ \Delta C_3 &= \underbrace{0.2\Delta C_2}_{0.192} = 0.0384 = b_3 \\ \Delta C_4 &= \underbrace{0.2\Delta C_3}_{0.0384} = \dots \end{aligned}$$

Verifichiamo:

$$\begin{aligned}
 b &= b_0 + b_1 + b_2 + \dots \\
 &= 0.3 + 0.26 + 0.192 (1 + 0.2 + 0.2^2 + \dots) \\
 &= 0.56 + \frac{0.192}{1 - 0.2} = 0.8
 \end{aligned}$$

grafico

7.3.4.1 L'equazione d'equilibrio di lungo periodo

Poniamo

$$\begin{cases} y_t = y^* \\ x_t = x^* \end{cases} \quad \forall t$$

Dalla formulazione di partenza (7.2)

$$\begin{aligned}
 \alpha(L)y_t &= a^* + \beta(L)x_t \\
 \alpha(L)y^* &= a^* + \beta(L)x^* \\
 \alpha(1)y^* &= a^* + \beta(1)x^*
 \end{aligned}$$

da cui otteniamo la relazione di lungo periodo

$$\begin{aligned}
 y^* &= \frac{a^*}{\alpha(1)} + \frac{\beta(1)}{\alpha(1)} x^* \\
 y^* &= a + bx^*
 \end{aligned} \tag{7.5}$$

dove b rappresenta l'effetto di lungo periodo. L'equazione (7.5) è derivabile direttamente dalla (7.1), ovvero

$$\begin{aligned}
 y_t &= a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \dots \\
 y^* &= a + b_0 x^* + b_1 x^* + b_2 x^* + \dots \\
 &= a + (b_0 + b_1 + b_2 + \dots) x^* \\
 &= a + bx^*
 \end{aligned}$$

Nell'esempio:

$$\begin{aligned}
 C^* &= 160 + 0.3R^* + 0.2R^* + 0.14R^* + 0.2C^* \\
 C^* &= \frac{160}{0.8} + \frac{0.64}{0.8} R^* \\
 C^* &= 200 + 0.8R^*
 \end{aligned}$$

Se in $t - 1$ ponessimo

$$\varepsilon_{t-1}^* = C_{t-1} - 200 - 0.8R_{t-1} \quad (7.6)$$

otterremmo il cosiddetto scarto dall'equilibrio di lungo termine (o errore) in $t - 1$.

Capitolo 8

Serie macroeconomiche non stazionarie

8.1 Stazionarietà e non stazionarietà

8.1.1 Stazionarietà

Una serie storica è stazionaria se soddisfa tre principi:

1. Tende a ritornare alla media
2. La varianza è costante (non dipende dal tempo)
3. La covarianza fra due termini della serie è unicamente funzione della distanza che separa i due termini e tende verso 0 quando questa distanza aumenta.

Grafico

Esempio 13. Rumore bianco (White Noise)

$$\varepsilon_t \sim iid(0, \sigma^2)$$

Esempio 14. Il modello AR(1)

$$\begin{cases} \varepsilon_t = \rho\varepsilon_{t-1} + u_t \\ u_t \sim iid(0, \sigma^2) \\ |\rho| < 1 \end{cases}$$

- $E(\varepsilon_t) = 0$.
- $V(\varepsilon_t) = \frac{\sigma^2}{1-\rho^2}$ indipendente da t .
- $Cov(\varepsilon_t, \varepsilon_{t-s}) = \frac{\sigma^2}{1-\rho^2} \rho^s \xrightarrow{s \rightarrow \infty} 0$.

Esempio 15. Il modello $ARMA(p, q)$

8.1.2 Non stazionarietà

Una serie storica è non stazionaria se:

1. Si scosta indefinitivamente dalla media
2. La varianza è proporzionale al tempo
3. La correlazione fra due termini della serie tende verso uno quando la distanza tra i due termini aumenta.

