

25/09

by EDL

MODelli ECONOMETRICI

CROSS- SECTION

NB. $\mu = 0 \approx m_0$

NON SEMPRE PERO'

OBETTIVO: stimare gli effetti causali. A patto che si possa avere endogenità dei regressori.

$$Y_i = \beta_0 + \beta_1 X_i + \nu_i$$

Regressore

- $\text{Cov}(X_i, \nu_i) \neq 0 \rightarrow X_i$ è un regressore endogeno
- $\text{Cov}(X_i, \nu_i) = 0 \rightarrow X_i$ è un regressore esogeno

$\hat{\beta}_0$ e $\hat{\beta}_1$ è distorto e inconsistente con gli OLS se il regressore è endogeno.

Es. farmaco nei patienti. Metodologie per garantire la causalità e isolare l'effetto del farmaco da altri effetti non misurabili. $X_i \perp \nu_i$.

$$N_i = \beta_2 z_{2i} + \beta_3 z_{3i} + V_i$$

$\neq 0$ $\neq 0$

Verificare di controllo

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 z_{2i} + \beta_3 z_{3i} + V_i$$

$$\begin{cases} \text{Cov}(X_i, z_{2i}) \neq 0 \\ \text{Cov}(X_i, z_{3i}) \neq 0 \end{cases}$$

$$\text{Cov}(X_i, V_i) = 0$$

Rendo l'errore e il regressore in correlati
così posso applicare gli OLS

Ci sono dei casi in cui non posso eliminare l'endogenità.

OBIEKTIVO: fare una previsione, in questo caso non mi interessa l'interpretazione dei coefficienti e di conseguente capire gli effetti causali o parziali. In questo caso posso usare modelli e pratica di serie storiche.

AR(1)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \quad t=1, \dots, T$$

^{WN}

→ Modello dinamico

Modello autoregressivo: perché regresso sul suo passato

PREVISIONE

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$$

VARC(1)

$$\begin{cases} Y_{1,t} = \delta_{1,0} + \beta_{1,1} + Y_{1,t-1} + \beta_{1,2} Y_{2,t-1} + \varepsilon_{1,t} \\ Y_{2,t} = \delta_{2,0} + \beta_{2,1} + Y_{1,t-1} + \beta_{2,2} Y_{2,t-1} + \varepsilon_{2,t} \end{cases}$$

Caso $m=2$

DATI PANEL

Osserviamo un insieme di unità statistiche al variare del tempo. Avremo quindi 2 indici.

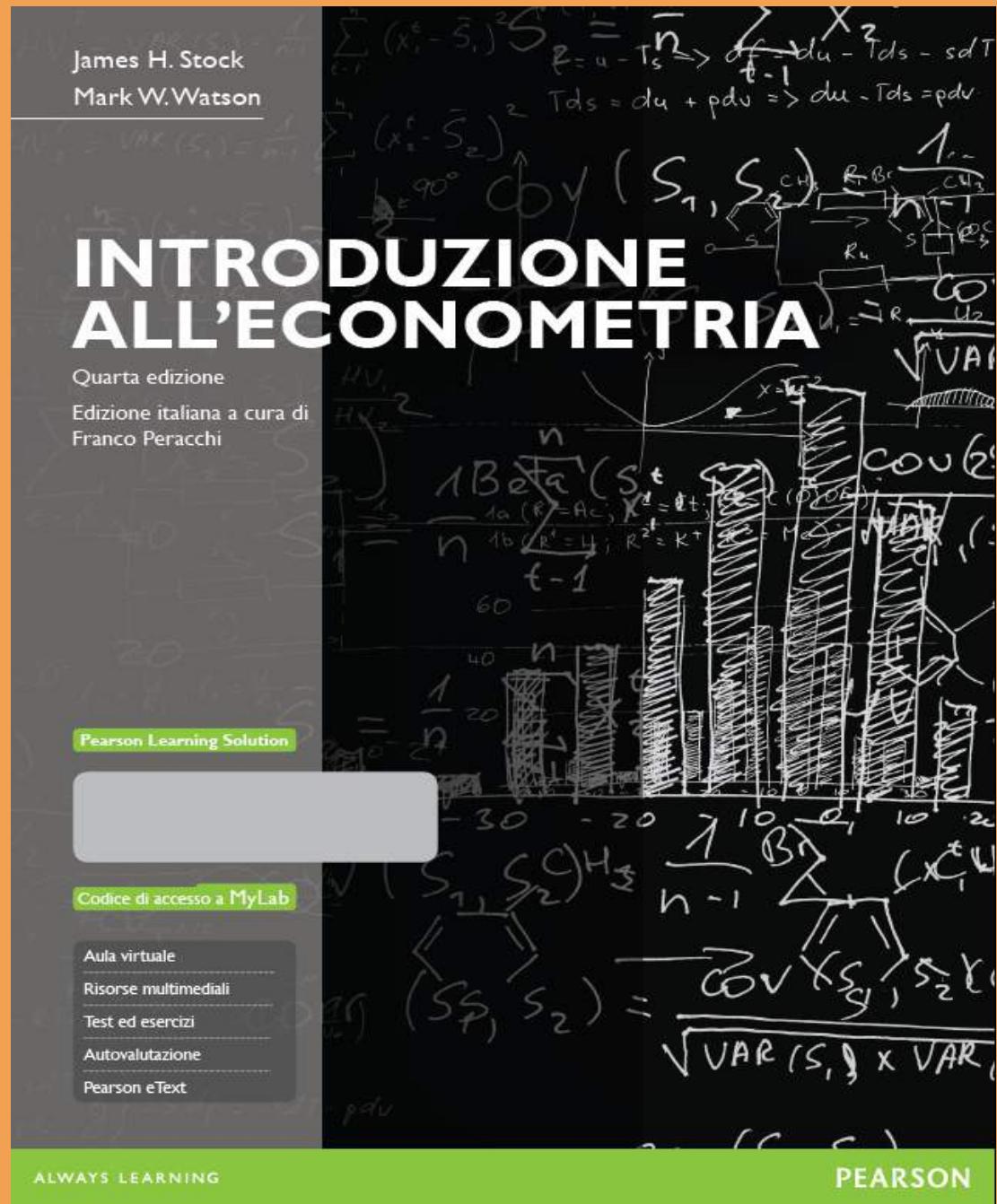
$$Y_{i,t} = \beta_0 + \beta_1 \cdot X_{i,t} + \mu_{i,t} = \alpha_i + \beta_1 X_{i,t} + \mu_{i,t}$$

