

Appunti (parziali) di Modelli Econometrici

2018/19

1. Indice fatto seguendo programma d'esame 2018\19

2. Libri:

- a. Stock e Watson capitoli:4, 5, 18
- b. Verbeek: capitoli 8,10, 12, "capitolo su endogeneità"

Sommario

Richiami.....	4
Capitolo 1. Richiami sulla funzione di regressione e sue proprietà	7
Il modello di regressione lineare multiplo.....	8
.1 Il modello di regressione lineare multiplo con uso della notazione matriciale	9
.2 Richiami sulle proprietà dello stimatore OLS dei parametri del modello di regressione lineare multipla	11
Normalità asintotica dei Beta	13
Funzione di regressione	15
Correttezza dei Beta OLS , consistenza dei Beta OLS	15
Operatore PLIM.....	16
Distribuzione Normale bivariata e multivariata	17
1.3 Inferenza robusta all'eteroschedasticità.....	21
Errori standard robusti rispetto all'eteroschedasticità	22
1.4 Test di corretta specificazione del modello di regressione multipla	23
Test di White	24
Formula di White (stimatore White di AVAR).....	25
Test di Breush-Pagan.....	25
Test F (per significatività del modello)	26
Test di Jarque–Bera	26
Test RESET: test di Ramsey	27
Test di Chow	29
Variabili strumentali.....	31
Metodo delle variabili strumentali.....	34
Metodo dei minimi quadrati a due stadi	35
Test di Sargan (o test per la sovraidentificazione 357/698 SeW (arrivato qui)	35
1.5 Esercitazioni al computer in GRET.....	36

Capitolo 2. Modelli econometrici per serie temporali: modelli statici e modelli dinamici	37
Le serie storiche.....	37
Il processo stocastico (a tempo discreto)	37
Stazionarietà del processo asintotico	38
Il processo White Noise	39
La notazione LAG	39
2.1 Proprietà dello stimatore OLS dei parametri del modello di regressione lineare statico	39
2.2 Inferenza robusta all'eteroschedasticità e alla correlazione seriale dei residui.....	40
2.3 Test di correlazione seriale dei residui	40
2.4 1) Il modello AR(1) ed il modello di regressione lineare dinamico.....	40
2.5 Proprietà dello stimatore OLS dei parametri del modello di regressione lineare dinamico.	40
2.6 Test di corretta specificazione del modello e criteri informativi per la scelta dei ritardi.	40
2.7 2) Cenni sui modelli ARMA(p,q) e l'approccio Box -Jenkins.	40
Il modello a media mobile	40
Il modello MA(1)	41
Stimatori per medie mobili generiche.....	42
Il modello MA(∞)	42
Il processo a media mobile con notazione LAG	43
Il teorema di Wold (Notazione ridotta)	43
Il modello Autoregressivo (AR)	43
Il modello AR(1).....	44
Il passaggio da AR(1) a MA(∞)	45
Stima dei parametri per Modelli ARMA	46
Stime nel modello AR	46
Stime nel modello MA	47
Selezione dell'ordine del modello ARMA con criteri informativi.....	48
Stazionarietà ed Invertibilità per processi ARMA	48
L'approccio di Box e Jenkins per la selezione del modello	49
Le previsioni sfruttando il modello	49
Previsioni con il modello autoregressivo	50
Il Q-test di Ljung e Box per la correlazione dei residui	51
Test dei moltiplicatori di Lagrange.....	51
2.8 Cenni sui processi non stazionari: il modello di passeggiata aleatoria (Random Walk).	51
2.9 Esercitazioni al computer	52
Capitolo 3 Modelli econometrici multivariati per serie storiche economiche stazionarie.....	53
3.1 Il modello di regressione lineare dinamico multivariato.....	53

3.2 Il modello autoregressivo vettoriale (VAR): specificazione, stima, test di corretta specificazione del modello ed inferenza sui parametri;	53
Modelli Autoregressivi Vettoriali (VAR)	53
3.3 Test di causalità di Granger nei modelli VAR stazionari	53
3.4 Le previsioni multiperiodali	54
3.5 Funzione di risposta all'impulso (innovation accounting)	54
3.6 Cenni sui VAR strutturali.	54
3.7 Esercitazioni al computer.	54
Capitolo 4 Modelli per dati panel	55
4.1 Vantaggi e potenziali problemi nell'uso di dati.....	55
4.2 Modelli lineari statici per dati panel: effetti	55
4.3 Lo stimatore Fixed effects (within group) e lo stimatore random effects	55
4.4 Regressori strettamente esogeni, debolmente esogeni ed endogeni; richiami sulla regressione con variabili strumentali.	55
4.5 Il metodo dei momenti ed il metodo dei momenti generalizzato (GMM)	55
4.6 Proprietà asintotiche dello stimatore GMM	55
4.7 Test sulle restrizioni di sovraidentificazione;	55
4.8 Applicazione ai modelli dinamici per dati panel;	56
4.9 Trasformazione alle differenze prime del modello e stimatore GMM di Arellano e Bond....	56
4.10 Esercitazioni al computer.	56

Richiami

Concetto chiave 3.1: stimatori e stime

Uno **stimatore** è una funzione di un campione di dati estratti casualmente da una popolazione. La **stima** è il valore numerico dello stimatore, quando questo viene calcolato usando i dati di uno specifico campione. Uno stimatore è una variabile casuale per effetto della casualità dovuta alla selezione del campione, mentre la stima è un numero.

Concetto chiave 3.2: distorsione, consistenza ed efficienza

Sia $\hat{\mu}_Y$ uno stimatore di μ_Y . Allora:

- $E(\hat{\mu}_Y) - \mu_Y$ misura la **distorzione** di $\hat{\mu}_Y$;
- $\hat{\mu}_Y$ è uno **stimatore non distorto o corretto** di μ_Y , se $E(\hat{\mu}_Y) = \mu_Y$;
- $\hat{\mu}_Y$ è uno **stimatore consistente** di μ_Y , se $\hat{\mu}_Y \xrightarrow{P} \mu_Y$;
- sia $\tilde{\mu}_Y$ un altro stimatore di μ_Y , anch'esso non distorto. Allora, $\hat{\mu}_Y$ è più **efficiente** di $\tilde{\mu}_Y$, se $\text{var}(\hat{\mu}_Y) < \text{var}(\tilde{\mu}_Y)$.

Concetto chiave 3.3: efficienza di \bar{Y}

Sia $\hat{\mu}_Y$ uno stimatore di μ_Y ottenuto come media ponderata di Y_1, \dots, Y_n , ovvero $\hat{\mu}_Y = \frac{1}{n} \sum_{i=1}^n a_i Y_i$, dove a_1, \dots, a_n sono costanti deterministiche. Se $\hat{\mu}_Y$ è non distorto, allora $\text{var}(\bar{Y}) < \text{var}(\hat{Y})$, a meno che $\hat{\mu}_Y = \bar{Y}$. Cioè, \bar{Y} è lo stimatore più efficiente di μ_Y tra tutti gli stimatori non distorti ottenuti come medie ponderate di Y_1, \dots, Y_n .

Concetto chiave 3.4: errore standard di \bar{Y}

L'**errore standard** di \bar{Y} è uno stimatore della deviazione standard di \bar{Y} . L'errore standard di \bar{Y} viene indicato con $\text{SE}(\bar{Y})$ o con $\hat{\sigma}_{\bar{Y}}$. Quando Y_1, \dots, Y_n sono i.i.d.,

$$\text{SE}(\bar{Y}) = \hat{\sigma}_{\bar{Y}} = s_Y / \sqrt{n}. \quad (3.14)$$

Concetto chiave 3.5: la terminologia della verifica di ipotesi

La probabilità prefissata di rifiutare l'ipotesi nulla quando questa è vera costituisce il **livello di significatività** del test. Il **valore critico** della statistica test è il valore per il quale il test passa dal non rifiuto al rifiuto dato un certo livello di significatività. L'insieme dei valori della statistica test per i quali il test rifiuta l'ipotesi nulla è detto **regione di rifiuto** e l'insieme dei valori per i quali il test non rifiuta l'ipotesi nulla è detto **regione di accettazione**. La probabilità che il test porti al rifiuto dell'ipotesi nulla quando questa è vera è detta **livello minimo** del test, e la probabilità che il test rifiuti correttamente l'ipotesi nulla quando è vera l'alternativa è detta **potenza del test**.

Il valore- p è la probabilità, nel caso di campionamento ripetuto, di ottenere una statistica test sfavorevole all'ipotesi nulla almeno quanto la statistica effettivamente osservata, assumendo che valga l'ipotesi nulla. Equivalentemente, il valore- p è il livello di significatività più basso per il quale si può rifiutare l'ipotesi nulla dato il valore osservato della statistica test.

Concetto chiave 3.6: verifica dell'ipotesi $E(Y) = \mu_{Y,0}$ contro l'alternativa $E(Y) \neq \mu_{Y,0}$

1. Si calcola l'errore standard di \bar{Y} , $\text{SE}(\bar{Y})$ (formula (3.14)).
2. Si calcola la statistica t (formula (3.10)).
3. Si calcola il valore- p (formula (3.13)). Si rifiuta l'ipotesi al livello di significatività del 5%, se il valore- p è minore di 0,05 (o, equivalentemente, se $|t^{act}| > 1,96$).

Eteroschedasticità: Gli econometri applicati utilizzano abitualmente errori standard robusti all'eteroschedasticità in modo da evitare di preoccuparsi se l'eteroschedasticità sia presente o meno. In questo libro andiamo oltre l'eteroschedasticità come eccezione o problema da affrontare; permettiamo, invece, la presenza di l'eteroschedasticità dall'inizio e semplicemente utilizziamo gli errori robusti all' eteroschedasticità presentiamo l'omoschedasticità come un caso speciale che fornisce una motivazione teorica per gli OLS.

Dato il modello lineare $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i \in \{1, \dots, n\}$

1.1 Variabile esogena

La variabile \underline{X} si dice esogena (nell'ambito di un modello lineare) se vale la relazione

$$\mathbf{E}(X_i \cdot \varepsilon_i) \stackrel{\Delta}{=} \text{COV}(X_i, \varepsilon_i) = 0, \forall i$$

Ovvero la variabile non è correlata con l'errore.

1.2 Variabile endogena

La variabile \underline{X} si dice endogena se vale la relazione

$$\exists \bar{k} \in \{1, \dots, n\} : \mathbf{E}(X_{\bar{k}} \cdot \varepsilon_{\bar{k}}) \neq 0$$

ovvero

$$\text{COV}(\underline{X}, \underline{\varepsilon}) \neq [0]$$

Il caso delle variabili endogene è importante nei modelli lineari perché va contro l'assunzione che le variabili non devono essere correlate con l'errore, causando così stimatori dei minimi quadrati distorti ed inconsistenti.

Concetto chiave 4.1: terminologia per il modello di regressione lineare con un singolo regressore

Il modello di regressione lineare è:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

dove:

il pedice i varia tra le osservazioni, $i = 1, \dots, n$;

Y_i è la *variabile dipendente* o semplicemente la variabile *di sinistra*;

X_i è la *variabile indipendente*, il *regressore* o semplicemente la variabile *di destra*;

$\beta_0 + \beta_1 X$ è la *retta di regressione della popolazione* o *funzione di regressione della popolazione*;

β_0 è l'*intercetta* della retta di regressione della popolazione;

β_1 è la *pendenza* della retta di regressione della popolazione; e

u_i è l'*errore*.

Concetto chiave 4.3: le assunzioni dei minimi quadrati

$Y_i = \beta_0 + \beta_1 X_i + u_i$, con $i = 1, \dots, n$, dove:

1. l'errore u_i ha media condizionata nulla data X_i , ovvero $E(u_i | X_i) = 0$;
2. (X_i, Y_i) , $i = 1, \dots, n$, sono estratti indipendentemente e identicamente distribuiti (i.i.d.) dalla loro distribuzione congiunta;
3. (X_i, u_i) hanno momenti quarti finiti non nulli.

Concetto chiave 4.4: le distribuzioni di $\hat{\beta}_0$ e $\hat{\beta}_1$ in grandi campioni

Se valgono le ipotesi dei minimi quadrati nel concetto chiave 4.3, $\hat{\beta}_0$ e $\hat{\beta}_1$ hanno congiuntamente distribuzione campionaria normale. La distribuzione normale in grandi campioni di $\hat{\beta}_1$ è $N(\beta_1, \sigma_{\hat{\beta}_1}^2)$, dove la varianza di questa distribuzione, $\sigma_{\hat{\beta}_1}^2$, è

$$\sigma_{\hat{\beta}_1}^2 = \frac{1}{n} \frac{\text{var}[(X_i - \mu_X)u_i]}{\text{var}(X_i)^2}. \quad (4.14)$$

La distribuzione normale in grandi campioni di $\hat{\beta}_0$ è $N(\beta_0, \sigma_{\hat{\beta}_0}^2)$, dove

$$\sigma_{\hat{\beta}_0}^2 = \frac{1}{n} \frac{\text{var}(H_i u_i)}{[\text{E}(H_i^2)]^2}, \text{ dove } H_i = 1 - \left(\frac{\mu_X}{\text{E}(X_i^2)} \right) X_i. \quad (4.15)$$

Concetto chiave 4.5: forma generale della statistica t

In generale, la statistica t ha la forma

$$t = \frac{\text{stimatore} - \text{valore ipotizzato}}{\text{errore standard dello stimatore}}. \quad (4.18)$$

Concetto chiave 4.6: verifica dell'ipotesi $\beta_1 = \beta_{1,0}$ contro l'alternativa $\beta_1 \neq \beta_{1,0}$

1. Si calcoli l'errore standard di $\hat{\beta}_1$, $\text{SE}(\hat{\beta}_1)$ (formula (4.17)).

2. Si calcoli la statistica t (formula (4.20)).

3. Si calcoli il valore- p (formula (4.22)). Si rifiuti l'ipotesi al livello di significatività del 5% se il valore- p è inferiore a 0,05 o, equivalentemente, se $|t^{\text{act}}| > 1,96$.

L'errore standard e (tipicamente) la statistica t e il valore- p per verificare l'ipotesi $\beta_1 = 0$ sono calcolati automaticamente dai pacchetti statistici.

Concetto chiave 4.7: intervalli di confidenza per β_1

Un intervallo di confidenza bilaterale del 95% per β_1 è un intervallo che contiene il vero valore di β_1 con probabilità 95%, ovvero contiene il vero valore di β_1 nel 95% di tutti i possibili campioni estratti casualmente. Equivalentemente, esso è anche l'insieme dei valori di β_1 che non possono essere rifiutati al 5% da un test d'ipotesi bilaterale. Quando la dimensione campionaria è elevata, l'intervalllo di confidenza di livello 95% per β_1 si costruisce come segue

$$(\hat{\beta}_1 - 1,96 \text{ SE}(\hat{\beta}_1), \hat{\beta}_1 + 1,96 \text{ SE}(\hat{\beta}_1)). \quad (4.28)$$

Concetto chiave 4.8: eteroschedasticità e omoschedasticità

L'errore di regressione u_i è omoschedastico se la varianza della distribuzione condizionata di u_i data X_i , $\text{var}(u_i | X_i = x)$, è costante per $i = 1, \dots, n$ e in particolare non dipende da x ; altrimenti, l'errore è eteroschedastico.

Capitolo 1. Richiami sulla funzione di regressione e sue proprietà

Concetto chiave 4.1: terminologia per il modello di regressione lineare con un singolo regressore

Il modello di regressione lineare è:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

dove:

il pedice i varia tra le osservazioni, $i = 1, \dots, n$;

Y_i è la *variabile dipendente* o semplicemente la variabile *di sinistra*;

X_i è la *variabile indipendente*, il *regressore* o semplicemente la variabile *di destra*;

$\beta_0 + \beta_1 X$ è la *retta di regressione della popolazione* o *funzione di regressione della popolazione*;

β_0 è l'*intercetta* della retta di regressione della popolazione;

β_1 è la *pendenza* della retta di regressione della popolazione; e

u_i è l'*errore*.