Grafico

Esempio 16. Cammino casuale o aleatorio (random walk). Nell'Esempio 14 poniamo $\rho = 1$:

$$\varepsilon_t = \varepsilon_{t-1} + u_t \sim \text{White Noise}$$

$$\begin{aligned} \varepsilon_0 &= 0 \\ \varepsilon_1 &= 0 + u_1 \\ \varepsilon_2 &= \varepsilon_1 + u_2 = u_1 + u_2 \\ \varepsilon_3 &= \varepsilon_2 + u_3 = u_1 + u_2 + u_3 \\ &\vdots \\ \varepsilon_t &= u_1 + u_2 + u_3 + \dots + u_t \end{aligned}$$

Possiamo allora calcolare valore atteso e varianza:

$$\begin{aligned} E(\varepsilon_t) &= 0 \\ V(\varepsilon_t) &= t\sigma^2 \\ Cor(\varepsilon_t, \varepsilon_{t-s}) &= \frac{Cov(\varepsilon_t, \varepsilon_{t-s})}{\sqrt{V(\varepsilon_t)V(\varepsilon_{t-s})}} = \frac{(t-s)}{\sqrt{t(t-s)}} \\ \lim_{t \rightarrow \infty} Cor(\varepsilon_t, \varepsilon_{t-s}) &= 1 \end{aligned}$$

Due autori sono arrivati alla conclusione che quasi tutte le variabili macroeconomiche sono non stazionarie (Nelson e Plosser). Si cerca quindi di rendere stazionarie le variabili prima di eseguire l'analisi econometrica.

Una serie $\{\varepsilon_t\}$ non stazionaria è detta integrata di ordine 1, e si scrive

$$\varepsilon_t \sim I(1)$$

se la differenza prima

$$d_t := \varepsilon_t - \varepsilon_{t-1}$$

è stazionaria (cioè $I(0)$). Nel nostro esempio è evidente che un Random Walk è $I(1)$ in quanto

$$\varepsilon_t - \varepsilon_{t-1} = u_t \sim WN$$

Per vedere se una serie è non stazionaria, stimo il modello $AR(1)$: $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$. Dopo aver stimato ρ conduco il test d'ipotesi $H_0 : \rho = 1$ (test della radice unitaria). Se accetto l'ipotesi, allora la serie è non stazionaria. C'è però un problema. Sotto l'ipotesi H_0 lo stimatore $\hat{\rho}$ non ha le proprietà statistiche che conosciamo. Torniamo al modello $AR(1)$:

$$\begin{aligned} \varepsilon_t &= \rho\varepsilon_{t-1} + u_t \\ \varepsilon_t - \rho\varepsilon_{t-1} &= u_t \\ (1 - \rho L)\varepsilon_t &= u_t \end{aligned}$$

una radice unitaria sarebbe quando trovo $\rho = 1$. In tal caso scriviamo anche $(1 - L) = \Delta$.

8.1.3 Una riparametrizzazione del modello dinamico classico (Jorgenson)

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \alpha_1^* y_{t-1} + \alpha_2^* y_{t-2} + \varepsilon_t \quad (8.1)$$

5 parametri da stimare.

- Effetto immediato: β_0
- Effetto di lungo termine: $b = \frac{\beta_0 + \beta_1 + \beta_2}{1 - \alpha_1^* - \alpha_2^*}$
- Relazione di lungo termine: $y_t = bx_t$.

Se in (8.1) adopero i valori stazionari:

$$\begin{cases} y_t = y^* \\ x_t = x^* \end{cases} \quad \forall t$$

$$\begin{aligned}
y^* &= (\beta_0 + \beta_1 + \beta_2) x^* + (\alpha_1^* + \alpha_2^*) y^* \\
y^* &= \frac{\beta_0 + \beta_1 + \beta_2}{1 - \alpha_1^* - \alpha_2^*} x^* \\
y^* &= bx^*
\end{aligned}$$

da cui otteniamo lo scarto dall'equilibrio di lungo termine (errore): $y_t - bx_t$.