\downarrow Effetto entità i-esima

In questi modelli c'è il bisogno di introdurre elementi dieterogenetà tra le differenti unità statistiche.

Capitolo 4

Regressione lineare con un singolo regressore



Sommario

1. Il modello di regressione lineare
2. Stima dei coefficienti del modello di regressione lineare
3. Misure di bontà dell'adattamento
4. Le assunzioni dei minimi quadrati
5. Distribuzione campionaria degli stimatori OLS
6. Conclusioni

Lo stimatore OLS (Paragrafo 4.2)

Come possiamo stimare β_0 e β_1 dai dati?

Si ricordi che lo stimatore OLS di μ_Y : \bar{Y} , è dato da

$$\min_m \sum_{i=1}^n (Y_i - m)^2$$

Per analogia, **ci concentreremo sullo stimatore dei minimi quadrati (OLS, “*ordinary least squares*”) dei parametri ignoti β_0 e β_1 .** Lo stimatore OLS è dato da

$$\min_{b_0, b_1} \sum_{i=1}^n [Y_i - (b_0 + b_1 X_i)]^2$$

Lo stimatore OLS:

$$\min_{b_0, b_1} \sum_{i=1}^n [Y_i - (b_0 + b_1 X_i)]^2$$

- Lo stimatore OLS minimizza la differenza quadratica media tra i valori reali di Y_i e la previsione ("valori predetti") basata sulla retta stimata.
- Questo problema di minimizzazione si può risolvere con il calcolo differenziale (App. 4.2).
- **Il risultato sono gli stimatori OLS di β_0 e β_1 .**

Gli stimatori OLS della pendenza β_1 e dell'intercetta β_0 sono:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2} \quad (4.7)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \quad (4.8)$$

I valori predetti \hat{Y}_i e i residui \hat{u}_i sono:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, \dots, n \quad (4.9)$$

$$\hat{u}_i = Y_i - \hat{Y}_i, \quad i = 1, \dots, n. \quad (4.10)$$

L'intercetta stimata ($\hat{\beta}_0$), la pendenza stimata ($\hat{\beta}_1$) e il residuo (\hat{u}_i) sono calcolati per un campione di n osservazioni di X_i e Y_i , con $i = 1, \dots, n$. Queste sono stime dell'intercetta β_0 , della pendenza β_1 e dell'errore u_i nella popolazione.