Concetto chiave 4.3: le assunzioni dei minimi quadrati

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \text{ con } i = 1, \dots, n, \text{ dove:}$$

1. l'errore u_i ha media condizionata nulla data X_i , ovvero $E(u_i | X_i) = 0$;
2. (X_i, Y_i) , $i = 1, \dots, n$, sono estratti indipendentemente e identicamente distribuiti (i.i.d.) dalla loro distribuzione congiunta;
3. (X_i, u_i) hanno momenti quarti finiti non nulli.

Il modello di regressione lineare multiplo

Concetto chiave 5.2: il modello di regressione multipla

Il modello di regressione multipla è

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n, \quad (5.7)$$

dove:

- Y_i è la i -esima osservazione della variabile dipendente, $X_{1i}, X_{2i}, \dots, X_{ki}$ sono le i -esime osservazioni di ciascuno dei k regressori e u_i è l'errore;
- la retta di regressione della popolazione è la relazione tra la Y e le X che vale in media nella popolazione:

$$\begin{aligned} E(Y | X_{1i} = x_1, X_{2i} = x_2, \dots, X_{ki} = x_k) \\ = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}; \end{aligned}$$

- β_1 è il coefficiente angolare di X_1 , β_2 è il coefficiente angolare di X_2 ecc. Il coefficiente β_1 rappresenta la variazione attesa di Y_i che deriva da una variazione unitaria in X_{1i} , tenendo costanti X_{2i}, \dots, X_{ki} . I coefficienti delle altre X si interpretano in maniera simile;
- l'intercetta β_0 è il valore atteso di Y , quando tutte le X sono pari a zero. L'intercetta può essere pensata come il coefficiente di un regressore, X_{0i} , che è uguale a uno per ogni i .

Concetto chiave 5.4: le assunzioni dei minimi quadrati relative al modello di regressione multipla

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$, con $i = 1, \dots, n$, dove:

1. u_i ha media condizionata nulla, date $X_{1i}, X_{2i}, \dots, X_{ki}$, ovvero $E(u_i | X_{1i}, X_{2i}, \dots, X_{ki}) = 0$;
2. $(X_{1i}, \dots, X_{ki}, Y_i)$, con $i = 1, \dots, n$, sono estratti indipendentemente e indenticamente distribuiti (i.i.d.) dalla propria distribuzione congiunta;
3. $(X_{1i}, \dots, X_{ki}, u_i)$ hanno momenti quarti finiti e non nulli;
4. non vi è collinearità perfetta.

.1 Il modello di regressione lineare multiplo con uso della notazione matriciale

Il modello di regressione multipla in notazione matriciale

Il modello di regressione multipla (concetto chiave 5.2) è

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \dots, n. \quad (16.1)$$

Per scrivere il modello in notazione matriciale, si devono definire i seguenti vettori e matrici:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \\ \vdots \\ \mathbf{X}'_n \end{pmatrix},$$

$$\text{e } \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad (16.2)$$

dove \mathbf{Y} è $n \times 1$, \mathbf{X} è $n \times (k+1)$, \mathbf{U} è $n \times 1$ e $\boldsymbol{\beta}$ è $(k+1) \times 1$. D'ora in poi matrici e vettori saranno riportati in grassetto. Secondo la seguente notazione,

- \mathbf{Y} è un vettore $n \times 1$ di n osservazioni della variabile dipendente;
- \mathbf{X} è una matrice $n \times (k+1)$ di n osservazioni dei $k+1$ regressori (incluso il “regressore” costante per l’intercetta);
- il vettore colonna \mathbf{X}_i di dimensioni $(k+1) \times 1$ è la i -esima osservazione sui $k+1$ regressori, cioè, $\mathbf{X}'_i = (1 \ X_{1i} \ \cdots \ X_{ki})$, dove \mathbf{X}'_i rappresenta la trasposta di \mathbf{X}_i ;
- \mathbf{U} è un vettore $n \times 1$ degli n termini d’errore;
- $\boldsymbol{\beta}$ è un vettore $(k+1) \times 1$ di coefficienti ignoti di regressione;

Il modello di regressione multipla (16.1) per la i -esima osservazione, scritto tramite i vettori $\boldsymbol{\beta}$ e \mathbf{X}_i , è

$$Y_i = \mathbf{X}'_i \boldsymbol{\beta} + u_i, \quad i = 1, \dots, n, \quad (16.3)$$

dove il primo regressore è il regressore “costante”, sempre pari ad 1, e il suo coefficiente è l’intercetta. Perciò l’intercetta non appare separatamente nella (16.3), ma è il primo elemento del vettore dei coefficienti $\boldsymbol{\beta}$.

Raggruppando tutte le n osservazioni della (16.3) si ottiene il modello di regressione multipla in forma matriciale:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}. \quad (16.4)$$

Concetto chiave 16.1: le assunzioni generalizzate dei minimi quadrati nel modello di regressione multipla

Il modello lineare di regressione con regressori multipli è

$$Y_i = \mathbf{X}'_i \boldsymbol{\beta} + u_i, \quad i = 1, \dots, n. \quad (16.5)$$

Le assunzioni estese dei minimi quadrati sono

1. $E(u_i | \mathbf{X}_i) = 0$ (u_i ha media condizionata nulla);
2. $(\mathbf{X}_i, Y_i), i = 1, \dots, n$ sono estrazioni indipendentemente e identicamente distribuite (i.i.d.) dalla loro distribuzione congiunta;
3. \mathbf{X}_i e u_i hanno quattro momenti finiti non nulli;
4. \mathbf{X} ha rango di colonna pieno (c'è assenza di perfetta collinearità);
5. $\text{var}(u_i | \mathbf{X}_i) = \sigma_u^2$ (omoschedasticità);
6. la distribuzione condizionata di u_i dato \mathbf{X}_i è normale (errori normali).

.2 Richiami sulle proprietà dello stimatore OLS dei parametri del modello di regressione lineare multipla

Concetto chiave 5.3: gli stimatori OLS, i valori previsti e i residui del modello di regressione multipla

Gli stimatori OLS $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ sono quei valori di b_0, b_1, \dots, b_k che minimizzano la somma dei quadrati degli errori di previsione $\sum_{i=1}^n (Y_i - b_0 - b_1 X_{1i} - \dots - b_k X_{ki})^2$. I valori predetti \hat{Y}_i e i residui \hat{u}_i degli OLS sono:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki}, \quad \text{con } i = 1, \dots, n, \quad \text{e} \quad (5.11)$$

$$\hat{u}_i = Y_i - \hat{Y}_i, \quad \text{con } i = 1, \dots, n. \quad (5.12)$$

Gli stimatori OLS $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ e il residuo \hat{u}_i sono calcolati per un campione di n osservazioni $(X_{1i}, \dots, X_{ki}, Y_i)$, con $i = 1, \dots, n$. Essi sono stimatori dei veri coefficienti ignoti della popolazione $\beta_0, \beta_1, \dots, \beta_k$ e dell'errore u_i .

Concetto chiave 5.5: la distribuzione di $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ in grandi campioni

Se valgono le assunzioni dei minimi quadrati (concetto chiave 5.4), gli stimatori OLS $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ sono, in grandi campioni, congiuntamente distribuiti secondo una normale e ogni $\hat{\beta}_j$ si distribuisce secondo una $N(\beta_j, \sigma_{\hat{\beta}_j}^2)$, con $j = 0, \dots, k$.

Concetto chiave 5.6: la verifica dell'ipotesi $\beta_j = \beta_{j,0}$ contro l'alternativa $\beta_j \neq \beta_{j,0}$

1. Si calcoli l'errore standardizzato di $\hat{\beta}_j$, $SE(\hat{\beta}_j)$.
2. Si calcoli la statistica t ,

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)}. \quad (5.14)$$

3. Si calcoli il valore- p ,

$$\text{valore-}p = 2\Phi(-|t^{act}|), \quad (5.15)$$

dove t^{act} è il valore effettivamente calcolato della statistica t . Si rifiuti l'ipotesi al livello di significatività 5%, se il valore- p è minore di 0,05 oppure, equivalentemente, se $|t^{act}| > 1,96$.

L'errore standard e (tipicamente) la statistica t e il valore- p per l'ipotesi nulla che $\beta_j = 0$ sono calcolati automaticamente dai software di regressione.

Concetto chiave 5.7: gli intervalli di confidenza per un singolo regressore nel modello di regressione multipla

Un intervallo di confidenza bilaterale di livello 95% per il coefficiente β_j è un intervallo che contiene il valore vero di β_j con probabilità 95%; in altre parole, esso contiene il vero valore di β_j nel 95% di tutti i campioni che è possibile estrarre. Equivalentemente, è anche l'insieme di valori di β_j che non possono essere rifiutati da un test d'ipotesi bilaterale al 5%. Quando il campione è grande, l'intervalllo di confidenza di livello 95% è:

$$\beta_j = (\hat{\beta}_j - 1,96 SE(\hat{\beta}_j), \hat{\beta}_j + 1,96 SE(\hat{\beta}_j)). \quad (5.17)$$

Un intervallo di confidenza di livello 90% si ottiene sostituendo 1,96 con 1,645 nella 5.17.

Concetto chiave 16.1: le assunzioni generalizzate dei minimi quadrati
nel modello di regressione multipla

Il modello lineare di regressione con regressori multipli è

$$Y_i = \mathbf{X}'_i \boldsymbol{\beta} + u_i, \quad i = 1, \dots, n. \quad (16.5)$$

Le assunzioni estese dei minimi quadrati sono

1. $E(u_i | \mathbf{X}_i) = 0$ (u_i ha media condizionata nulla);
2. $(\mathbf{X}_i, Y_i), i = 1, \dots, n$ sono estrazioni indipendentemente e identicamente distribuite (i.i.d.) dalla loro distribuzione congiunta;
3. \mathbf{X}_i e u_i hanno quattro momenti finiti non nulli;
4. \mathbf{X} ha rango di colonna pieno (c'è assenza di perfetta collinearità);
5. $\text{var}(u_i | \mathbf{X}_i) = \sigma_u^2$ (omoschedasticità);
6. la distribuzione condizionata di u_i dato \mathbf{X}_i è normale (errori normali).

Risolvendo il sistema di equazioni (16.10) si ottiene lo stimatore OLS $\hat{\boldsymbol{\beta}}$ in forma matriciale:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}, \quad (16.11)$$

dove $(\mathbf{X}' \mathbf{X})^{-1}$ è la matrice inversa di $(\mathbf{X}' \mathbf{X})$.

Concetto chiave 16.2: il teorema limite centrale multivariato

Si supponga che $\mathbf{W}_1, \dots, \mathbf{W}_n$ siano variabili casuali m dimensionali con vettore delle medie $E(\mathbf{W}_i) = \boldsymbol{\mu}_{\mathbf{W}}$ e matrice di covarianza $E[(\mathbf{W}_i - \boldsymbol{\mu}_{\mathbf{W}})(\mathbf{W}_i - \boldsymbol{\mu}_{\mathbf{W}})'] = \boldsymbol{\Sigma}_{\mathbf{W}}$, dove $\boldsymbol{\Sigma}_{\mathbf{W}}$ è definita positiva e finita. Sia $\bar{\mathbf{W}} = \frac{1}{n} \sum_{i=1}^n \mathbf{W}_i$, allora $\sqrt{n}(\bar{\mathbf{W}} - \boldsymbol{\mu}_{\mathbf{W}}) \xrightarrow{d} N(\mathbf{0}_m, \boldsymbol{\Sigma}_{\mathbf{W}})$.

Normalità asintotica dei Beta

Normalità asintotica di $\hat{\beta}$

Per grandi campioni, lo stimatore OLS ha distribuzione asintotica normale multivariata

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}_{k+1}, \Sigma_{\sqrt{n}(\hat{\beta} - \beta)}), \text{ dove } \Sigma_{\sqrt{n}(\hat{\beta} - \beta)} = \mathbf{Q}_X^{-1} \Sigma_V \mathbf{Q}_X^{-1}, \quad (16.12)$$

dove \mathbf{Q}_X è la matrice $(k+1) \times (k+1)$ dei momenti secondi dei regressori, cioè, $\mathbf{Q}_X = E(\mathbf{X}_i \mathbf{X}_i')$ e Σ_V è la matrice $(k+1) \times (k+1)$ di covarianza di $\mathbf{V}_i = \mathbf{X}_i u_i$, cioè $\Sigma_V = E(\mathbf{V}_i \mathbf{V}_i')$. Si noti che la seconda assunzione dei minimi quadrati del concetto chiave 16.1 implica che $\mathbf{V}_i, i = 1, \dots, n$, siano i.i.d.

Scritta in termini di $\hat{\beta}$ piuttosto che di $\sqrt{n}(\hat{\beta} - \beta)$, l'approssimazione normale nella (16.12) è

$$\begin{aligned} \text{per grandi campioni } \hat{\beta} \text{ si distribuisce come } N(\beta, \Sigma_{\hat{\beta}}), \\ \text{dove } \Sigma_{\hat{\beta}} = \Sigma_{\sqrt{n}(\hat{\beta} - \beta)}/n = \mathbf{Q}_X^{-1} \Sigma_V \mathbf{Q}_X^{-1} / n. \end{aligned} \quad (16.13)$$

La matrice di covarianza $\Sigma_{\hat{\beta}}$ nella (16.13) è la matrice di covarianza dell'approssimazione normale alla distribuzione di $\hat{\beta}$, mentre $\Sigma_{\sqrt{n}(\hat{\beta} - \beta)}$ nell'equazione (16.12) è la matrice di covarianza della distribuzione asintotica normale di $\sqrt{n}(\hat{\beta} - \beta)$. Queste due matrici di covarianza differiscono per un fattore pari a n , a seconda che gli OLS siano riscalati per \sqrt{n} oppure no.

Derivazione dell'equazione (16.12). Per derivare la (16.12), si utilizzano anzitutto la (16.4) e la (16.11) per scrivere $\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X}\beta + \mathbf{U})$, in modo che

$$\hat{\beta} = \beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{U}. \quad (16.14)$$

Perciò $\hat{\beta} - \beta = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{U}$, e quindi

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}' \mathbf{U}}{\sqrt{n}} \right) \quad (16.15)$$

Per derivare l'equazione (16.12) è necessario prima mostrare che la matrice "denominatore", $\mathbf{X}' \mathbf{X}/n$, è consistente e poi che la matrice "numeratore", $\mathbf{X}' \mathbf{U}/\sqrt{n}$, obbedisce al teorema limite centrale multivariato del concetto chiave 16.2. I dettagli sono riportati nell'appendice 16.3.

Funzione di regressione

Considerati due numeri aleatori (n.a.) Y e X , è detta "funzione di regressione di Y rispetto X " il n.a. $E(Y/X)$ dipendente da X secondo una funzione $g(.)$ [cioè $E(Y/X) = g(X)$] determinata dalla distribuzione subordinata di Y rispetto X .

I valori del n.a. $g(X)$ [ossia i valori del n.a. $E(Y/X)$] sono le speranze matematiche condizionate $E(Y/X = x)$. Pertanto l'evento $\{g(X) = E(Y/X = x)\}$ coincide con l'evento "Il n.a. $E(Y/X)$ assume il valore $E(Y/X = x)$ " ed esso ha probabilità pari a $\text{Prob}(X = x)$ quando il n.a. X è discreto; altrimenti ad esso è associata la densità di probabilità marginale del n.a. X nel punto x , cioè $f(x)$.