- Prima di affrontare il caso particolare dato dall'equazione stimabile (8.1) consideriamo il caso generale dato dalla (7.4), ovvero

$$y_t = a^* + \underbrace{\beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_s x_{t-s}}_{p1} + \underbrace{\alpha_1^* y_{t-1} + \dots + \alpha_r^* y_{t-r}}_{p2} + \varepsilon_t$$

Consideriamo dapprima la riparametrizzazione della parte 1:

$$\begin{aligned}
p1 &= \beta_0 x_t + (\beta_0 + \beta_1 + \dots + \beta_s - \beta_0 - \beta_2 - \dots - \beta_s) x_{t-1} + \\
&\quad + (\beta_2 + \beta_3 + \dots + \beta_s - \beta_3 - \dots - \beta_s) x_{t-2} \\
&\quad + (\beta_3 + \dots + \beta_s - \beta_4 - \dots - \beta_s) x_{t-3} \\
&\quad \vdots \\
&\quad + (\beta_{s-1} + \beta_s - \beta_s) x_{t-(s-1)} \\
&\quad + \beta_s x_{t-s} \\
&= \beta_0 \Delta x_t + (\beta_0 + \beta_1 + \dots + \beta_s) x_{t-1} \\
&\quad - (\beta_2 + \dots + \beta_s) \Delta x_{t-1} \\
&\quad - (\beta_3 + \dots + \beta_s) \Delta x_{t-2} \\
&\quad \vdots \\
&\quad - (\beta_{s-1} + \beta_s) \Delta x_{t-(s-2)} \\
&\quad - \beta_s \Delta x_{t-(s-1)}
\end{aligned}$$

e ridefinendo i coefficienti delle nuove variabili esplicative otteniamo per questa prima parte

$$p1 = \delta_0 \Delta x_t + \gamma_1 x_{t-1} + \delta_1 \Delta x_{t-1} + \dots + \delta_{s-1} \Delta x_{t-(s-1)}.$$

La riparametrizzazione della parte 2 è quasi identica alla precedente.

Abbiamo infatti

$$\begin{aligned}
p2 &= (\alpha_1^* + \alpha_2^* + \dots + \alpha_r^* - \alpha_2^* - \dots - \alpha_r^*) y_{t-1} \\
&+ (\alpha_2^* + \alpha_3^* + \dots + \alpha_r^* - \alpha_3^* - \dots - \alpha_r^*) y_{t-2} \\
&+ (\alpha_3^* + \alpha_4^* + \dots + \alpha_r^* - \alpha_4^* - \dots - \alpha_r^*) y_{t-3} \\
&\vdots \\
&+ (\alpha_{r-1}^* + \alpha_r^* - \alpha_r^*) y_{t-(r-1)} \\
&+ \alpha_r^* y_{t-r} \\
&= (\alpha_1^* + \alpha_2^* + \dots + \alpha_r^*) y_{t-1} \\
&- (\alpha_2^* + \dots + \alpha_r^*) \Delta y_{t-1} \\
&- (\alpha_3^* + \dots + \alpha_r^*) \Delta y_{t-2} \\
&\vdots \\
&- (\alpha_{r-1}^* + \alpha_r^*) \Delta y_{t-(r-2)} \\
&- \alpha_r^* \Delta y_{t-(r-1)}
\end{aligned}$$

Anche in questo caso ridefiniamo i coefficienti dei Δy

$$p2 = (\alpha_1^* + \dots + \alpha_r^*) y_{t-1} + q_1 \Delta y_{t-1} + q_2 \Delta y_{t-2} + \dots + q_{r-1} \Delta y_{t-(r-1)}$$