Misure di bontà dell'adattamento (Paragrafo 4.3)

Due statistiche di regressione forniscono misure complementari della bontà dell'adattamento della regressione ai dati:

- L' **R^2 della regressione** misura la frazione della varianza (devianza) di Y spiegata da X ; è priva di unità e può variare tra zero (nessun adattamento) e uno (adattamento perfetto)
- L'***errore standard della regressione (SER)*** misura la dimensione di un tipico residuo di regressione nelle unità (di misura) di Y .

L' R^2 della regressione è la frazione della varianza campionaria di Y_i “spiegata” dalla regressione.

$$Y_i = \hat{Y}_i + \hat{u}_i = \text{stima OLS} + \text{residuo OLS}$$

$$\rightarrow \text{var camp. } (Y) = \text{var camp. } (\hat{Y}_i) + \text{var camp. } (\hat{u}_i) \text{ (perché?)}$$

$$\rightarrow \text{somma dei quadrati(SS) Totali} = \text{SS “spiegata”} + \text{SS “residua”}$$

Definizione di R^2 :

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

- $R^2 = 0$ significa $ESS = 0$
- $R^2 = 1$ significa $ESS = TSS$
- $0 \leq R^2 \leq 1$
- Per la regressione con una singola X , $R^2 =$ il quadrato del coefficiente di correlazione tra X e Y

L'errore standard della regressione (SER)

Il SER misura la dispersione della distribuzione dei residui. È (quasi) la deviazione standard campionaria dei residui OLS:

$$\begin{aligned} SER &= \sqrt{\frac{1}{n-2} \sum_{i=1}^n (\hat{u}_i - \bar{\hat{u}})^2} \\ &= \sqrt{\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2} \end{aligned}$$

La seconda uguaglianza vale perché $\bar{\hat{u}} = \frac{1}{n} \sum_{i=1}^n \hat{u}_i = 0$.

$$SER = \sqrt{\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2}$$

Il *SER*:

ha le unità di u , che sono le unità di Y

misura la “dimensione” tipica del residuo OLS (l’“errore” tipico della retta di regressione OLS)

La ***radice dell’errore quadratico medio*** (*RMSE*, *Root Mean Squared Error*) è strettamente legata al *SER*:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{u}_i^2}$$

Misura la stessa cosa del *SER* – la differenza sta nel fattore $1/n$ anziché $1/(n-2)$.

Nota tecnica: perché dividere per $n-2$ anziché per $n-1$?

$$SER = \sqrt{\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2}$$

- La divisione per $n-2$ è una correzione “dei gradi di libertà” – esattamente come la divisione per $n-1$, con la differenza che per il SER sono stati stimati due parametri (β_0 e β_1 , da $\hat{\beta}_0$ e $\hat{\beta}_1$), mentre in s_Y^2 ne è stato stimato solo uno (μ_Y , da \bar{Y}).
- Quando n è grande non importa se si utilizzi n , $n-1$ o $n-2$ – anche se la formula convenzionale utilizza $n-2$ quando c’è un singolo regressore.
- Per i dettagli, cfr. il Paragrafo 17.4

Le assunzioni dei minimi quadrati

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n$$

1. La distribuzione di u condizionata a X ha media nulla, cioè $E(u|X = x) = 0$ per ogni x .
 - Questo implica, insieme all'ass. 2, che $\hat{\beta}_1$ è non distorto
2. $(X_i, Y_i), i = 1, \dots, n$, sono i.i.d.
 - Questo è vero se (X, Y) sono ottenuti mediante campionamento casuale semplice
 - Questo fornisce la distribuzione camp. di $\hat{\beta}_0$ e $\hat{\beta}_1$ in grandi campioni
3. Gli outlier in X e/o Y sono rari.
 - Tecnicamente, X e Y hanno momenti quarti finiti
 - Gli outlier possono risultare in valori privi di senso di $\hat{\beta}_1$

Assunzione dei minimi quadrati n. 2: $(X_i, Y_i), i = 1, \dots, n$ sono i.i.d.

Questo si verifica automaticamente se l'unità (individuo, distretto) è campionata mediante campionamento casuale semplice:

- Le unità sono scelte dalla stessa popolazione, perciò (X_i, Y_i) sono *identicamente distribuite* per ogni $i = 1, \dots, n$.
- Le unità sono scelte a caso, perciò i valori di (X, Y) per unità diverse sono *indipendentemente distribuite*.

I campionamenti non i.i.d. si incontrano principalmente quando si registrano dati nel tempo per la stessa unità (dati panel e serie temporali) – affronteremo tale complicazione quando tratteremo i dati panel.