La conoscenza della funzione di regressione $E(Y/X)$ permette quindi di conoscere per ogni possibile valore x assunto dal n.a. X quale sarà il valore medio (condizionato) di Y . Inoltre al variare di x , saremo in grado di sapere come varia in media (condizionata) Y .

Alcune proprietà della funzione di regressione sono le seguenti:

1. $E(aY + bZ / X) = aE(Y/X) + bE(Z/X)$, se a e b sono numeri certi (ossia delle costanti) e X , Y e Z sono numeri aleatori (proprietà di linearità della funzione di regressione);
2. $E[E(Y/X)] = E(Y)$ (legge del valore atteso iterato);
3. $E[Y h(X) / X] = h(X) E(Y/X)$: come caso particolare della 3) vale
- 3.1 $E[E(Y/X) / X] = E[g(X) / X] = g(X) = E(Y/X)$, essendo $E(Y/X)$ una funzione $g(.)$ di X .
4. se i numeri aleatori X e Y sono stocasticamente indipendenti, $E(Y/X) = E(Y)$.
5. $E[Y - E(Y/X)]^2 = E[Y - (X)]^2$ per ogni funzione reale $(.)$ tale che $E[(X)]^2$.

Una definizione assiomatica di funzione di regressione $E(Y/X)$ è la seguente: essa è quel numero aleatorio dipendente da X che verifica la condizione per ogni funzione per la quale esista la speranza matematica a primo membro.

La nozione di funzione di regressione è fondamentale nella teoria e calcolo delle probabilità; è il caso di osservare che in lingua inglese $E(Y/X)$ è detta "conditional expectation", ma con lo stesso nome si indicano anche i suoi possibili valori $E(Y/X=x)$ e ciò può ingenerare fraintendimenti. In questi appunti indicheremo con il nome di funzione di regressione il numero aleatorio $E(Y/X)$ e con quello di valor medio condizionato le possibili determinazioni $E(Y/X=x)$ di $E(Y/X)$ che sono numeri certi.

E' immediato generalizzare la definizione di funzione di regressione di Y rispetto X al caso in cui ci si condiziona rispetto a più di un n.a. Ad esempio, la funzione di regressione di Y rispetto X e Z è indicata con $E(Y/X,Z)$ ed è una funzione $g(X,Z)$. Quindi $E(Y/X, Z)$ è un n.a. i cui valori sono le speranze matematiche condizionate $E(Y/X = x, Z=z)$. L'evento "Il n.a. $E(Y/X, Z)$ assume il valore $E(Y/X = x, Z=z)$ " ha probabilità pari alla probabilità congiunta $\text{Prob}(X = x, Z=z)$ quando entrambi i n.a. X e Z sono discreti; se X e Z sono entrambi (assolutamente) continui ad esso è associata la densità di probabilità congiunta dei n.a. X e Z nel punto (x,z) , cioè $f(x,z)$.

Le proprietà di $E(Y/X)$ sopra esposte sono immediatamente generalizzabili. In particolare, se X e Y sono stocasticamente indipendenti risulta $E(Y/X, Z) = E(Y/Z)$. Un'ultima importante proprietà è la seguente generalizzazione della legge del valore atteso iterato

6. $E[E(Y/X) / X, Z] = E[E(Y / X, Z) / X] = E(Y/X) :$

Correttezza dei Beta OLS, consistenza dei Beta OLS.

Operatore PLIM

Si riferisce a "probability limit", cioè alla convergenza in probabilità $X_n \xrightarrow{p} X$ cioè

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1.$$

Valgono le proprietà:

$$p_lim(XY) = p_lim(X) \times p_lim(Y)$$

$$p_lim(X \pm Y) = p_lim(X) \pm p_lim(Y)$$

Distribuzione Normale bivariata e multivariata

coefficient to be

$$R = \frac{(\sigma_{12}\Sigma_{22}^{-1}\sigma_{21})^{\frac{1}{2}}}{\sigma_{11}^{\frac{1}{2}}} \geq \text{Corr}(X_1, \alpha'X_2), \quad 0 \leq R \leq 1. \quad (15.8)$$

In the case where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \mathbf{X}_1: k \times 1, \quad \mathbf{X}: (n-k) \times 1, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad k \geq 1,$$

we could define the r.v.'s $Z_1 = \alpha'_1 \mathbf{X}_1$ and $Z_2 = \alpha'_2 \mathbf{X}_2$ whose correlation coefficient is

$$\text{Corr}(Z_1, Z_2) = \frac{\alpha'_1 \Sigma_{12} \alpha_2}{(\alpha'_1 \Sigma_{11} \alpha_1)^{\frac{1}{2}} (\alpha'_2 \Sigma_{22} \alpha_2)^{\frac{1}{2}}}. \quad (15.9)$$

From the above inequality it follows that for $\alpha_2 = \Sigma_{22}^{-1} \Sigma_{21} \alpha_1$

$$C_{12} = \frac{(\alpha'_1 \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha_1)^{\frac{1}{2}}}{(\alpha'_1 \Sigma_{11} \alpha_1)^{\frac{1}{2}}} \geq \text{Corr}(Z_1, Z_2). \quad (15.10)$$

$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ has at most k non-zero eigenvalues which measure the association between \mathbf{X}_1 and \mathbf{X}_2 and are called *canonical correlations*.

Let

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \quad \text{where } \mathbf{X}_3: (n-2) \times 1$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \Sigma_{33} \end{pmatrix}.$$

Another form of correlation of interest in this context is the correlation between X_1 and X_2 given that the effect of \mathbf{X}_3 is taken away. For this we form the r.v.'s

$$Y_1 = X_1 - \mathbf{b}'_1 \mathbf{X}_3, \quad \text{and} \quad Y_2 = X_2 - \mathbf{b}'_2 \mathbf{X}_3, \quad \text{and} \quad \text{Corr}(Y_1, Y_2)$$

is maximised by $\mathbf{b}_1 = \Sigma_{33}^{-1} \sigma_{31}$ and $\mathbf{b}_2 = \Sigma_{33}^{-1} \sigma_{32}$ as seen above. Hence we define the *partial correlation coefficient* between X_1 and X_2 given \mathbf{X}_3 to be

$$\rho_{12.3} = \frac{\sigma_{12} - \sigma_{13} \Sigma_{33}^{-1} \sigma_{32}}{[\sigma_{11} - \sigma_{13} \Sigma_{33}^{-1} \sigma_{31}]^{\frac{1}{2}} [\sigma_{22} - \sigma_{23} \Sigma_{33}^{-1} \sigma_{32}]^{\frac{1}{2}}} \geq \text{Corr}(Y_1, Y_2). \quad (15.11)$$

†

15.2 The multivariate normal distribution

The univariate normal density function discussed above was of the form

$$f(x; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}. \quad (15.12)$$

The density function of $\mathbf{X}=(X_1, X_2, \dots, X_n)'$ when the X_i 's are IID normally distributed r.v.'s was shown to be of the form

$$\begin{aligned} f(\mathbf{x}; \mu, \sigma^2) &= \prod_{i=1}^n f(x_i; \mu, \sigma^2) \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}. \end{aligned} \quad (15.13)$$

Similarly, the density function of \mathbf{X} when the X_i 's are only independent, i.e. $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$, takes the form

$$\begin{aligned} f(\mathbf{x}; \mu_1, \dots, \mu_n, \sigma_1^2, \dots, \sigma_n^2) &= \prod_{i=1}^n f(x_i; \mu_i, \sigma_i^2) \\ &= (2\pi)^{-n/2} (\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right\}. \end{aligned} \quad (15.14)$$

Comparing the above three density functions we can discern a pattern developing which is very suggestive for the density function of an arbitrary normal vector \mathbf{X} with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, which takes the form

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} (\det \boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right\}, \quad (15.15)$$

and we write $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If the X_i 's are IID r.v.'s $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n$ and $(\det \boldsymbol{\Sigma}) = (\sigma^2)^n$. On the other hand, if the X_i 's are independent but not identically distributed

$$\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \quad \text{and} \quad (\det \boldsymbol{\Sigma}) = \prod_{i=1}^n (\sigma_i^2) = (\sigma_1^2 \sigma_2^2 \dots \sigma_n^2).$$

In the case of $n=2$

$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \text{where } \rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}, \\ \Rightarrow (\det \boldsymbol{\Sigma}) &= \sigma_1^2\sigma_2^2(1 - \rho^2) > 0 \quad \text{for } -1 < \rho < 1 \end{aligned}$$

and

$$\Sigma^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1 \sigma_2} \\ \frac{-\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}. \quad (15.16)$$

Thus the bivariate normal density function is

$$f(x_1, x_2; \mu, \Sigma) = \frac{[\sigma_1^2 \sigma_2^2 (1 - \rho^2)]^{-\frac{1}{2}}}{2\pi} \times \exp \left\{ -\frac{(1 - \rho^2)^{-1}}{2} \right. \\ \left. \times \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} \quad (15.17)$$

(see Chapter 6). The standard bivariate density function can be obtained by defining the new r.v.'s

$$Z_i = \left(\frac{X_i - \mu_i}{\sigma_i} \right), \quad i = 1, 2,$$

whose density function is

$$f(z_1, z_2; \rho) = \frac{(1 - \rho^2)^{-\frac{1}{2}}}{2\pi} \exp \left\{ -\frac{(1 - \rho^2)^{-1}}{2} (z_1^2 - 2\rho z_1 z_2 + z_2^2) \right\}. \quad (15.18)$$

(1) Properties

- (N1) Let $\mathbf{X} \sim N(\mu, \Sigma)$ then $\mathbf{Y} = (\mathbf{A}\mathbf{X} + \mathbf{b}) \sim N(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}')$ for $\mathbf{A}: m \times n$ and $\mathbf{b}: m \times 1$ constant matrices, e.g. if $\mathbf{Y} = c\mathbf{X}$, $c \neq 0$, $\mathbf{Y} \sim N(c\mu, c^2\Sigma)$.

This property shows that if \mathbf{X} is normally distributed then any linear function of \mathbf{X} is also normally distributed.

- (N2) Let $\mathbf{X}_t \sim N(\mu_t, \Sigma_t)$, $t = 1, 2, \dots, T$, be independently distributed random vectors, then for any arbitrary constant matrices \mathbf{A}_t , $t = 1, 2, \dots, T$,

$$\left(\sum_{t=1}^T \mathbf{A}_t \mathbf{X}_t \right) \sim N \left(\sum_{t=1}^T \mathbf{A}_t \mu_t, \sum_{t=1}^T (\mathbf{A}_t \Sigma_t \mathbf{A}_t') \right).$$

The converse also holds. If the \mathbf{X}_t s are IID then $\mu_t = \mu$, $\Sigma_t = \Sigma$, $t = 1, 2, \dots, T$, and

$$\left(\frac{1}{T} \sum_{t=1}^T \mathbf{X}_t \right) \sim N \left(\mu, \frac{1}{T} \Sigma \right).$$

- (N3) Let $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then the X_i s are independent if and only if $\sigma_{ij}=0$, $i \neq j$, $i, j = 1, 2, \dots, n$, i.e. $\boldsymbol{\Sigma} = \text{diag}(\sigma_{11}, \dots, \sigma_{nn})$. In general, zero covariance does not imply independence but in the case of normality the two are equivalent.
- (N4) If $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then the marginal distribution of any $k \times 1$ subset \mathbf{X}_1 where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

$\mathbf{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$. This follows from property N1 for $\mathbf{A} = (\mathbf{I}_k : \mathbf{0})$, $(k \times n)$, $\mathbf{b} = \mathbf{0}$. Similarly, $\mathbf{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

These can be verified directly using

$$f(\mathbf{x}_1; \boldsymbol{\theta}_1) = \int f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}_2 \quad \text{and} \quad f(\mathbf{x}_2; \boldsymbol{\theta}_2) = \int f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}_1,$$

although the manipulations involved are rather too cumbersome. Taking $k=1$, this property implies that each component of $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is also normally distributed; the converse, however, is not true.

- (N5) For the same partition of \mathbf{X} considered in N4 the conditional distribution of \mathbf{X}_1 given \mathbf{X}_2 takes the form

$$(\mathbf{X}_1/\mathbf{X}_2), \quad N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \quad \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

This follows from property N1 for

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_k & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \mathbf{0} \quad (15.19)$$

since

$$\mathbf{AX} = \begin{pmatrix} \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2 \\ \mathbf{X}_2 \end{pmatrix}, \quad \text{Cov}(\mathbf{AX}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \quad (15.20)$$

From this we can deduce that if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ then \mathbf{X}_1 and \mathbf{X}_2 are independent since $(\mathbf{X}_1/\mathbf{X}_2) \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$. Moreover, for any $\boldsymbol{\Sigma}_{12}$, $(\mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2)$ and \mathbf{X}_2 are independent given that their covariance is zero. Similarly, $(\mathbf{X}_2/\mathbf{X}_1) \sim N(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$. In the case $n=2$

$$(X_1/X_2) \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (X_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right). \quad (15.21)$$

These results can be verified using the formula

$$f(\mathbf{x}_1/\mathbf{x}_2; \boldsymbol{\phi}) = \frac{f(\mathbf{x}; \boldsymbol{\theta})}{f(\mathbf{x}_2; \boldsymbol{\theta}_2)}. \quad (15.22)$$

1.3 Inferenza robusta all'eteroschedasticità

Errori standard robusti rispetto all'eteroschedasticità

Errori standard robusti all'eteroschedasticità

Lo stimatore robusto all'eteroschedasticità di $\Sigma_{\sqrt{n}(\hat{\beta}-\beta)}$ è ottenuto sostituendo i momenti della popolazione che appaiono nella sua definizione (equazione (16.12)) con i momenti campionari. Quindi, lo stimatore robusto all'eteroschedasticità della matrice di covarianza di $\sqrt{n}(\hat{\beta}-\beta)$ è

$$\Sigma_{\sqrt{n}(\hat{\beta}-\beta)} = \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \hat{\Sigma}_{\hat{\mathbf{V}}} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1}, \text{ dove } \hat{\Sigma}_{\hat{\mathbf{V}}} = \frac{1}{n-k-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' \hat{u}_i^2. \quad (16.16)$$

Lo stimatore $\hat{\Sigma}_{\hat{\mathbf{V}}}$ incorpora la stessa correzione per i gradi di libertà che si ha nel SER (errore standard di regressione) per il modello di regressione multipla (sezione 5.10) quando si devono correggere potenziali distorsioni verso il basso causate dalla stima dei $k+1$ coefficienti di regressione.

La dimostrazione che $\hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)} \xrightarrow{P} \Sigma_{\sqrt{n}(\hat{\beta}-\beta)}$ è concettualmente simile alla dimostrazione, presentata nella sezione 15.3, della consistenza degli errori standard robusti all'eteroschedasticità per il modello con un singolo regressore.

Errori standard robusti all'eteroschedasticità. Lo stimatore robusto all'eteroschedasticità della matrice di covarianza di $\hat{\beta}$, $\hat{\Sigma}_{\hat{\beta}}$, è

$$\hat{\Sigma}_{\hat{\beta}} = n^{-1} \hat{\Sigma}_{\sqrt{n}(\hat{\beta}-\beta)}. \quad (16.17)$$

L'errore standard robusto all'eteroschedasticità per il j -esimo coefficiente di regressione è la radice quadrata del j -esimo elemento diagonale di $\hat{\Sigma}_{\hat{\beta}}$. Cioè, l'errore standard robusto all'eteroschedasticità del j -esimo coefficiente è

$$SE(\hat{\beta}_j) = \sqrt{(\hat{\Sigma}_{\hat{\beta}})_{jj}}, \quad (16.18)$$

dove $(\hat{\Sigma}_{\hat{\beta}})_{jj}$ è l'elemento (j, j) di $\hat{\Sigma}_{\hat{\beta}}$.