Ricapitolando abbiamo che

$$y_t = a^* + \gamma_1 x_{t-1} + (\alpha_1^* + \dots + \alpha_r^*) y_{t-1} + \sum_{i=0}^{s-1} \delta_i \Delta x_{t-i} + \sum_{i=1}^{r-1} q_i \Delta y_{t-i} + \varepsilon_t$$

e sottraendo da entrambi i lati y_{t-1} si ottiene la forma finale

$$\begin{aligned}
y_t - y_{t-1} &= a^* + \gamma_1 x_{t-1} + \underbrace{(\alpha_1^* + \dots + \alpha_r^* - 1)}_{\gamma_2} y_{t-1} + \sum_{i=0}^{s-1} \delta_i \Delta x_{t-i} + \sum_{i=1}^{r-1} q_i \Delta y_{t-i} + \varepsilon_t \\
\Delta y_t &= a^* + \gamma_1 x_{t-1} + \gamma_2 y_{t-1} + \sum_{i=0}^{s-1} \delta_i \Delta x_{t-i} + \sum_{i=1}^{r-1} q_i \Delta y_{t-i} + \varepsilon_t
\end{aligned}$$

- Dalla (8.1):

$$\begin{aligned}
y_t &= \beta_0 x_t + (\beta_0 + \beta_1 + \beta_2 - \beta_0 - \beta_2) x_{t-1} + \beta_2 x_{t-2} + \\
&(\alpha_1^* + \alpha_2^* - \alpha_2^*) y_{t-1} + \alpha_2^* y_{t-2} + \varepsilon_t \\
y_t &= \underbrace{\beta_0}_{\delta_0} \underbrace{(x_t - x_{t-1})}_{\Delta x_t} + \underbrace{(\beta_0 + \beta_1 + \beta_2)}_{=\gamma_1} x_{t-1} - \underbrace{\beta_2}_{\delta_1} \underbrace{(x_{t-1} - x_{t-2})}_{\Delta x_{t-1}} + \\
&(\alpha_1^* + \alpha_2^*) y_{t-1} - \underbrace{\alpha_2^* (y_{t-1} - y_{t-2})}_{\substack{q_1 \\ \Delta y_{t-1}}} + \varepsilon_t \\
\underbrace{y_t - y_{t-1}}_{\Delta y_t} &= \gamma_1 x_{t-1} + \underbrace{(\alpha_1^* + \alpha_2^* - 1)}_{\gamma_2} y_{t-1} + \delta_0 \Delta x_t + \delta_1 \Delta x_{t-1} + q_1 \Delta y_{t-1} \quad (8.2) \\
\Delta y_t &= \gamma_1 x_{t-1} + \gamma_2 y_{t-1} + \gamma_3 \Delta x_t + \gamma_4 \Delta x_{t-1} + \gamma_5 \Delta y_{t-1} + \varepsilon_t \quad (8.3)
\end{aligned}$$

Quest'equazione comporta 5 parametri come la (8.1) identificabili. Stimata con i *m.q.o.* produce gli stessi risultati (trasformati) dei coefficienti e lo stesso *SS*.

$$\begin{aligned} y &= X\beta + \varepsilon \\ \beta &= A\gamma + L \end{aligned}$$

dove A rappresenta una matrice nota non singolare ed L un vettore noto. E' possibile risalire ad A ed L confrontando le due formulazioni (8.1) e (8.2) del modello. Infatti

$$\begin{aligned} y &= X(A\gamma + L) + \varepsilon \\ y &= \underbrace{XA}_{W}\gamma + \underbrace{XL}_{\text{conosciuto}} + \varepsilon \\ (y - XL) &= W\gamma + \varepsilon \end{aligned}$$

confrontando i termini dell'equazione (8.2) con ques'ultima espressione è possibile evincere il contenuto della matrice A e del vettore L .

$$L = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}; A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Inoltre dall'equazione (8.2) è possibile derivare direttamente il moltiplicatore di lungo termine come

$$b = \frac{\gamma_1}{-\gamma_2} = \frac{\beta_0 + \beta_1 + \beta_2}{1 - \alpha_1^* - \alpha_2^*} = \frac{\beta(1)}{\alpha(1)}$$