Media e varianza della distribuzione campionaria di $\hat{\beta}_1$

Un po' di algebra preliminare: In termini matriciali:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad Y = X \beta + u$$

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$$

perciò $Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + (u_i - \bar{u})$

Quindi

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\begin{aligned}\hat{\beta}_1 &= (X^\top X)^{-1} X^\top Y \\ &= (X^\top X)^{-1} X^\top [X \beta + u] \\ &\stackrel{\text{new page}}{=} \frac{\sum_{i=1}^n (X_i - \bar{X})[\beta_1(X_i - \bar{X}) + (u_i - \bar{u})]}{\sum_{i=1}^n (X_i - \bar{X})^2}\end{aligned}$$

$$\hat{\beta}_1 = \beta_1 \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$


 $\hat{\beta} = (\cancel{X^\top X})^{-1} \cancel{X^\top} \beta + (\cancel{X^\top X})^{-1} \cancel{X^\top} u$
 $\hat{\beta} - \beta = (\cancel{X^\top X})^{-1} \cancel{X^\top} u$

perciò $\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2}$

Ora $\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^n (X_i - \bar{X})u_i - \left[\sum_{i=1}^n (X_i - \bar{X}) \right] \bar{u}$

$$= \sum_{i=1}^n (X_i - \bar{X})u_i - \left[\left(\sum_{i=1}^n X_i \right) - n\bar{X} \right] \bar{u}$$

$$= \sum_{i=1}^n (X_i - \bar{X})u_i$$

Sostituiamo $\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^n (X_i - \bar{X})u_i$ nella
espressione per $\hat{\beta}_1 - \beta_1$:

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

perciò

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Ora possiamo calcolare $E(\hat{\beta}_1)$ e $\text{var}(\hat{\beta}_1)$:

$$\begin{aligned} E(\hat{\beta}_1) - \beta_1 &= E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})u_i}{\sum_{i=1}^n (X_i - \bar{X})^2}\right] \\ &= E\left\{E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})u_i}{\sum_{i=1}^n (X_i - \bar{X})^2} \middle| X_1, \dots, X_n\right]\right\} \end{aligned}$$

Velore atteso iterato

$= 0$ poiché $E(u_i | X_i = x) = 0$ per l'assunzione OLS 1 (e grazie all'ass. 2)

- Quindi l'assunzione 1 e 2 implicano insieme che $E(\hat{\beta}_1) = \beta_1$
- Cioè $\hat{\beta}_1$ è **uno stimatore non distorto di β_1** .
- Per i dettagli cfr. Appendice 4.3

Ora calcoliamo $\text{var}(\hat{\beta}_1)$:

scriviamo

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\frac{1}{n} \sum_{i=1}^n v_i}{\left(\frac{n-1}{n}\right) s_X^2}$$

$\text{var}(\hat{\beta}_1) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} s_X^2$

$\text{var}(\hat{\beta}_1) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} s_X^2$

dove $v_i = (X_i - \bar{X}) u_i$. Se n è grande, $s_X^2 \approx \sigma_X^2$, $\frac{n-1}{n} \approx 1$, e
 $\bar{X} \approx \mu_X$, perciò

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n} \sum_{i=1}^n v_i}{\sigma_X^2},$$

dove $v_i = (X_i - \mu_X) u_i$ (cfr. Appendice 4.3). Quindi

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n} \sum_{i=1}^n v_i}{\sigma_x^2} \rightarrow \text{Var}(\hat{\beta}_1) + \text{Var}(\beta_1) + \text{cov}(\dots)$$

perciò $\text{var}(\hat{\beta}_1 - \beta_1) = \text{var}(\hat{\beta}_1)$

$$= \text{var}\left(\frac{1}{n} \sum_{i=1}^n v_i\right) / (\sigma_x^2)^2 = \frac{\text{var}(v_i) / n}{(\sigma_x^2)^2}$$

dove l'uguaglianza finale usa l'assunzione 2. Quindi

$$\text{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\text{var}[(X_i - \mu_x)u_i]}{(\sigma_x^2)^2}$$

Riepilogo

- $\hat{\beta}_1$ è non distorto: $E(\hat{\beta}_1) = \beta_1$ – proprio come \bar{Y} !
- $\text{Var}(\hat{\beta}_1)$ è inversamente proporzionale a n – proprio come \bar{Y} !

Qual è la distribuzione campionaria di $\hat{\beta}_1$?