L'effetto atteso di questa variazione in \mathbf{X}_i è pari a $\mathbf{d}' \beta$ e lo stimatore di questo effetto è $\mathbf{d}' \hat{\beta}$. Poiché combinazioni lineari di variabili casuali normalmente distribuite sono anch'esse normalmente distribuite, $\sqrt{n}(\mathbf{d}' \hat{\beta} - \mathbf{d}' \beta) = \mathbf{d}' \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbf{d}' \Sigma_{\sqrt{n}(\hat{\beta}-\beta)} \mathbf{d})$. Perciò, l'errore standard di questo effetto predetto è $(\mathbf{d}' \hat{\Sigma}_{\hat{\beta}} \mathbf{d})^{1/2}$. Un intervallo di confidenza di livello 95% per questo effetto predetto è

$$\mathbf{d}' \hat{\beta} \pm 1,96 \sqrt{\mathbf{d}' \hat{\Sigma}_{\hat{\beta}} \mathbf{d}}. \quad (16.19)$$

1.4 Test di corretta specificazione del modello di regressione multipla

Scopo di questi test diagnostici è verificare che le ipotesi classiche del modello di regressione lineare siano plausibili. Questo viene fatto controllando efficienza e correttezza della stima OLS dei parametri del modello.

Tutti questi test si basano sulla analisi dei residui dei MQI (=min quad ord). In tali contesti il modello non vincolato prende il nome di equazione (o modello) ausiliaria.

LOGICA: I test diagnostici si interpretano come problemi di variabili omesse, ovvero non corretta specificazione del modello.

SPECIFICAZIONE:

$y = X\beta + \varepsilon \rightarrow$ Modello stimato (potenzialmente scorretto)

$y = X\beta + Wy + \eta \rightarrow$ Modello generale (ausiliario)

dove W sono i regressori omessi, e γ sono i coefficienti associati ai regressori omessi.

Se il modello ausiliario è quello giusto (cioè se $\gamma \neq 0$), allora lo stimatore dei MQO del modello stimato è distorto:

$$E(\hat{\beta}|X, W) = E[(X'X)^{-1}X'(X\beta^*W\gamma + \eta)|X, W] = \beta^* + (X'X)^{-1} + X'W\gamma + X'E(\eta|X, W) = \beta^* + (X'X)^{-1} + X'W\gamma \neq \hat{\beta}$$

STATISTICA TEST: per verificare la significatività dei regressori W si applica un test di significatività congiunta di tutti i parametri.

$$\hat{\varepsilon} = y - X\hat{\beta}^* + W\gamma + \eta - X\hat{\beta} = X(\beta^* - \hat{\beta}) + W\gamma + \eta = Xb + W\gamma + \eta, \text{ con } b = \beta^* - \hat{\beta}$$

Ricordiamo che

$y = X\beta + \varepsilon \rightarrow$ Modello stimato (potenzialmente scorretto) (Vincolato)

$y = X\beta + Wy + \eta \rightarrow$ Modello generale (ausiliario) (Non Vincolato)

$$I \text{ residui del modello non vincolato sono dati dall'equazione ausiliaria } F = \frac{(SQR_v - SQR_{nv})/h}{SQR_{nv}/(n - k - h)}$$

In pratica si sottopone a verifica l'ipotesi nulla: $H_0 : \gamma = 0$

APPLICABILITÀ DEI TEST: ci sono problemi se...

- se $h > (n - k)$ il test non è calcolabile
- se le variabili W spesso (e in pratica) non sono note a priori
- se le variabili W sono considerate delle proxy, cioè variabili pensate solo allo scopo di cogliere possibili errori di specificazione

Esempio di output in GRETL: la diagnostica del modello è la seguente

$P-value[F_{(0,0577;2,492)}] = 0,9439$ si accetta l'ipotesi nulla in quanto il valore di probabilità è altissimo e pertanto il modello è correttamente specificato

I test in questione sono

Test di White	Test di Breush-Pagan	Test F
Test su eteroschedasticità degli errori	Test su eteroschedasticità degli errori	Modello ridotto

Test di Jarque – Bera	Test RESET di Ramsey	Test di Chow
Test di normalità degli errori	non-linearietà del modello	o di stabilità dei parametri

Test di Sargan	
Test su omoschedasticità?	

Test di White

E' un test di eteroschedasticità: riguarda la ipotesi che $\text{Var}(\varepsilon_i | X) = \sigma^2$.

IDEA: se $E(\varepsilon^2 | X) = \sigma^2$, allora la v.c. ε^2 è incorrelata sia con X che con sue trasformazioni quali i quadrati o i prodotti incrociati

H_0 : eteroschedasticità non presente

Una delle ipotesi basi dei modelli lineari è $\mathbf{V}(\varepsilon_i | \underline{X}_i = \underline{x}_i) = \sigma^2, \forall i \in \{1, \dots, n\}$, ovvero viene supposta l'omoschedasticità del modello. Se ciò non accadesse, però, avremmo che $\mathbf{V}(\varepsilon_i | \underline{X}_i = \underline{x}_i) = \sigma_i^2$, ovvero ogni residuo avrebbe varianza diversa. Ciò si può formulare anche in un'altra maniera: poiché la varianza dei residui è dipendente dall'indice i , allora potremmo supporre che esista una funzione $h(\cdot)$ tale da collegare le covariate \underline{x}_i alle varianze dei residui: in pratica $h(\underline{x}_i) = \sigma_i^2$.

Il test di White introduce una regressione "ausiliaria" per gli errori al quadrato: siccome so, dalle ipotesi del modello, che $E(\varepsilon_i | \underline{X}_i = \underline{x}_i) = 0$, ne consegue che

$$h(\underline{x}_i) = \mathbf{V}(\varepsilon_i | \underline{X}_i = \underline{x}_i) = E(\varepsilon_i^2 | \underline{X}_i = \underline{x}_i) - E(\varepsilon_i | \underline{X}_i = \underline{x}_i)^2 = E(\varepsilon_i^2 | \underline{X}_i = \underline{x}_i)$$

essendo la media al quadrato pari a 0.

Ciò equivale a dire che il piano di regressione per gli errori al quadrato è pari alla $h(\underline{x}_i)$, essendo $h(\underline{x}_i) = E(\varepsilon_i^2 | \underline{X}_i = \underline{x}_i)$, il che introduce un modello per gli errori al quadrato:

$$\varepsilon_i^2 = h(\underline{x}_i) + \nu_i \quad (1)$$

A questo punto, il modello per gli errori al quadrato viene assunto considerando tutte le combinazioni possibili fra covariate. Nel caso semplificato con due covariate abbiamo:

$$\varepsilon_i^2 = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i1}^2 + \gamma_4 x_{i2}^2 + \gamma_5 x_{i1} \cdot x_{i2} + \nu_i$$

Si noti che, con tre covariate, avremmo 15 coefficienti γ .

Se introduco la stima solita degli ε_i , stima che chiamo $e_i \triangleq y_i - \underline{x}_i^T \hat{\beta}$, dove \underline{x}_i^T è il vettore riga dell' i -esima unità, posso calcolare i vari $\hat{\gamma}_j, \forall j \in \{1, \dots, K\}$. Nel nostro caso bivariato $K = 5$.

White introduce una statistica test $H = R^2 \cdot n$, dove R^2 è il coefficiente di correlazione lineare. Si dimostra che $H \sim \chi_K^2$. Se H ha valori alti, allora vuol dire che la regressione ausiliaria spiega bene la variabilità degli errori al quadrato, e quindi è **leccito dire che gli errori dipendono dalle covariate**, che sarebbe come dire che vi è eteroschedasticità. Al contrario, un H basso mi permette di dire che gli errori non dipendono dalle covariate, la relazione diventa quindi

$$\varepsilon_i^2 = \gamma_0$$

e non vi è eteroschedasticità. Un test del genere si adatta quindi ad un sistema di ipotesi

$$\begin{cases} H_0 : \gamma_1 = \dots = \gamma_K = 0 \\ H_1 : \exists k : \gamma_k \neq 0 \end{cases}$$

Il test di White è uno dei diversi test sull'eteroschedasticità dei residui. Si tratta di test importanti in quanto il loro risultato consente all'utente di discriminare fra due metodi di stima diversi:

- Errori omoschedastici → posso usare gli stimatori dei minimi quadrati ordinari (OLS, spesso detti semplicemente LS): sono stimatori BLUE (*best linear unbiased estimator*), ovvero corretti, efficienti e consistenti;
- Errori eteroschedastici → se uso gli OLS ottengo stimatori corretti e consistenti, ma non efficienti; anzi, non sono per niente efficienti (varianza “esplosiva”) e quindi possono dare risultati approssimativi per quel che riguarda intervalli di confidenza e verifiche d’ipotesi, con possibili falsi positivi. In questo caso la strada da seguire è quella dei GLS, di cui fanno parte i WLS (*weighted least squares*) e gli stimatori di White. Entrambi sono più efficienti degli OLS con eteroschedasticità. Gli stimatori di White sono da preferire per grandi campioni, perché più efficienti dei WLS; tuttavia, di loro conosciamo solo risultati asintotici (correttezza asintotica, consistenza debole), di conseguenza, per piccoli campioni, la strada dei WLS è preferibile.

1. **Stimatore LS:** data X matrice delle osservazioni,

$$\underline{\beta}_{LS} = (X^T X)^{-1} X^T \underline{y} \quad (2)$$

2. **Stimatore WLS:** data X matrice delle osservazioni, Σ matrice di varianza e covarianza delle osservazioni,

$$\underline{\beta}_{WLS} = \left(X^T \hat{\Sigma}^{-1} X \right)^{-1} X^T \hat{\Sigma}^{-1} \underline{y} \quad (3)$$

Si noti come non è nient’altro che uno stimatore LS pesato da una matrice che “approssima” il reciproco delle varianze.

3. **Stimatore di White:** sostituisce la varianza degli stimatori OLS in caso di eteroschedasticità con una “varianza asintotica”:

$$\Sigma_{\underline{\beta}} = (X^T X)^{-1} \left(\sum_{i=1}^n e_i^2 \underline{x}_i^T \underline{x}_i \right) (X^T X)^{-1} \quad (4)$$

Così facendo, si evita di ottenere varianze “esplosive”, ma dovrò tener conto del fatto che i risultati che otterrò in seguito saranno tutti di tipo asintotico.

Formula di White (stimatore White di AVAR)

Test di Breush-Pagan

Viene utilizzato per testare l’eteroschedasticità in un modello di regressione lineare.

Verifica se la varianza degli errori di una regressione dipende dai valori delle variabili indipendenti. In tal caso, l’eteroschedasticità è presente.

Esso è valido per grandi campioni, assume che gli errori siano indipendenti e normalmente distribuiti e che la loro varianza sia funzione lineare del tempo t secondo $\ln(\sigma_t^2) = a + bt$ ciò implica che la varianza aumenti o diminuisca al variare di t , a seconda del segno di b .

Se si ha l'omoschedasticità, si realizza l'ipotesi nulla: $H_0: b = 0$

Per la sua verifica, si calcola una regressione lineare, a partire da un diagramma di dispersione che:

- sull'asse delle ascisse riporta il tempo t
- sull'asse delle ordinate il valore dei residui corrispondente

Si ottiene una retta di regressione, la cui devianza totale (SQR) è in rapporto alla devianza d'errore precedente (SQE) calcolata con i dati originari secondo una relazione di tipo quadratico che, se è vera l'ipotesi nulla, al crescere del numero delle osservazioni si distribuisce secondo una variabile casuale chi quadro con un grado di libertà.

PROCEDURA:

Passaggio 1: applicare OLS nel modello $y = X\beta + \varepsilon$ e calcolare i residui di regressione

Passaggio 2: eseguire la regressione ausiliaria $e_i^2 = \gamma_1 + \gamma_2 z_{2i} + \dots + \gamma_p z_{pi} + \eta_i$; z potrebbe essere parzialmente sostituito da variabili indipendenti x

Fase 3: la statistica del test è il risultato del coefficiente di determinazione della regressione ausiliaria nella fase 2 e della dimensione del campione n con $LM = nR^2$.

La statistica del test è asintoticamente distribuita come sotto χ_{p-1}^2 l'ipotesi nulla di omoschedasticità.

Test F (per significatività del modello)

Per verificare la significatività dell'intero modello si utilizza il test F.

$H_0: \beta_1 = 0, \dots, \beta_k = 0$. H_1 : almeno uno dei parametri sia diverso da zero.

IDEA: si deve decidere se un modello di regressione si adatta ai dati in modo significativamente migliore rispetto a un suo sotto-modello, cioè si vuole determinare se il modello con più regressori (modello 1) abbia un adattamento significativamente migliore (cioè un errore inferiore) ai dati rispetto al modello con meno regressori (modello 2). Date n osservazioni:

$$F = \frac{\left(\frac{RSS_1 - RSS_2}{p_2 - p_1} \right)}{\left(\frac{RSS_2}{n - p_2} \right)} \sim^{H_0} F_{(p_2 - p_1, n - p_2)}$$

L'ipotesi nulla viene rifiutata se la F calcolata dai dati è maggiore del valore critico della distribuzione F per una probabilità di falsa rigetto desiderata.

Test di Jarque–Bera

H_0 : normalità della distribuzione da cui è estratto campione.

Si basa sull'analisi dei momenti terzo (asimmetria) e quarto (curtosi) dal valor medio dei residui.

$JB = \frac{n}{6} (S^2 + \frac{(K-3)^2}{4})$, con n numero di osservazioni, S è la asimmetria del campione, K è la curtosi del campione definiti come $S = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\sigma^2)^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^3}{(\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2)^{3/2}}$

$K = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\sigma^2)^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^4}{(\frac{1}{n} \sum_{i=1}^n (x - \bar{x})^2)^2}$ dove μ_3 e μ_4 sono il terzo e quarto momento centrale, \bar{x} è la media campionaria e σ^2 è la varianza. $JB \sim_{n>>0} \chi_2^2$

CRITERIO: l'ipotesi nulla viene rigettata per valori di JB troppo grandi.

Test RESET: test di Ramsey

REgression Specification Error Test

Verifica:

- la non appropriata scelta della forma funzionale
- la possibile omissione di variabili nel modello

Intuizione: se combinazioni non lineari delle variabili esplicative hanno qualche potere nello spiegare la variabile risposta, allora il modello è errato. Nel senso che il processo di generazione dei dati potrebbe essere approssimato in modo migliore da un polinomio o da un altro funzionale non lineare.

Il test RESET non indica alternative specifiche al modello dato: si dice che è un "test non costruttivo", cioè non fornisce indicazioni su come agire per superare il problema.