Riscriviamo ora la (8.2) nel seguente modo:

$$\begin{aligned} \Delta y_t &= \gamma_2 \left(y_{t-1} + \frac{\gamma_1}{\gamma_2} x_{t-1} \right) + \sum_{i=0}^1 \delta_i \Delta x_{t-i} + \sum_{i=1}^1 r_i \Delta y_{t-i} \\ \Delta y_t &= \gamma_2 (y_{t-1} - b x_{t-1}) + \sum_{i=0}^1 \delta_i \Delta x_{t-i} + \sum_{i=1}^1 r_i \Delta y_{t-i} \end{aligned} \quad (8.4)$$

- γ_2 : coefficiente di correzione dell'errore.
- $y_{t-1} - b x_{t-1}$: scarto (errore) dall'equilibrio di lungo termine, ε_{t-1}
- $\sum_{s=0}^1 \delta_s \Delta x_{t-s} + \sum_{s=1}^1 r_s \Delta y_{t-s}$: dinamica di corto periodo: tanti Δx_{t-i} quanti i numeri di ritardi (s), tanti Δy_{t-i} quanti i numeri di ritardi in y meno 1 ($r - 1$).

L'equazione (8.4) è chiamata modello a correzione d'errore: descrive come l'errore di equilibrio di lungo termine è corretto ed esprime la dinamica di corto periodo. Questa equazione è alla base dell'approccio delle serie storiche per il concetto di cointegrazione. L'equazione (8.4) non è direttamente stimabile in quanto il parametro b non è conosciuto.

L'equazione (8.4) non è direttamente stimabile in quanto il parametro b non è conosciuto.

8.1.4 Introduzione alla cointegrazione

Due serie non stazionarie possono ammettere una relazione stabile di lungo periodo, vale a dire che sebbene prese singolarmente le due serie tendano a divergere indefinitamente, lo scarto dell'una rispetto all'altra è qualcosa di stabile o più propriamente detto qualcosa di stazionario.

Grafico

Definizione 4. Due serie non stazionarie ed integrate d'ordine 1

$$\begin{aligned}y_t &\sim I(1) \\x_t &\sim I(1)\end{aligned}$$

si dicono cointegrate se esiste una relazione lineare tra le due

$$y_t - bx_t$$

che è stazionaria

$$y_t - bx_t \sim I(0)$$

8.1.4.1 La procedura di stima

La procedura di stima di serie storiche cointegrate comprende le seguenti tappe:

1. Verificare attraverso il test di radice unitaria che le due serie sono effettivamente $I(1)$.
2. Stimare la relazione stazionaria di lungo termine

$$y_t = bx_t + \varepsilon_t$$

e verificare che l'errore stimato $\hat{\varepsilon}_t$ è stazionario (che non abbia radice unitaria). Come visto precedentemente ε_t rappresenta lo scarto dall'equilibrio di lungo periodo. Generalmente l'errore ε_t non è *i.i.d.* ma può avere una dinamica anche complessa. Ciò che è richiesto affinché si possa parlare di cointegrazione è che ε_t sia un processo stazionario: nel corto periodo y_t e bx_t possono divergere ma nel lungo periodo devono nuovamente riavvicinarsi. Da qui la necessità di verificare (testare tramite test appropriati) la stazionarietà dell'errore stimato $\hat{\varepsilon}_t$.

3. Se l'ipotesi di stazionarietà del termine d'errore non è rigettata si introduce $\hat{\varepsilon}_{t-1}$ nel modello a correzione d'errore (vedi equazione (8.4)). In questo modello, tutte le variabili esplicative sono stazionarie:

$$\begin{aligned}\Delta x_{t-s} &\sim I(0) \\ \Delta y_{t-r} &\sim I(0) \\ \hat{\varepsilon}_{t-1} &\sim I(0)\end{aligned}$$

e quindi può essere stimato con i *m.q.o.*.

Osservazione 7. Questa procedura garantisce la convergenza in probabilità degli stimatori. Per ottenere l'efficienza asintotica occorre utilizzare il metodo di massima verosimiglianza (metodo di Johanson).