Determinare la distribuzione campionaria esatta è complicato – dipende dalla distribuzione di (Y, X) – ma quando n è grande otteniamo alcune buone (e semplici) approssimazioni:

- 1) Poiché $\text{var}(\hat{\beta}_1) \propto 1/n$ e $E(\hat{\beta}_1) = \beta_1$, $\hat{\beta}_1 \xrightarrow{P} \beta_1$
- 2) Quando n è grande, la distribuzione campionaria di $\hat{\beta}_1$ è ben approssimata da una distribuzione normale (TLC)

Ricordiamo il TLC: sia $\{v_i\}$, $i = 1, \dots, n$ i.i.d. con

$E(v) = 0$ e $\text{var}(v) = \sigma^2$ finita. Allora, quando n è grande, la distribuzione di $\frac{1}{n} \sum_{i=1}^n v_i$ è approssimata da $N(0, \sigma_v^2 / n)$.

Approssimazione per n grande della distribuzione di $\hat{\beta}_1$:

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n} \sum_{i=1}^n v_i}{\left(\frac{n-1}{n}\right) s_X^2} \approx \frac{\frac{1}{n} \sum_{i=1}^n v_i}{\sigma_X^2}, \text{ dove } v_i = (X_i - \bar{X})u_i$$

- Quando n è grande, $v_i = (X_i - \bar{X})u_i \approx (X_i - \mu_X)u_i$, che è i.i.d. (*perché?*) e $\text{var}(v_i) < \infty$ (*perché?*). Perciò, per il TLC, la distribuzione di $\frac{1}{n} \sum_{i=1}^n v_i$ è approssimata da $N(0, \sigma_v^2 / n)$.
- Quindi, per n grande, la distribuzione di $\hat{\beta}_1$ è approssimata da

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_v^2}{n(\sigma_X^2)^2}\right), \text{ dove } v_i = (X_i - \mu_X)u_i$$

Maggiore è la varianza di X , minore è la varianza di $\hat{\beta}_1$

Calcoli

$$\text{var}(\hat{\beta}_1 - \beta_1) = \frac{1}{n} \times \frac{\text{var}[(X_i - \mu_x)u_i]}{(\sigma_x^2)^2}$$

Dove $\sigma_x^2 = \text{var}(X_i)$. La varianza di X appare (al quadrato) al denominatore – perciò aumentando la dispersione di X diminuisce la varianza di $\hat{\beta}_1$.

Ragionamento intuitivo

Se vi è più variazione in X , allora vi sono più informazioni nei dati che si possono utilizzare per l'adattamento della retta di regressione. Lo si vede meglio in una figura...

Riepilogo della distribuzione campionaria di $\hat{\beta}_1$:

Se valgono le tre assunzioni dei minimi quadrati, allora

- La distribuzione campionaria esatta (campione finito) di $\hat{\beta}_1$ ha:

- $E(\hat{\beta}_1) = \beta_1$ (cioè $\hat{\beta}_1$ è non distorto)

- $\text{var}(\hat{\beta}_1) = \frac{1}{n} \times \frac{\text{var}[(X_i - \mu_x)u_i]}{\sigma_x^4} \propto \frac{1}{n}$.

- A parte media e varianza, la distribuzione esatta di $\hat{\beta}_1$ è complessa e dipende dalla distribuzione di (X, u)

- $\hat{\beta}_1 \xrightarrow{P} \beta_1$ (cioè $\hat{\beta}_1$ è consistente) *Qte Pisot*

- Quando n è grande,
$$\frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{\text{var}(\hat{\beta}_1)}} \stackrel{\text{CLT}}{\sim} N(0,1)$$

- Segue in parallelo la distribuzione campionaria di \bar{Y} .

26/09

$$Y_i = \beta_0 + \beta_1 X_i + \nu_i \quad i=1, \dots, n$$

Proprietà dei momenti degli errori

$$1) E[\nu_i | X_i = x] = 0, \quad \forall x$$

\hookrightarrow Valore certo perché conosco il valore delle x

$$\text{sse } E[\nu_i | X_i] = g(X_i) = 0 \quad \begin{matrix} \nearrow \text{Variabile casuale} \\ \searrow \text{Costante, non varia al variare di } X \end{matrix}$$

q.c. \rightarrow con prob. 1

\hookrightarrow f.d. di regressione ν_i rispetto a X_i .