Non è un test generale di non-linearità del modello; mette a confronto diverse alternative:

- relazione fra \underline{x} e y lineare;
- relazione fra \underline{x} e y quadratica;
- relazione fra \underline{x} e y cubica;
- relazione fra \underline{x} e y mista;

Esempio di modello quadratico completo a due covariate:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1}^2 + \beta_4 x_{i2}^2 + \beta_5 x_{i1} x_{i2} + \varepsilon_i \quad (5)$$

Per il test di Ramsay, la relazione quadratica è semplificata dalla regressione ausiliaria:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma \hat{y}_i^2 + \varepsilon_i \quad (6)$$

Questa semplificazione è giustificata dal seguente ragionamento: in una regressione lineare semplice, i valori previsti \hat{y}_i sono ottenuti tramite

$$\hat{y}_i = \mathbf{X} \underline{\hat{\beta}}$$

che, nel modello a due covariate, corrisponde a

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$

di conseguenza, tale quantità, elevata al quadrato, dà

$$\hat{y}_i^2 = \hat{\beta}_0^2 + \hat{\beta}_1^2 x_{i1}^2 + \hat{\beta}_2^2 x_{i2}^2 + 2\hat{\beta}_1 \hat{\beta}_2 x_{i1} x_{i2} + 2\hat{\beta}_0 \hat{\beta}_1 x_{i1} + 2\hat{\beta}_0 \hat{\beta}_2 x_{i2}$$

sostituisco all'interno della (6)

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma(\hat{\beta}_0^2 + \hat{\beta}_1^2 x_{i1}^2 + \hat{\beta}_2^2 x_{i2}^2 + 2\hat{\beta}_1 \hat{\beta}_2 x_{i1} x_{i2} + 2\hat{\beta}_0 \hat{\beta}_1 x_{i1} + 2\hat{\beta}_0 \hat{\beta}_2 x_{i2}) + \varepsilon_i = \\ &= \beta_0 + \gamma \hat{\beta}_0^2 + (\beta_1 + 2\gamma \hat{\beta}_0 \hat{\beta}_1) x_{i1} + (\beta_2 + 2\gamma \hat{\beta}_0 \hat{\beta}_2) x_{i2} + \gamma \hat{\beta}_1 x_{i1}^2 + \gamma \hat{\beta}_2 x_{i2}^2 + 2\gamma \hat{\beta}_1 \hat{\beta}_2 x_{i1} x_{i2} + \varepsilon_i \end{aligned}$$

a seguito di attività di riparametrizzazione, ottengo:

$$y_i = y_i = \beta_0^* + \beta_1^* x_{i1} + \beta_2^* x_{i2} + \beta_3^* x_{i1}^2 + \beta_4^* x_{i2}^2 + \beta_5^* x_{i1} x_{i2} + \varepsilon_i$$

che è del tutto uguale alla (5).

Identici sono i ragionamenti per il modello cubico:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma \hat{y}_i^3 + \varepsilon_i \quad (7)$$

e misto:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + \varepsilon_i \quad (8)$$

In generale, posso riassumere le componenti tipo γ in un vettore $\underline{\gamma}$. La statistica test per il test di Ramsay è la stessa H del test di White:

$$H = R^2 \cdot n; H \sim \chi^2_{\dim(\underline{\gamma})}$$

ricordando che la relazione è asintotica se abbiamo usato le stime della varianza di White. Il test di Ramsay risponde quindi al seguente set d'ipotesi:

$$\begin{cases} H_0 : \underline{\gamma} = \underline{0} \\ H_1 : \underline{\gamma} \neq \underline{0} \end{cases}$$

Importante: il test di Ramsay può dare falsi positivi in caso di eteroschedasticità.
In tal caso, quindi, la strada migliore da seguire è quella delle analisi grafiche (con una covariata: grafico y su x , con più covariate: grafico residui su covariate prese singolarmente oppure residui su valori previsti).

Test di Chow

Un test effettuabile per verificare la stabilità strutturale (ovvero la stabilità dei parametri nel tempo). Esso ricerca se esiste ed è significativa una data di rottura strutturale, ovvero un momento temporale dove si verifica un cambiamento netto dei parametri della regressione. Per farlo si modifica la regressione inserendo una variabile dummy che assuma valore 0 in un certo intervallo di tempo e 1 in un altro.

Ipotesi nulla: la funzione di regressione è la stessa prima e dopo la data di rottura, ossia i coefficienti della variabile binaria sono nulli.

CONTESTO: Per verificare la presenza di rotture strutturali in una funzione di regressione temporale, si assume che il campione termini prima di quando effettivamente accada e si valutano le previsioni che si sarebbero fatte in questo caso. Le *rotture strutturali* sono individuate dal fatto che la prestazione delle previsioni è sostanzialmente peggiore di quanto ci si aspettava. Se si ha una rottura strutturale nella regressione durante il periodo campionario, allora le stime OLS per l'intero campione stimeranno una relazione valida in media, nel senso che la stima combinerà i due differenti periodi. A seconda della posizione e dell'ampiezza della rottura, la funzione di regressione media può essere molto diversa dalla vera funzione di regressione alla fine del campione, e ciò porta a cattive previsioni.

Cioè, sotto l'ipotesi nulla di assenza di rotture, $\gamma_0 = \gamma_1 = \gamma_2 = 0$.

Sotto l'ipotesi alternativa che ci sia una rottura, la funzione di regressione è diversa prima e dopo la data della rottura τ , nel qual caso almeno uno dei γ è non nullo. Perciò l'ipotesi di una rottura può essere verificata utilizzando la statistica F che verifica l'ipotesi $\gamma_0 = \gamma_1 = \gamma_2 = 0$ contro l'ipotesi che almeno uno dei γ sia non nullo.

Se ci sono predittori multipli o più ritardi, allora questo test può essere esteso costruendo variabili binarie d'interazione per tutti i regressori e verificando l'ipotesi che tutti i coefficienti dei termini che coinvolgono $D_t(\tau)$ siano nulli.

Questo approccio può essere modificato per verificare la presenza di una rottura in un sottoinsieme dei coefficienti includendo solo le interazioni con le variabili binarie per il sottoinsieme di regressori d'interesse.

Verificare una rottura strutturale ad una data ignota. Spesso la data di una possibile rottura è ignota o è conosciuta solo entro un intervallo. Si supponga, ad esempio, che la rottura sia avvenuta ad una data compresa tra τ_0 e τ_1 . Il test di Chow può in questo caso essere modificato verificando la presenza di una rottura per tutte le possibili date τ tra τ_0 e τ_1 , e poi utilizzando la più grande delle statistiche F risultanti per verificare la presenza di una rottura ad una data ignota. Questo test di Chow modificato è chiamato in vari modi, **statistica del rapporto delle verosimiglianze** o **QLR** (acronimo dall'inglese *Quandt Likelihood Ratio*) (Quandt, 1960), o più oscuramente, **statistica sup di Wald** (in inglese *sup-Wald statistic*).

È un test originariamente pensato per le serie storiche; testa la presenza o meno di un break strutturale all'interno del campione: ci si chiede se sia lecito pensare che, all'interno del nostro campione, vi sia un numero sostanziale di unità con coefficienti comuni fra loro ma significativamente diversi dal resto del campione. Formalmente, dato il campione C :

$$\begin{aligned} C &= C_1 \cup C_2, C_1 \cap C_2 = \emptyset \\ C_1 &\longmapsto \underline{\beta}_1 \\ C_2 &\longmapsto \underline{\beta}_2 \\ \underline{\beta}_1 &\neq \underline{\beta}_2 \end{aligned}$$

Il campione è diviso in due sottocampioni C_1, C_2 , che conducono a due stime diverse dei coefficienti. Lo scopo del test di Chow è valutare se questa differenza sia o meno statisticamente significativa.

Per discriminare fra i due gruppi, introduco una variabile indicatrice (dummy) che identificherà le unità del sottocampione C_1 :

$$z_i = \begin{cases} 0, & \text{se } i \in C_1 \\ 1, & \text{se } i \notin C_1 \end{cases}$$

La regressione ausiliaria introdotta da Chow è la seguente:

$$y_i = \underline{x}_i^T \underline{\beta} + z_i \cdot \underline{x}_i^T \underline{\gamma} + \varepsilon_i \quad (9)$$

vengono quindi introdotti tanti coefficienti γ tanti quanti sono i β . Nella semplificazione per due covariate, la regressione diviene

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \gamma_0 z_i + \gamma_1 x_{i1} z_i + \gamma_2 x_{i2} z_i + \varepsilon_i$$

Raccogliendo si ottiene:

$$y_i = (\beta_0 + \gamma_0 z_i) + (\beta_1 + \gamma_1 z_i) x_{i1} + (\beta_2 + \gamma_2 z_i) x_{i2} + \varepsilon_i$$

abbiamo così evidenziato il ruolo additivo dei coefficienti della regressione ausiliaria fra elementi di C_2 e C_1 .

$$y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}_{\text{Parte baseline, comune a } C_1, C_2} + \underbrace{\gamma_0 z_i + \gamma_1 x_{i1} z_i + \gamma_2 x_{i2} z_i}_{\text{Parte additiva portata da } C_1} + \varepsilon_i$$

L'assenza di *break* strutturale si ha se tutti i coefficienti γ sono nulli, ovvero se non vi è

regressione con variabili strumentali. Introduciamo la regressione con variabili strumentali come metodo generale per trattare la correlazione tra errore e regressore, la quale può nascere per diverse ragioni, inclusa la causalità simultanea.

Oss. Simultanea causalità (o interdipendenza) delle variabili nel modello regressione: Y ha effetto sulla X ed X ha effetto sulla Y.

Variabili strumentali

Un'ipotesi standard del modello classico di regressione lineare è che le variabili esplicative non siano correlate con l'errore; laddove tale ipotesi viene meno, la regressione con il consueto metodo dei minimi quadrati non consentirà di ottenere stime consistenti (cioè asintoticamente corrette e con varianza asintoticamente nulla). Se tuttavia è disponibile una variabile strumentale, allora è ancora possibile ottenere stime consistenti.

Il metodo di stima di un modello lineare tramite variabili strumentali è anche noto come *metodo dei minimi quadrati a due stadi* (o **2SLS**, dall'inglese Two-Stages Least Squares).

Concetto chiave 10.1: *il modello generale di regressione con variabili strumentali e la sua terminologia*

Il modello generale di regressione IV è

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \cdots + \beta_{k+r} W_{ri} + u_i, \quad (10.12)$$

$i = 1, \dots, n$, dove:

- Y_i è la variabile dipendente;
- u_i è l'errore, che rappresenta errori di misura e/o fattori omessi;
- X_{1i}, \dots, X_{ki} sono k regressori endogeni, potenzialmente correlati con u_i ;
- W_{1i}, \dots, W_{ri} sono r regressori esogeni inclusi, incorrelati con u_i ;
- $\beta_0, \beta_1, \dots, \beta_{k+r}$ sono coefficienti di regressione ignoti;
- Z_{1i}, \dots, Z_{mi} sono m variabili strumentali.

I coefficienti sono sovraidentificati se ci sono più strumenti che regressori endogeni ($m > k$); sono sottoidentificati se $m < k$; e sono esattamente identificati se $m = k$. La stima del modello di regressione IV richiede l'identificazione esatta o la sovraidentificazione.

Concetto chiave 10.2: minimi quadrati a due stadi (TSLS)

Lo stimatore TSLS per il modello generale di regressione IV (10.12) con più variabili strumentali si calcola in due stadi.

1. **Regressione(i) del primo stadio.** Si effettua una regressione di X_{1i} sulle variabili strumentali (Z_{1i}, \dots, Z_{mi}) e sulle variabili esogene incluse (W_{1i}, \dots, W_{ri}) tramite gli OLS. Si calcolano i valori predetti da questa regressione, indicati con \hat{X}_{1i} . Si ripete lo stesso procedimento per tutti i regressori endogeni X_{2i}, \dots, X_{ki} , calcolando quindi i valori predetti $\hat{X}_{1i}, \dots, \hat{X}_{ki}$.
2. **Regressione del secondo stadio.** Si effettua una regressione di Y_i sui valori predetti delle variabili endogene ($\hat{X}_{1i}, \dots, \hat{X}_{ki}$) e sulle variabili endogene incluse (W_{1i}, \dots, W_{ri}) tramite gli OLS. Gli stimatori TSLS $\hat{\beta}_0^{TSLS}, \dots, \hat{\beta}_{k+r}^{TSLS}$ sono gli stimatori ottenuti dalla regressione del secondo stadio.

In pratica, i moderni software econometrici accorpano i due stadi nel comando per la stima TSLS.

Concetto chiave 10.3: le due condizioni per la validità degli strumenti

Un insieme di m strumenti, Z_{1i}, \dots, Z_{mi} , deve soddisfare le due condizioni seguenti per essere valido:

1. **Rilevanza degli strumenti**
 - In generale, sia \hat{X}_{1i}^* il valore predetto di X_{1i} dalla regressione di X_{1i} sugli strumenti (le Z) e i regressori esogeni inclusi (i W) e si indichi con “1” un regressore che è uguale a “1” per tutte le osservazioni (il suo coefficiente è l’intercetta). Allora $(\hat{X}_{1i}^*, \dots, \hat{X}_{ki}^*, W_{1i}, \dots, W_{ri}, 1)$ non sono perfettamente collineari.
 - Se c’è una sola X , allora almeno una delle Z deve essere inclusa nella regressione di X sulle Z e i W .
2. **Esogeneità degli strumenti** Gli strumenti sono incorrelati con l’errore, ovvero $\text{corr}(Z_{1i}, u_i) = 0, \dots, \text{corr}(Z_{mi}, u_i) = 0$.

Concetto chiave 10.4: le assunzioni della regressione IV

Le variabili e gli errori nel modello di regressione IV del concetto chiave 10.1 soddisfano:

1. $E(u_i | W_{1i}, \dots, W_{ri}) = 0$;
2. $(X_{1i}, \dots, X_{ki}, W_{1i}, \dots, W_{ri}, Z_{1i}, \dots, Z_{mi}, Y_i)$ sono estratti i.i.d. dalla loro distribuzione congiunta;
3. le X , i W , le Z e la u hanno momenti quarti non nulli e finiti;
4. i W non sono perfettamente collineari;
5. valgono le due condizioni del concetto chiave 10.3 per la validità di uno strumento.

Concetto chiave 10.5: una regola del pollice per valutare la debolezza degli strumenti

La statistica F del primo stadio è la statistica F per verificare l'ipotesi che i coefficienti degli strumenti Z_{1i}, \dots, Z_{mi} siano uguali a zero nel primo stadio dei minimi quadrati a due stadi. Quando c'è un singolo regressore endogeno, una F del primo stadio minore di 10 indica che gli strumenti sono deboli, nel qual caso lo stimatore TSLS è distorto (anche in grandi campioni) e la statistica t e gli intervalli di confidenza per i TSLS sono inaffidabili.

Concetto chiave 10.6: il test delle restrizioni di sovraidentificazione (la statistica J)

Siano \hat{u}_i^{TSLS} i residui dalla stima TSLS (10.12). Usiamo gli OLS per stimare i coefficienti nella regressione

$$\hat{u}_i^{TSLS} = \delta_0 + \delta_1 Z_{1i} + \dots + \delta_m Z_{mi} + \delta_{m+1} W_{1i} + \dots + \delta_{m+r} W_{ri} + e_i, \quad (10.17)$$

dove e_i è l'errore della regressione. Si indichi con F la classica statistica F per verificare l'ipotesi che $\delta_1 = \dots = \delta_m = 0$. La statistica test per le restrizioni di sovraidentificazione è $J = mF$. Sotto l'ipotesi nulla che tutti gli strumenti siano esogeni, J si distribuisce in grandi campioni come una χ^2_{m-k} , dove $m - k$ è il “grado di sovraidentificazione”, ovvero la differenza tra il numero di strumenti e il numero di regressori endogeni.

Applicazione del metodo: La correlazione tra regressori e disturbi in un modello di regressione lineare può insorgere in una serie di circostanze. Alcuni casi notevoli, generalmente menzionati in letteratura, sono:

- Omissione di variabili rilevanti, se il modello di regressione (multivariata) non include tra i regressori una variabile, che pure avrebbe rilevante potere esplicativo nei confronti della variabile dipendente;
- Errore nelle variabili esplicative, laddove i dati relativi a uno o più regressori sono affetti da un errore di misura, distinto dal disturbo ε
- Equazioni simultanee, nei casi in cui il sistema oggetto di analisi mette insieme diversi modelli statistici che operano simultaneamente.