$$E[Z|W] = g(W) \quad \begin{matrix} \nearrow \text{t.c.} \\ \searrow \text{b.v.c.} \end{matrix} \quad W=w \quad g(w) = E[Z|W=w] + \nu_w$$

$$2) (Y_i, X_i) \sim \text{i.i.d. } f_{Y_i}$$

- Indipendente: $(Y_i, X_i) \perp (Y_j, X_j) \quad i \neq j \rightarrow Y_i \perp Y_j \text{ e } X_i \perp X_j \quad Y_i \perp X_j$
- Identicamente distribuite: (Y_i, X_i) ha le stesse distib. congiunte delle coppie (Y_j, X_j)

$$F_i(g|X=x) = F_j(g|X=x)$$

$$3) 0 < E[Y_i^4] < +\infty$$

$$0 < E[X_i^4] < +\infty$$

\rightarrow Outliers poco probabili

$$z_i = \frac{Y_i - E[Y_i]}{\sqrt{\text{Var}[Y_i]}}$$

$K = E[z_i^4] \rightarrow$ indice di curtosi:

$$= E[(\nu_i - E[\nu_i])^4 - (X_i - E[X_i])^4] = \text{Cov}(\nu_i, X_i)$$

Ritorniamo alle 1). $E[\nu_i | X_i = 0]$

$$I) E[\nu_i] = 0$$

II) $\text{Cov}(\nu_i, X_i) = 0 \rightarrow X_i$ è regressore esogeno

$$III) \text{Cov}(\nu_i, m(x_i)) = 0$$

$$E[z] = E_w [E[z|w]] \rightarrow E[\nu_i] = E_x [E[\nu_i | X_i]] = 0$$

$$\text{Cov}(\nu_i, m(x_i)) = E[(\nu_i - E[\nu_i])(m(x_i) - E[m(x_i)])]$$

$$= E[\nu_i \cdot m(x_i)] - E[\nu_i] \cdot E[m(x_i)]$$

$$= E[\nu_i \cdot m(x_i)] = 0$$

$$E[\nu_i \cdot m(x_i)] = E_x [E[w_i | X_i]] = E_x [E[\nu_i | X_i]]$$

$$= E_x [m(x_i) \cdot E[\nu_i | X_i]] = 0$$

costante

$$E[Y - m(x_i) | X_i] = m(x_i) \cdot E[Y | X_i]$$

$$\begin{cases} \textcircled{i)} \quad Y_i = \beta_0 + \beta_1 X_i + \nu_i \\ \textcircled{ii)} \quad E[\nu_i | X_i] = 0 \end{cases} \rightarrow E[Y_i | X_i] = g(X_i) = \beta_0 + \beta_1 X_i \rightarrow E[Y_i | X_i = x] = \beta_0 + \beta_1 x$$

$$E[Y_i | X_i] = E[(\beta_0 + \beta_1 X_i) + \nu_i | X_i] = E[\underbrace{\beta_0 + \beta_1 X_i}_{g(x_i)} | X_i] + E[\nu_i | X_i] = \beta_0 + \beta_1 X_i$$

$$\Delta E[Y_i | X_i] = \beta_1 \cdot \Delta X_i$$

$$E[Y_i | X_i = x + \Delta X] - E[Y_i | X_i = x] = \beta_1 \cdot \Delta X$$

Previsore ottimale secondo il criterio dell'errore quadratico medio

Mean Square Error
MSE

$Y_{m(x)}$ min $E[Y - m(x)]^2$ Se scelgo come $m(x) = E[Y | X]$. $E[m(x)^2]$ finito
 $E[Y | X]$ il predittore è ottimale

$$E[Y_i | X_i] = \beta_0 + \beta_1 \cdot X_i$$

Assunzioni:

i) Assunzione del modello

$$\text{ii)} \quad E[\nu_i | X_i] = 0 \quad \rightarrow E[\nu_i | X_1, \dots, X_n] = E[\nu_i | X_i] = 0 \quad \text{Correttezza di } \hat{\beta}_{OLS}$$

$$\text{iii)} \quad (Y_i, X_i) \sim \text{i.i.d.} \quad \text{see } (\nu_i, X_i) \sim \text{i.i.d.}$$

↳ Così lavora direttamente sugli errori

$$\nu_i = Y_i - \beta_0 - \beta_1 X_i = f(Y_i, X_i) \quad (Y_i, X_i) \perp (Y_j, X_j)$$

$$\nu_j = Y_j - \beta_0 - \beta_1 X_j = f(Y_j, X_j) \quad \nu_i \perp \nu_j \quad (Y_i, X_i) \perp (Y_j, X_j)$$

$$Y \perp X \quad E[Y | X] = f(X) \cdot f(Y | X) \xrightarrow{X \perp Y} f(Y | X) = f(Y)$$