Il metodo delle variabili strumentali è spesso applicato con una procedura di stima con i minimi quadrati a due stadi (in inglese, Two-Stages Least Squares, o 2SLS). Nell'approccio 2SLS, in un primo stadio di stima i regressori sono regrediti sulle variabili strumentali, ottenendo dei valori di previsione di primo stadio. Nel secondo stadio, la variabile dipendente è regedita sui valori di previsione di primo stadio, ottenendo le stime $\hat{\beta}_{IV}$.

Metodo delle variabili strumentali

È una tecnica che sfrutta il metodo dei momenti al fine di stimare i parametri di un modello lineare con variabile esplicativa endogena:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i; \mathbf{E}(x_i \varepsilon_i) \neq 0, X \text{ endogena}$$

Il metodo prevede prima di tutto di designare una variabile Z , detta **strumento**, esogena rispetto all'errore, ma fortemente correlata con la variabile X .

$$\mathbf{COV}(z_i, \varepsilon_i) = 0, \forall i ; |\mathbf{COR}(z_i, x_i)| \simeq 1, \forall i$$

Uno strumento che rispetta la seconda condizione indicata si dice **strumento forte**. Le condizioni di ortogonalità sono quindi

$$\begin{aligned} \mathbf{E}(\varepsilon_i) &= 0 \\ \mathbf{E}(z_i \varepsilon_i) &= 0 \end{aligned}$$

Le due condizioni rappresentano il seguente sistema

$$\begin{cases} \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \\ \sum (y_i - \beta_0 - \beta_1 x_i) z_i = 0 \end{cases}$$

che è leggermente diverso da quello visto in (12). Alla fine, esso conduce alle seguenti soluzioni, che vengono dette delle variabili strumentali, IV (*instrumental variables*):

$$\hat{\beta}_{0IV} = \bar{y} - \hat{\beta}_{1IV} \bar{x} \tag{13}$$

$$\hat{\beta}_{1IV} = \frac{\hat{\mathbf{COV}}(z_i, y_i)}{\hat{\mathbf{COV}}(z_i, x_i)} \tag{14}$$

Gli stimatori ottenuti tramite il metodo IV non sono, in generale, corretti, ma sono consistenti in probabilità.

Metodo dei minimi quadrati a due stadi

Affine al metodo delle variabili strumentali è il metodo dei minimi quadrati a due stadi. Esso si applica nel caso in cui si vuole utilizzare un numero di strumenti superiore alle variabili da sostituire, ovvero vi è un caso di sovraidentificazione fra strumenti e regressori endogeni. È piuttosto facile notare che, in questo caso, applicare il metodo dei momenti equivarrebbe ad ottenere un sistema di tre equazioni in due incognite, che è usualmente un sistema con infinite soluzioni. Possiamo quindi agire nei seguenti modi:

- Sopprimere uno dei due strumenti, ma perdere l'informazione portata dallo strumento eliminato;
- Utilizzare altri metodi che permettano di conservare entrambi gli strumenti senza perdere alcuna informazione derivante da questi.

Uno dei metodi indicato è il metodo dei minimi quadrati a due stadi. Si avvale di una regressione ausiliaria che formula un modello lineare fra il regressore endogeno (qui chiamato X) ed i suoi strumenti (Z_1, Z_2):

$$x_i = \pi_0 + \pi_1 z_{i1} + \pi_2 z_{i2} + \nu_i$$

Il procedimento più semplice è, a questo punto, stimare π_0, π_1, π_2 con il metodo dei minimi quadrati. Il risultato così ottenuto ci permetterà di calcolare anche i valori previsti di $x_i, \forall i$, ovvero gli \hat{x}_i . Nota:

$$\text{COV}(\hat{X}_i, \varepsilon_i) = \underbrace{\hat{\pi}_0 \mathbf{E}(\varepsilon_i)}_{\text{La media degli errori è } 0} + \underbrace{\hat{\pi}_1 \text{COV}(Z_{i1}, \varepsilon_i)}_{0 \text{ per def. di var. esogena}} + \underbrace{\hat{\pi}_2 \text{COV}(Z_{i2}, \varepsilon_i)}_{0 \text{ per def. di var. esogena}} = 0$$

Quindi \hat{X} è una variabile endogena e può essere utilizzata al posto della X stessa, essendone una buona approssimazione:

$$y_i = \beta_0 + \beta_1 \hat{x}_{i1} + \varepsilon_i$$

posso quindi procedere stimando i $\underline{\beta}$ tramite il metodo dei minimi quadrati ordinari, ottenendo così uno stimatore $\hat{\underline{\beta}}_{2SLS}$ (2SLS=two stages least squares).

Test di Sargan (o test per la sovraidentificazione 357/698 SeW (arrivato qui)

Questo test verifica se tutti gli strumenti, esclusi e inclusi, sono effettivamente incorrelati con il termine di errore dell'equazione strutturale.

In particolare il test di Sargan è un test di sovraidentificazione, l'ipotesi da verificare è quella che gli strumenti siano incorrelati con il termine d'errore, ovvero che gli strumenti siano validi.

Per calcolare il test dobbiamo usare una regressione ausiliaria, la quale consiste nel regredire i residui del modello stimato su tutti gli strumenti, a questo punto possiamo calcolare il valore della statistica test, per fare questo ci sono due metodi equivalenti, uno più teorico e l'altro più pratico.

Il primo metodo considera la seguente statistica test:

$$S = \frac{\hat{\epsilon}' z(z'z)^{-1} z' \hat{\epsilon}}{\hat{\sigma}^2} \sim \chi^2_{R-K}, \text{ sotto } H_0$$

dove R è il numero totale delle variabili strumentali e K è il numero delle esplicative.

Il secondo metodo è di più facile applicazione, sarà quello che useremo in questa relazione.

Il calcolo della statistica test è il seguente:

$$S = NR^2 \sim \chi^2_{R-K}, \text{ sotto } H_0$$

dove N è la numerosità del campione, mentre R^2 è l'indicatore della bontà di *fitting* della regressione ausiliaria.

1.5 Esercitazioni al computer in GRET

Capitolo 2. Modelli econometrici per serie temporali: modelli statici e modelli dinamici

Le serie storiche

Lo scopo dello studio delle serie storiche è la **previsione** a partire dai dati passati con la mediazione di **modelli statistici per serie storiche**.

Le ipotesi alla base sono le seguenti:

- i. $COR(Y_t, Y_{t-1}) \neq 0 \rightarrow$ È prevista correlazione fra i dati della serie (*correlazione seriale*)
- ii. Date le osservazioni del carattere di studio nel tempo y_1, \dots, y_T realizzate tutte in momenti diversi, le osservazioni sono assunte come una delle infinite **realizzazioni di un processo stocastico** (vedi sez. 8)

Il processo stocastico (a tempo discreto)

8 Il processo stocastico

Alla base della serie storica osservata y_1, \dots, y_T presuppongo essere un processo stocastico Y_1, \dots, Y_T formato da variabili aleatorie e riassunto con $\{Y_t\}_{t \in \{1, \dots, T\}} = \{Y_t\}_{t \in \mathcal{T}}$, dove $\mathcal{T} = \{1, \dots, T\}$, ovvero insieme di istanti temporali.

Poiché il processo stocastico è composto di variabili aleatorie, queste possono essere diverse fra loro nella distribuzione ed, in generale, hanno media e varianza differenti, quindi la realizzazione di un processo non presenta, di solito, osservazioni simili.

Si definisce **famiglia delle distribuzioni di probabilità congiunte** del processo stocastico il seguente insieme:

$$\mathcal{L} = \{F(y_{t_1}, \dots, y_{t_n}), \forall t_1 \neq \dots \neq t_n \in \mathcal{T}, \forall n \in \mathbb{N}\}$$

Se fosse $\{Y_t\}_{t \in \mathcal{T}}$ un processo stocasticamente indipendente, ovvero in cui le Y_i sono i.i.d., allora potremmo dire $f(y_1, y_2, \dots, y_T) = \prod_{t=1}^T f(y_t)$, ma ciò, in generale, non vale perché le T.v.a. non sono i.i.d. ma sono dipendenti fra loro, quindi vale $f(y_1, y_2, \dots, y_T) = f(y_1) \cdot f(y_2|y_1) \cdot f(y_3|y_2, y_1) \cdots f(y_T|y_{T-1}, \dots, y_1)$.

Poiché ogni terminazione del processo stocastico è una v.a., possiamo calcolare un valore atteso ed una varianza per ognuna di esse, avremo quindi una **funzione valore atteso** e una **funzione varianza** per ogni processo.

- **Funzione valore atteso:** è definita da $\mathbf{E}(Y_t) \stackrel{\Delta}{=} \mu_t \stackrel{\Delta}{=} \{\mathbf{E}(Y_t) | t \in \mathcal{T}\}$, ovvero l'insieme dei valori attesi al variare di t in \mathcal{T} .
- **Funzione varianza:** è definita dalla funzione valore atteso tramite la relazione $\mathbf{V}(Z) = \mathbf{E}(Z^2) - \mathbf{E}^2(Z)$ con Z v.a. Si ha quindi $\mathbf{V}(Y_t) \stackrel{\Delta}{=} \sigma_t^2 = \mathbf{E}(Y_t^2) - \mathbf{E}^2(Y_t)$.
- Si ricava di conseguenza la **Funzione di [auto]covarianza**: $\text{COV}(Y_t, Y_u) \stackrel{\Delta}{=} \gamma_{t,u} = \mathbf{E}[(Y_t - \mathbf{E}(Y_t))(Y_u - \mathbf{E}(Y_u))]$.
- Da cui ricaviamo quindi la **Funzione di [auto]correlazione**: $\rho_{t,u} = \frac{\gamma_{t,u}}{\sqrt{\mathbf{V}(Y_t)\mathbf{V}(Y_u)}}$. Il grafico dell'autocorrelazione è detto *autocorrelogramma* o *ACF* dall'inglese *AutoCorrelation Function*.

Teorema di Kolmogorov: per descrivere il processo aleatorio non occorre fornire la distribuzione congiunta di infinite (numerabili variabili aleatorie), ma è sufficiente fornire tutte le distribuzioni congiunte finito-dimensionali

Stazionarietà del processo asintotico

$\{Y_t\}_{t \in \mathcal{T}}$ è definito **processo stocastico stazionario in senso forte** (*p.s.s. forte*) se, data la successione di istanti temporali consecutivi t_1, \dots, t_n e data la medesima successione *shiftata* di un valore temporale h tale che $h > n$, vale $F(y_{t_1}, \dots, y_{t_n}) = F(y_{t_1+h}, \dots, y_{t_n+h})$. Conseguentemente a ciò $F(y_1) = \dots = F(y_T)$.

Il processo è definito **stazionario in senso debole** se rispetta le seguenti caratteristiche:

- 1) $\mathbf{E}(Y_t) = \mu \in \mathbb{R}; \forall t \in \mathcal{T}$, quindi il valore atteso è uguale in tutti gli istanti ed è un numero reale;
- 2) $\mathbf{V}(Y_t) = \sigma^2 \in \mathbb{R}^+ \cup \{0\}; \forall t \in \mathcal{T}$, ovvero la varianza è uguale in tutti gli istanti ed è un numero reale.

Il che implica che il processo si attesti su una media e una varianza fissi. Ai punti 1) e 2) consegue:

3.1)

$$\text{COV}(y_t, y_{t-\tau}) = \gamma_{|\tau|}; \tau \in \mathbb{Z} \quad (17)$$

ovvero l'autocorrelazione non dipende dalla posizione degli istanti ma solamente dalla differenza temporale fra questi due. Poiché la differenza è calcolata in valore assoluto, allora vale $\gamma_\tau = \gamma_{-\tau}$ quindi γ è una **funzione pari**; di seguito il tutto viene semplificato indicando solo γ_τ senza i valori assoluti;

3.2) Stessa cosa applicata a ρ anziché γ .

Nota bene: p.s.s. forte \Leftrightarrow p.s.s. debole: infatti dalla seconda non prescinde la prima dato che non vale sempre che dall'uguaglianza dei primi due momenti possa risalere all'uguaglianza della funzione di ripartizione; un p.s.s. forte non è debole perché può accadere che $\mathbf{E}(Y_t) = +\infty$ o $\mathbf{V}(Y_t) = +\infty$.

Un esempio: i residui ε_t genericamente sono definiti come aventi media nulla e varianza finita σ^2 e formano quindi un p.s.s. debole. Posso calcolare:

$$\gamma_\tau = \begin{cases} 0 & \text{se } \tau \neq 0 \\ \sigma^2 & \text{se } \tau = 0 \end{cases} \Rightarrow \rho_\tau = \begin{cases} 0 & \text{se } \tau \neq 0 \\ 1 & \text{se } \tau = 0 \end{cases}$$

Riassumendo le condizioni precedenti:

$$\mathbf{E}(\underline{\varepsilon}) = \underline{0}; \mathbf{V}(\underline{\varepsilon}) = \sigma^2 \cdot \mathbf{I}; \varepsilon_t \perp \varepsilon_u \forall t \neq u$$

Di conseguenza, i residui non sono utili al fine previsivo dato che non c'è correlazione col passato (è un fattore comune a tutti i processi i.i.d.).

Il processo White Noise

$$\{\varepsilon_t\}_t \sim \text{WN}(\sigma^2) \Leftrightarrow [(\mu_t = 0, \forall t \in \mathcal{T}) \wedge (\sigma_t^2 = \sigma^2 \in \mathbb{R}, \forall t \in \mathcal{T}) \wedge (\gamma_\tau = 0, \forall \tau \neq 0)]$$

La differenza fra un processo White Noise e un processo i.i.d. come prima definito è che il WN non prevede indipendenza stocastica fra le varie terminazioni, ma solo correlazione nulla. Di norma, però, il processo white noise sottintende che $\varepsilon_t \sim N(0; \sigma^2)$, di conseguenza la non correlazione implica indipendenza stocastica e la caratterizzazione della funzione di densità (e quindi di quella di ripartizione) è data dai primi due momenti centrali (valore atteso e varianza); quindi, nel caso dei WN normali e dei processi gaussiani in generale, p.s.s. forte \Leftrightarrow p.s.s debole.

La notazione LAG

Lag significa *ritardo* in inglese.

Per un processo stocastico, vale la seguente definizione:

$$L(y_t) = Ly_t = y_{t-1} \quad (23)$$

$$(24)$$

2.1 Proprietà dello stimatore OLS dei parametri del modello di regressione lineare statico

2.2 Inferenza robusta all'eteroschedasticità e alla correlazione seriale dei residui.

2.3 Test di correlazione seriale dei residui.

2.4 1) Il modello AR(1) ed il modello di regressione lineare dinamico.

2.5 Proprietà dello stimatore OLS dei parametri del modello di regressione lineare dinamico.

2.6 Test di corretta specificazione del modello e criteri informativi per la scelta dei ritardi.

2.7 2) Cenni sui modelli ARMA(p,q) e l'approccio Box - Jenkins.

Il modello a media mobile

Un processo stocastico è in media mobile di ordine q se:

$$\begin{aligned} \{y_t\}_t \sim MA(q) \iff & y_t = \varepsilon_t + a_{t-1}\varepsilon_{t-1} + \dots + a_{t-q}\varepsilon_{t-q} \\ & = \sum_{j=1}^q a_j \varepsilon_{t-j}; \text{ con } a_0 = 1 \text{ e } \varepsilon_t \sim WN(\sigma^2) \end{aligned} \tag{25}$$

Il valore attuale è composto da una somma pesata di *shock* presenti e passati.

Il modello MA(1)

$$y_t = \varepsilon_t + a_1 \varepsilon_{t-1}$$

È un p.s.s. debole, infatti:

- ▷ $\mathbf{E}(Y_t) = \mathbf{E}(\varepsilon_t + a_1 \varepsilon_{t-1}) = \mathbf{E}(\varepsilon_t) + a_1 \mathbf{E}(\varepsilon_{t-1}) = 0$
- ▷ $\mathbf{V}(Y_t) = \mathbf{V}(Y_t)(\varepsilon_t + a_1 \varepsilon_{t-1}) \stackrel{\text{cov}=0}{=} \mathbf{V}(\varepsilon_t) + \mathbf{V}(a_1 \varepsilon_{t-1}) = \sigma^2 + a_1^2 \sigma^2 = \sigma^2(a_1^2 + 1)$

È subito visibile come il processo MA(1) sia stazionario. Possiamo controllare anche come questo porti ad avere valori di covarianza non dipendenti da t , ma solo da τ :

$$\begin{aligned} \mathbf{COV}(Y_t, Y_{t-\tau}) &= \mathbf{E}[(Y_t - \mathbf{E}(Y_t))(Y_{t-\tau} - \mathbf{E}(Y_{t-\tau}))] = \\ &\stackrel{\mathbf{E}(Y_t)=0}{=} \mathbf{E}(Y_t \cdot Y_{t-\tau}) = \\ &= \mathbf{E}(\varepsilon_t \varepsilon_{t-\tau}) + a_1 \mathbf{E}(\varepsilon_{t-1} \varepsilon_{t-\tau-1}) + a_1^2 \mathbf{E}(\varepsilon_t \varepsilon_{t-\tau}) + a_1 \mathbf{E}(\varepsilon_{t-1} \varepsilon_{t-\tau}) \end{aligned} \quad (26)$$

Si notino due cose:

- $\mathbf{COV}(X, Z) = \mathbf{E}[(X - \mathbf{E}(X))(Z - \mathbf{E}(Z))]$, per cui, se $\mathbf{E}(X) = \mathbf{E}(Z) = 0$ la relazione diventa $\mathbf{COV}(X, Z) = \mathbf{E}[XZ]$
- $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X)$, per cui, se $\mathbf{E}(X) = 0 = \mathbf{E}^2(X)$, la relazione diventa $\mathbf{V}(X) = \mathbf{E}(X^2)$

Tenendo presente il risultato b., noto che $\mathbf{V}(\varepsilon_t) = \mathbf{E}(\varepsilon_t^2)$. Parimenti, ricordando a., si vede che $\mathbf{COV}(\varepsilon_t, \varepsilon_{t-\tau}) = \mathbf{E}(\varepsilon_t \varepsilon_{t-\tau})$, essendo il processo $\{\varepsilon_t\}_t$ a media nulla. Considerato ciò, la (26) può essere separata in casi:

- Se $\tau = 0$ ho

$$\mathbf{COV}(Y_t, Y_t) = \mathbf{V}(Y_t) = a_1^2 \sigma^2$$

- Se $\tau = 1$ ho

$$\mathbf{E}(\varepsilon_t \varepsilon_{t-1}) + a_1 \mathbf{E}(\varepsilon_{t-1} \varepsilon_{t-2}) + a_1^2 \mathbf{E}(\varepsilon_t \varepsilon_{t-1}) + a_1 \mathbf{E}(\varepsilon_{t-1} \varepsilon_{t-1})$$

È facile notare come i primi tre termini siano termini *rettangoli*, ovvero in cui l'indice non corrisponde. Per a., essi sono una covarianza fra due *shock* diversi, covarianza che è nulla essendo il processo degli *shock* un rumore bianco. L'ultimo termine è invece un termine *quadrato*, ovvero in cui i due indici corrispondono e si "uniscono" a formare un $\mathbf{E}(\varepsilon_{t-1}^2)$, che, per b., è pari a $\mathbf{V}(\varepsilon_t) = \sigma^2$, di conseguenza

$$\mathbf{COV}(Y_t, Y_{t-1}) = 0 + 0 + 0 + a_1 \sigma^2$$

- Se $\tau > 1$ si nota facilmente come i termini che si ripresentano all'interno della (26) siano tutti *rettangoli*: la covarianza è quindi nulla.

Riassumendo

$$\text{COV}(Y_t, Y_{t-\tau}) = \begin{cases} \sigma^2 & \text{se } \tau = 0 \\ a_1\sigma^2 & \text{se } \tau = 1 \\ 0 & \text{se } \tau > 1 \end{cases} \quad (27)$$

Noto che l'autocovarianza non dipende mai da valori di t , ma solo da τ, σ^2, a_1 .

Calcolo dell'autocorrelazione:

$$\rho_\tau = \begin{cases} 1 & \text{se } \tau = 0 \\ a_1 & \text{se } \tau = 1 \\ 0 & \text{se } \tau > 1 \end{cases}$$

Si vede quindi come MA(1) sia un processo con memoria finita e limitata al primo ritardo.

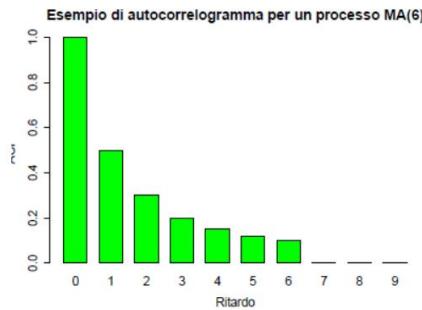
Stimatori per medie mobili generiche

Si noti che il processo MA è una somma pesata di *white noise*

$$\mu = 0$$

$$\mathbf{V}(Y_t) = \sigma^2(1 + a_1^2 + \dots + a_q^2) = \sigma^2 \sum_{j=1}^q a_j^2, \text{ dove } a_0 = 1$$

La soluzione si ricava facilmente dal caso MA(1). Per quel che riguarda la covarianza, è difficile trovarne una forma semplice: si vede quindi, ragionando in base al caso MA(1), che essa avrà valori positivi fino al q -esimo ritardo, dopodiché nulli, come nella figura sotto:



Il modello MA(∞)

Non è più una somma, ma una serie:

$$\{y_t\}_t \sim \text{MA}(\infty) \Leftrightarrow y_t = \sum_{j \geq 0} a_j \varepsilon_{t-j}; \text{ con } a_0 = 1 \quad (28)$$

È un processo a memoria infinita; inoltre la media è nulla. La varianza invece è $\sigma^2 \sum_{j \geq 0} a_j^2$, che è convergente solamente se impongo la quadrato sommabilità sulla successione dei coefficienti a_j , solo in tal caso MA(∞) è p.s.s.; poiché gli a_j sono quadrato sommabili, la successione è quindi monotona decrescente (proprietà delle serie) e quindi posso dire che il correlogramma tende a 0.

Il processo a media mobile con notazione LAG

Di conseguenza possiamo scrivere $Y \sim \text{MA}(q)$ come:

$$y_t = L^0 \varepsilon_t + a_1 L^1 \varepsilon_t + \dots + a_q L^q \varepsilon_t = \varepsilon_t \sum_{j=0}^q L^j a_j$$

Definisco quindi $a_q(L) = a_0 L^0 + a_1 L^1 + a_2 L^2 + \dots + a_q L^q$. Sostituisco nell'equazione:

$$y_t = a_q(L) \varepsilon_t$$

Un $\text{MA}(\infty)$ si può scrivere come $y_t = a_\infty(L) \varepsilon_t$.

Il teorema di Wold (Notazione ridotta)

Tutti i p.s.s. deboli sono rappresentabili come un processo $\text{MA}(\infty)$.

$$y_t = \sum_{j \geq 1} a_j \varepsilon_{t-j}, \quad a_0 = 1 \vee \sum_{j \geq 1} a_j \in \mathbb{R} \quad (29)$$

Ad esempio, un $\text{MA}(q)$ si può ottenere considerando tutti gli $a_j = 0, j > q$.

Il modello Autoregressivo (AR)

Un processo stocastico è autoregressivo di ordine p se:

$$\begin{aligned} \{y_t\}_t \sim \text{AR}(p) \iff & y_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t \\ & = \sum_{i=1}^p \theta_i y_{t-i} + \varepsilon_t; \text{ con } \varepsilon_t \sim \text{WN}(\sigma^2) \end{aligned} \quad (30)$$

Il valore attuale è quindi visto come combinazione lineare dei valori osservati nel passato più uno shock contemporaneo.

Il modello AR(1)

Suppongo di avere la successione di istanti $0, \dots, t$. Secondo la (29), $y_h = \theta y_{h-1} + \varepsilon_h$, $\forall h \in \{1, \dots, t\}$, di conseguenza posso, a partire dal primo valore della serie storica, risalire a tutti gli altri:

$$\begin{aligned}
 y_1 &= \theta y_0 + \varepsilon_1 \\
 y_2 &= \theta y_1 + \varepsilon_2 = (\text{sostituisco } y_1) \theta^2 y_0 + \theta \varepsilon_1 + \varepsilon_2 \\
 &\vdots \\
 y_t &= \theta^t y_0 + \sum_{i=1}^{t-1} \theta^i \varepsilon_{t-i} \\
 \text{usando la notazione LAG } &= \theta^t y_0 + \sum_{i=1}^{t-1} \theta^i L^i \varepsilon_t
 \end{aligned}$$

Quindi ho $\mu_t = \mathbf{E}(y_t) = \theta^t \mathbf{E}(y_0) = \theta^t \mu_0$ perché la parte degli *shock* è composta solo da coefficienti e *white noise*; tuttavia si vede che $\mu_h \neq \mu_l, \forall h \neq l$, di conseguenza l'AR(1) non è un p.s.s. a meno che non ponga dei valori particolari per μ_0 e σ_0^2 !!

La formula generica per la covarianza è:

$$\gamma_{t,t-\tau} = \theta^{2t-\tau} \cdot \mathbf{V}(y_0) + \theta^{-\tau} \sigma^2 \cdot \frac{1 - \theta^{2t}}{1 - \theta^2}$$

Si notano i seguenti risultati:

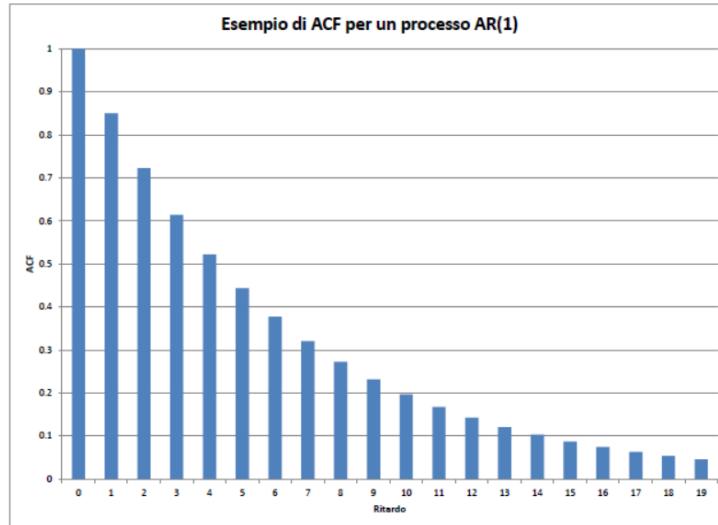
$$\lim_{t \rightarrow +\infty} \mu_t = \lim_{t \rightarrow +\infty} \mu_0 = \begin{cases} 0 \text{ se } |\theta| < 1 \\ +\infty \text{ se } |\theta| \geq 1 \end{cases} \quad (31)$$

$$\lim_{t \rightarrow +\infty} \gamma_{t,t-\tau} = \lim_{t \rightarrow +\infty} \gamma_{t,t} \quad (\text{perché } t \text{ grande}) = \lim_{t \rightarrow +\infty} \mathbf{V}(Y_t) = \frac{\sigma^2}{1 - \theta^2} \quad (32)$$

Di conseguenza si può notare che la stazionarietà del processo è *asintotica* per $|\theta| < 1$ dato che, in questo caso, i valori di media e varianza tendono a valori singoli all'aumentare di t . Nella teoria classica, la stazionarietà viene forzata obbligando media e varianza a assumere i valori limite indicati in (30) e (31): si nota, infatti, che se $\mu = 0$ e $\sigma_y^2 = \sigma^2/(1 - \theta^2)$ allora il processo autoregressivo diviene stazionario.

Se $\theta = 1$ allora il processo diventa $y_t = y_{t-1} + \varepsilon_t$ ed è detto **random walk** ed è tale che il valore attuale sia il risultato del valore precedente più un errore *white noise*. È un processo che riassume gli andamenti dei prezzi giornalieri in borsa.

Tornando agli AR con $\theta < 1$, sì dimostra che:



Il passaggio da AR(1) a MA(∞)

Ricordando la (??), mi pongo nell'ottica di avere un t crescente, di conseguenza posso effettuare la seguente approssimazione:

$$(t \rightarrow +\infty) \quad \theta^t y_0 \sim \theta^t \sim \theta^t \varepsilon_0, \text{ dato che } \varepsilon_0 \in \mathbb{R}, \theta \in]0; 1[$$

Stima dei parametri per Modelli ARMA

Stime nel modello AR

Per semplicità viene trattato solamente il caso del modello AR(1), ovvero $y_t = \theta y_{t-1} + \varepsilon_t$. Per i modelli AR ci si avvale degli stimatori dei minimi quadrati (LS), ma è necessario che vengano soddisfatte due condizioni:

- i. $\varepsilon_t \sim WN$
- ii. La covarianza fra l'errore e le variabili esplicative dev'essere nulla: nel nostro caso, deve accadere che $\text{COV}(\varepsilon_t; y_{t-1}) = 0$

La prima è rispettata dalle ipotesi del modello AR; la seconda si dimostra rapidamente, basti ricordare che un processo AR invertibile si può scrivere come processo MA:

$$y_t = \sum_{j \geq 0} \theta^j \varepsilon_{t-j}, \text{ di conseguenza, } y_{t-1} = \sum_{j \geq 0} \theta^j \varepsilon_{t-j-1}$$

Quindi posso scrivere

$$\text{COV}(\varepsilon_t; y_{t-1}) = \text{COV}(\varepsilon_t; \varepsilon_{t-1} + \varepsilon_{t-2} + \dots) = 0$$

provenendo gli errori da un processo white noise.

Abbiamo quindi uno stimatore θ_{LS} che è consistente, ovvero $\theta_{LS} \xrightarrow{P} \theta$. Nel caso avessimo incluso una costante nel modello, anche lo stimatore della costante sarebbe consistente.

Nota: per calcolare gli stimatori dei minimi quadrati devo minimizzare la quantità $\sum \varepsilon_t^2 = \sum (y_t - \theta y_{t-1})^2$. Io avrei a disposizione T dati su cui ho osservato i valori, ma poiché all'interno dell'equazione ho sia y_t che y_{t-1} , allora devo rinunciare al primo dato osservato. Nel caso in cui stessi stimando un modello AR di ordine p , dovrei rinunciare ai primi p dati. Mentre il campione *originario* contiene T dati, il campione c.d. *di stima* ha una numerosità pari a $T - p$. Quindi avrò:

$$\min_{\theta} \left(\sum_{t=p+1}^T \varepsilon_t^2 \right)$$

per un modello AR di ordine p .

Per quel che riguarda la correttezza di θ_{LS} , devo dimostrare che $\mathbf{E}(\varepsilon_t | \underline{y}) = 0$, ovvero che i residui non siano correlati con la matrice delle osservazioni. Dire ciò equivale a dire che tutte le osservazioni y_t siano incorrelate con il residuo ε_t , e ciò ripetuto su ogni t del campione di stima. Poiché l'equazione dei modelli autoregressivi impongono che la ε_t abbia un certo peso sulla determinazione del valore di y_t , allora vale $\mathbf{E}(\varepsilon_t | \underline{y}) \neq 0$, quindi si ha che θ_{LS} è uno stimatore **distorto**.