$$E[Y | X] = E[Y] \quad \text{vale sempre}$$

$$E[X | Y] = E[X]$$

$$(\nu_i, X_i) \perp (\nu_j, X_j)$$

$$\rightarrow \nu_i \perp X_j \quad i \neq j$$

$(\nu_i, X_i) \sim \text{i.i.d.}$

$$\text{Var}(\nu_i) = \sigma^2 \quad \text{if:}$$

$$\text{Cov}(\nu_i, \nu_j) = 0 \quad \nu_i \perp \nu_j$$

$$\text{Var}(\nu_i | X_1, \dots, X_n) = ?$$

$$\text{Cov}(\nu_i, \nu_j | X_1, \dots, X_n) = ?$$

$$E[\nu_i | X_1, \dots, X_n] = E[\nu_i | X_i] = \overset{\text{Ass. 1}}{0}$$

$$\text{Var}(\nu_i) = E[(\nu_i - E[\nu_i])^2] = E[\nu_i^2] - E[\nu_i]^2 = E[\nu_i^2]$$

$$\begin{aligned} \text{Var}(\nu_i | X_1, \dots, X_n) &= E[(\nu_i - E[\nu_i | X_i])^2 | X] = E[\nu_i^2 | X] = E[\nu_i^2 | X_i] \\ &= h(X_i) \end{aligned}$$

potenzialmente possiamo avere omoschedasticità

$$\text{Cov}(\nu_i, \nu_j | X) = E[\nu_i \nu_j | X] = E[\nu_i | X] \cdot E[\nu_j | X]$$

29/09

$$\text{sub } \underline{X} = \text{condizionante} \quad \rightarrow \nu_i \perp \nu_j \rightarrow \text{Cov}(\nu_i, \nu_j | \underline{X}) = 0$$

$$\nu_i = f(Y_i, X_i)$$

$$\nu_j = f(Y_j, X_j)$$

$$\nu_i \perp \nu_j \quad \text{perché } (Y_i, X_i) \perp (Y_j, X_j)$$

costanti

ν_i, ν_j sono indipendenti anche se condizionati solo X_i, X_j . $E[\nu_i \nu_j | X] = E[\nu_i \nu_j | X_i, X_j] = 0$

$$\begin{aligned} \text{Var}(\nu_i | \underline{X}) &= \text{Var}(\nu_i | X_i) = h(X_i) \quad \text{eteroschedasticità} \\ &= \text{Var}(Y_i | X_i) \end{aligned}$$

$$\begin{aligned} \text{Var}(Y_i | \underline{X}) &= E[(Y_i - \underbrace{E[Y_i | \underline{X}]}_{E[Y_i | X_i]})^2 | \underline{X}] = E[\nu_i^2 | \underline{X}] \\ \nu_i &= (Y_i - \beta_0 - \beta_1 X_i) \end{aligned}$$

$$\text{Cov}(Y_i, Y_j | \underline{X}) \triangleq E[(Y_i - E[Y_i | \underline{X}]) \cdot (Y_j - E[Y_j | \underline{X}]) | \underline{X}]$$

$$Y_i = \beta_0 + \beta_1 X_i + \nu_i$$

$$\hookrightarrow \text{Cov}(X_i, \nu_i) \neq 0$$

$$\begin{cases} (1) Z_i \text{ contribuisce a spiegare } Y_i \Rightarrow Z_i \in \nu_i: & \nu_i = \beta_2 Z_i + V_i \\ (2) \text{Cov}(X_i, Z_i) \neq 0 \end{cases}$$

Fattori omessi da porre attenzione che sono correlati con X e Y

$$\text{Cov}(X_i, \nu_i) = \text{Cov}(X_i, \beta_2 Z_i + V_i) = \text{Cov}(X_i, \beta_2 Z_i) + \text{Cov}(X_i, V_i) =$$

La soluzione allora potrebbe essere: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + V_i$. Il problema è se X e V sono ancora correlati che mi dice che devo ancora cercare variabili omesse

In notazione matriciale

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \nu_i$$

Assunzioni:

$$1) E[\nu_i | X_{1i}, \dots, X_{ki}] = 0 \text{ con prob. 1} \rightarrow \begin{cases} 1) E[\nu_i] = 0 \text{ con prob. 1} \\ 2) \text{Cov}(\nu_i, X_{ij}) = 0 \forall j \end{cases}$$

$$2) (Y_i, X_{1i}, \dots, X_{ki}) \sim i.i.d.$$

$$3) 0 < E[Y_i^4] < +\infty, \dots, 0 < E[X_{ki}^4] < +\infty$$

$$4) \text{Assenza di collinearità perfetta } (X_{0i}, \dots, X_{ki})$$

$$\text{Var}(\nu_i) = \sigma_\nu^2 \quad \forall i$$

$$\text{Var}(\nu_i | X) = \text{Var}(\nu_i | X_{1i}, \dots, X_{ki}) = h(X_{1i}, \dots, X_{ki})$$

$$\text{Cov}(\nu_i, \nu_j | X) = \text{Cov}(\nu_i, \nu_j | X_{1i}, \dots, X_{ki}, X_{1j}, \dots, X_{kj}) = 0$$