Rimanendo nel caso AR(1), se aggiungo l'ipotesi di normalità ai residui, posso calcolare gli stimatori di **massima verosimiglianza** (ML). Devo prima di tutto poter trovare la funzione di verosimiglianza a partire dalla densità marginale $f(y_1, \dots, y_T)$, ma è difficile trovarne la forma analitica. Posso quindi scrivere

$$f(y_1, \dots, y_T) = f(y_1) \cdot f(y_2, \dots, y_T | y_1)$$

isolando così il primo dato, appartenente al cosiddetto *precampione*, ovvero quel dato che non appartiene al campione di stima. Dopodiché avrò:

$$f(y_1, \dots, y_T) = f(y_1) \cdot f(y_2, \dots, y_T | y_1) = f(y_1) \prod_{t=2}^T f(y_t | y_{t-1}, \dots, y_1) = f(y_1) \cdot f(y_t | y_{t-1}) = f(y_1) \cdot N(\theta y_{t-1}; \sigma_\varepsilon^2)$$

Posso quindi decidere di ricavare la funzione di verosimiglianza da tutta la densità, oppure utilizzando solo la seconda parte (la normale); così facendo, ottengo delle stime ML condizionali, che coincidono con le stime LS; altrimenti tale equivalenza non vale e devo massimizzare la verosimiglianza (che sarà quindi una verosimiglianza “esatta”) tramite metodi iterativi non analitici. Si dimostra che $\hat{\theta}_{MLCOND} \xrightarrow{P} \hat{\theta}_{ML}$.

Stime nel modello MA

Per il modello MA è impossibile affidarsi alle stime LS, infatti, nel modello MA(1):

$$\sum \varepsilon_t^2 = \sum (y_t - a\varepsilon_{t-1})^2 \quad (34)$$

Di conseguenza dovremmo minimizzare una quantità di cui non conosciamo tutti i valori, poiché contiene i residui (che sono per definizione valori latenti) al suo interno.

La soluzione si ha passando da MA(1) a AR(∞): infatti possiamo scrivere

$$\begin{aligned} y_t &= \varepsilon_t + a\varepsilon_{t-1} \\ &= a_1(L)\varepsilon_t \text{ usando l'operatore ritardo} \\ \varepsilon_t &= a_1^{-1}(L)(y_t) \\ \text{di conseguenza } \varepsilon_{t-1} &= a_1^{-1}(L)(y_{t-1}) \end{aligned}$$

Posso quindi sostituire la forma appena trovata all'interno della 33 ottenendo così una equazione del tipo

$$\sum_{t=2}^T \varepsilon_t^2 = \sum (y_t - a(a_1^{-1}(L)(y_{t-1})))^2 = \sum_t \left(y_t - a \left(\sum_{j \geq 0} (-a)^j y_{t-j} \right) \right)^2$$

Devo ridurre la serie ai dati osservati

$$= \sum_{t=2}^T \left(y_t - a \left(\sum_{j=0}^{T-2} (-a)^j y_{t-j} \right) \right)^2$$

A questo punto devo minimizzare questa quantità; tuttavia, poiché a appare in diverse forme (non solo lineari), otterrò delle stime a minimi quadrati non lineari; inoltre, la minimizzazione deve avvenire con metodi numerici e non analitici. Come nell'AR, lo stimatore è distorto ma consistente; tuttavia, più T è grande, più la serie viene approssimata con precisione e più lo stimatore è corretto. Per le stime del modello ARMA vengono prese le stesse decisioni finora evidenziate e si ottengono, quindi degli stimatori dei minimi quadrati non lineari distorti e consistenti.

Selezione dell'ordine del modello ARMA con criteri informativi

Sono dei criteri di tipo statistico che danno un risultato numerico con scopi simili a quello di R² corretto, al fine di offrire uno strumento di scelta fra modelli parsimoniosi e sovraparametrati senza perdere troppa informazione. I criteri sono due:

1. Criterio di Akaike: $AIC = \ln(\hat{\sigma}_{ML}^2) + 2\frac{p+q+1}{T^*}$, dove T^* è la numerosità del campione di stima.
2. Criterio di Schwartz: $SIC = \ln(\hat{\sigma}_{ML}^2) + 2 \ln(T^*) \frac{p+q+1}{T^*}$.

Il criterio di Schwartz è ritenuto essere migliore in quanto tende a privilegiare modelli più parsimoniosi.

In entrambi i casi si procede come di seguito:

- a) Fisso un \bar{p} ed un \bar{q} che rappresentano il massimo di parametri AR e MA che voglio ottenere;
- b) Parto da ARMA(0,0) e calcolo AIC/SIC;
- c) Aggiungo 1 al parametro AR e calcolo AIC/SIC; procedo così finchè raggiungo \bar{p} ;
- d) Reitero il passaggio precedente con le componenti MA, fino a raggiungere \bar{q} ;
- e) Seleziono il modello con AIC/SIC minimo.

Si noti che AIC/SIC viene confrontato fra modelli non equi: infatti, ogni volta che aggiungo un parametro, il campione di stima diminuisce: a tal scopo, è necessario predisporre di un campione di stima abbastanza numeroso al fine di evitare la non rappresentatività del campione stesso e quindi modelli con stime molto distorte, soprattutto in modelli con molti parametri. *Nota: nei modelli ARMA $T^* = T - \max\{p, q\}$.*

Stazionarietà ed Invertibilità per processi ARMA

9.6 Stazionarietà ed Invertibilità per processi ARMA

Dato il processo ARMA(p, q)

$$y_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \epsilon_t + a_1 \epsilon_{t-1} + \dots + a_q \epsilon_{t-q}$$

posso riscriverlo usando la notazione lag:

$$\theta_p(L) = a_q(L) + \epsilon_t$$

L'approccio di Box e Jenkins per la selezione del modello

Si basa su 4 fasi:

- I. Rendere la serie stazionaria (se è stazionaria in media tramite differenziazione, se è stazionaria in varianza tramite trasformazioni logaritmiche o radicali)
- II. Identificazione delle componenti ARMA, tramite il confronto dei correlogrammi oppure tramite criteri informativi (vedi sezione 9.5)
- III. Controllo della bontà del modello (test vari ed analisi dei residui)
- IV. Previsioni

Le previsioni sfruttando il modello

Definizione: si dice set informativo all'istante t l'insieme delle informazioni a disposizione da un istante temporale t fino all'infinito passato: $\mathcal{I}_t = \{y_t, y_{t-1}, \dots\}$.

Si dimostra che, se un processo è invertibile, $\mathcal{I}_t = \mathcal{F}_t \triangleq \{\varepsilon_t, \varepsilon_{t-1}, \dots\}$.

Definizione: si dice previsione statica una previsione effettuata aggiornando il set informativo ogni istante successivo; una previsione dinamica non aggiorna il set informativo e, per stimare i momenti successivi al primo, utilizza le previsioni stesse per gli istanti precedenti.

Definizione: si dice processo a differenze di Martingale un processo tale che:

- a. $E(|X_t|) \in \mathbb{R}, \forall t$, ovvero è un processo a media finita.
- b. $E(X_t | \mathcal{I}_{t-1}) = 0, \forall t$, ovvero è un processo non prevedibile dal passato.
- c. $V(X_t) = c, \forall t$, ovvero è un processo omoschedastico.

Si dimostra che un processo white noise è anche un processo a differenze di Martingale.

Definizione: $\hat{y}_{T+k|T}$ è detto previsore per l'istante $T + k$ dato il set informativo \mathcal{I}_T .
[$\hat{y}_{T+k|T} = g(\mathcal{I}_t) \wedge [V(\hat{y}_{T+k|T}) < +\infty]$: il previsore è funzione del set informativo ed è a varianza finita.]

La scelta del previsore avviene tramite minimizzazione:

$$\hat{y}_{T+k|T} : \min E[(y_{T+k} - \hat{y}_{T+k|T})^2 | \mathcal{I}_t] = \min MSE[\hat{y}_{T+k|T}]$$

In pratica il previsore *ottimale* minimizza l'errore quadratico medio (Mean Square Error - MSE) fra se stesso ed il valore reale della serie nell'istante $T + k$.

Si dimostra che lo stimatore funzione di regressione ($E(y_t + k | \mathcal{I}_T)$) rispetta tale caratteristica.

Previsioni con il modello autoregressivo

Ammetto, per la serie Y_{tt} , di aver considerato il modello AR(1) come valido. Voglio quindi passare alle previsioni del modello sfruttando il set \mathcal{I}_t ed effettuando una previsione dinamica. In base al risultato qui sopra ottenuto, ho che:

$$\hat{y}_{T+1|T} = \mathbf{E}[y_{T+1}|\mathcal{I}_t] = \mathbf{E}[\theta y_T + \varepsilon_{T+1}] = \theta \mathbf{E}(y_T) = \theta y_T$$

visto che $\mathbf{E}(Y_T) = y_T$ essendo un valore osservato.

Posso ripetere a catena il risultato ora trovato, ma cercando $\hat{y}_{T+2|T+1}$. Trovo

$$\hat{y}_{T+2|T} = \theta \hat{y}_{T+1} = \theta^2 y_T$$

A partire da questo risultato, risalgo al generico previsore:

$$\hat{y}_{T+k|T} = \theta^k y_T, \forall k > 0 \quad (35)$$

In più l'ipotesi di stazionarietà $|\theta| < 1$ implica che

$$\theta^k y_T = \hat{y}_{T+k|T} \rightarrow 0, (k \rightarrow +\infty)$$

Inoltre, si dimostra, passando al processo MA(∞) che

$$\lim_{k \rightarrow +\infty} \mathbf{V}[(y_{t+k} - \hat{y}_{T+k})|\mathcal{I}_T] = \mathbf{V}(\mathcal{I}_T) = \frac{\sigma_\varepsilon^2}{1 - \theta^2}$$

Nelle previsioni per un modello stimato, dovrò utilizzare $\hat{\theta}$ al posto di θ .

Per il processo AR(2), si dimostra facilmente che vale

$$\hat{y}_{T+1|T} = \theta_1 y_T + \theta_2 y_{T-1}$$

Per il momento successivo avrò

$$\hat{y}_{T+2|T} = \theta_1 \hat{y}_{T+1} + \theta_2 y_T = \theta_1(\theta_1 y_T + \theta_2 y_{T-1}) + \theta_2 y_T = y_T(\theta_1^2 + \theta_2) + \theta_2 y_{T-1}$$

Man mano che avanzo di momenti avrò polinomi di grado k sempre più complessi.

Previsioni con il modello a media mobile

Ho il modello $y_T = \varepsilon_T + a\varepsilon_{T-1}$. Ricavo quindi il previsore per il primo momento:

$$\hat{y}_{T+1|T} = \mathbf{E}(\varepsilon_{T+1}|\mathcal{I}_T) + a\mathbf{E}(\varepsilon_T|\mathcal{I}_T)$$

Ora, poiché gli errori sono un processo a differenze di Martingale, $\mathbf{E}(\varepsilon_{T+1}|\mathcal{I}_T) = 0$. Quindi ottengo:

$$\hat{y}_{T+1|T} = a\mathbf{E}(\varepsilon_T|\mathcal{I}_T) = a\mathbf{E}(\varepsilon_T|\mathcal{F}_T) = a\varepsilon_T$$

Passando al momento successivo:

$$\hat{y}_{T+2|T} = \mathbf{E}(\varepsilon_{T+2} | \mathcal{I}_T) + a\mathbf{E}(\varepsilon_{T+1} | \mathcal{I}_T)$$

Poiché gli errori sono un processo a differenze di Martingale, entrambi i valori attesi sono nulli. Si verifica che vale lo stesso per tutti i momenti successivi.

Quindi

$$\hat{y}_{T+k|T} = \begin{cases} a\varepsilon_T, & \text{se } k = 1 \\ 0, & \text{se } k > 1 \end{cases} \quad (36)$$

Quindi, la previsione si basa su un solo valore. Nelle previsioni reali userò \hat{a} al posto di a e e_T al posto di ε_T .

9.8.3 Previsioni con il modello ARMA

Per i modelli ARMA(1,1), si dimostra facilmente che vale

$$\hat{y}_{T+k|T} = \theta^k + \theta^k y_T + \theta^{k-1} a\varepsilon_T \quad (37)$$

Il Q-test di Ljung e Box per la correlazione dei residui

Definita $\rho_\varepsilon(\tau)$ la funzione di autocorrelazione dei residui al ritardo τ , il test ha come ipotesi

$$\begin{cases} H_0 : \rho_\varepsilon(1) = \dots = \rho_\varepsilon(K) = 0 \\ H_1 : \exists \bar{k} : \rho_\varepsilon(\bar{k}) \neq 0 \end{cases}$$

Definisce una statistica $Q(k) \stackrel{\circ}{\sim} \chi^2_{k-p-q-1}$. Ci possono essere problemi con tanti parametri stimati, quindi per valori bassi del ritardo, di norma la statistica Q non viene fornita.

Per valori alti di Q rifiuto l'ipotesi per il ritardo k .

Test dei moltiplicatori di Lagrange

Funziona solo per i modelli AR e può essere una valida alternativa al Q-test. Prevede una regressione ausiliaria:

$$\varepsilon_t = \xi_0 + \xi_1 y_{t-1} + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \dots + \omega_k \varepsilon_{t-k} + \epsilon_t$$

ad esempio, per un AR(1). Le ipotesi sono

$$\begin{cases} H_0 : \omega_1 = \dots = \omega_k = 0 \\ H_1 : \exists \bar{i} : \omega_{\bar{i}} \neq 0 \end{cases}$$

La statistica test è sempre $T \cdot R^2 \stackrel{\circ}{\sim} \chi^2_k$.

2.8 Cenni sui processi non stazionari: il modello di passeggiata aleatoria (Random Walk).

2.9 Esercitazioni al computer

Capitolo 3 Modelli econometrici multivariati per serie storiche economiche stazionarie.

3.1 Il modello di regressione lineare dinamico multivariato.

3.2 Il modello autoregressivo vettoriale (VAR): specificazione, stima, test di corretta specificazione del modello ed inferenza sui parametri;

Modelli Autoregressivi Vettoriali (VAR)

Rispondono alla necessità di prevedere il futuro combinando più serie storiche. la variabile risposta si presenta sotto forma di vettore:

$$\underline{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{pmatrix}$$

Avrò quindi una matrice di coefficienti:

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{p1} & \beta_{m2} & \dots & \beta_{pm} \end{bmatrix}$$

Ho quindi:

$$\underline{y}_t = B \cdot \underline{y}_{t-1} = \begin{cases} y_{1t} = \beta_{11}y_{1,t-1} + \dots + \beta_{1m}y_{m,t-1} \\ \vdots \\ y_{mt} = \beta_{p1}y_{1,t-1} + \dots + \beta_{pm}y_{m,t-1} \end{cases} \quad (38)$$

3.3 Test di causalità di Granger nei modelli VAR stazionari

3.4 Le previsioni multiperiodali

3.5 Funzione di risposta all'impulso (innovation accounting)

3.6 Cenni sui VAR strutturali.

3.7 Esercitazioni al computer.

Capitolo 4 Modelli per dati panel

4.1 Vantaggi e potenziali problemi nell'uso di dati Panel

4.2 Modelli lineari statici per dati panel: effetti fissi ed effetti casuali ed altre forme di eterogeneità

4.3 Lo stimatore Fixed effects (within group) e lo stimatore random effects

4.4 Regressori strettamente esogeni, debolmente esogeni ed endogeni; richiami sulla regressione con variabili strumentali.

4.5 Il metodo dei momenti ed il metodo dei momenti generalizzato (GMM)

4.6 Proprietà asintotiche dello stimatore GMM

4.7 Test sulle restrizioni di sovraidentificazione;

4.8 Applicazione ai modelli dinamici per dati panel;

4.9 Trasformazione alle differenze prime del modello e
stimatore GMM di Arellano e Bond.

4.10 Esercitazioni al computer.