Outline of the talk

1 Matrices

2 Matrix types

3 Operations with matrices

4 Matrix Rank

5 Inverse Matrix



Università degli Studi di Trieste

Matrices, vectors and scalars-1

A *matrix* is a collection or array of numbers

$$\begin{pmatrix} 1 & -2 & 4 \\ 1.2 & 3 & 5 \end{pmatrix}$$

2 Rows by 3 Columns (2x3) : *size* of the matrix

-This is a *rectangular* matrix

A *square* matrix: number of Rows = number of Columns = *dimension*

$$A_{(2 \times 2)} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$$



Matrices, vectors and scalars-2

$$A_{(2 \times 3)} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

Element (i,j) of matrix A: a_{ij}

$i=1,2$ (row index); $j=1,\dots,3$ (column index).



Matrices, vectors and scalars–3

Column vector

$$\begin{matrix} a \\ (2 \times 1) \end{matrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$$

or more simply

$$\begin{matrix} a \\ (2 \times 1) \end{matrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Row vector $\begin{matrix} a \\ (1 \times 3) \end{matrix} = (a_{11} \quad a_{12} \quad a_{13})$

or more simply

$$\begin{matrix} a \\ (1 \times 3) \end{matrix} = (a_1 \quad a_2 \quad a_3)$$

A Scalar: $\begin{matrix} c \\ (1 \times 1) \end{matrix} \in R$



Università degli Studi di Trieste

Outline of the talk

1 Matrices

2 Matrix types

3 Operations with matrices

4 Matrix Rank

5 Inverse Matrix



Università degli Studi di Trieste

Zero, Symmetric, Diagonal Matrices

A zero matrix

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

A symmetric (square) matrix: $a_{ij} = a_{ji}$

$$\begin{pmatrix} 1 & 3 & -1 \\ 3 & 0 & 5 \\ -1 & 5 & -2 \end{pmatrix}$$

A diagonal (square) matrix: $a_{ij} = 0$ for $i \neq j$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$



Scalar and Identity Matrix

A scalar matrix: a diagonal matrix with $a_{ii} = c \forall i$

e.g.
$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

The identity matrix: a scalar matrix with $a_{ii} = 1 \forall i$

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

It is the matrix equivalent of the number one!



Università degli Studi di Trieste

Outline of the talk

- 1 Matrices
- 2 Matrix types
- 3 Operations with matrices
- 4 Matrix Rank
- 5 Inverse Matrix



Università degli Studi di Trieste

Addition and Subtraction of Matrices-1

In order to perform operations with matrices, matrices must be *conformable*!

Specifically, for Addition and Subtraction matrices must have the same size or dimension. Then the operation is performed element by element.

Therefore, given matrices A and B , both of size $(R \times C)$, $A + B = D$ means that the sum of A and B is equal to the matrix D of size $(R \times C)$, where the element

$$d_{ij} = a_{ij} + b_{ij} \quad \forall ij$$

$$A - B = D$$

means that the difference between A and B is equal to the matrix D of size $(R \times C)$, where the element

$$d_{ij} = a_{ij} - b_{ij} \quad \forall ij$$



Addition and Subtraction of Matrices-2

If $A = \begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix}$

and $B = \begin{pmatrix} 0.2 & -0.1 \\ 0 & 0.3 \end{pmatrix}$ then:

$$A + B = \begin{pmatrix} 0.5 & 0.5 \\ -0.1 & 1.0 \end{pmatrix}$$

$$\text{and } A - B = \begin{pmatrix} 0.1 & 0.7 \\ -0.1 & 0.4 \end{pmatrix}$$



Università degli Studi di Trieste

Multiplication of a matrix by a scalar-1

Given a scalar c and a Matrix A , multiplying A by c means

$$cA = D$$

where

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2 \cdot I_2$$

$$d_{ij} = ca_{ij} \quad \forall ij$$

$$2 \begin{pmatrix} 0.5 & 0.5 \\ -0.1 & 1.0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -0.2 & 2 \end{pmatrix}$$



Properties-1

$$A + B = B + A$$

$$cA = Ac$$

$$c(A + B) = cA + cB$$



Università degli Studi di Trieste

Multiplying two matrices together-1

Matrices A and B must be conformable:

$$\begin{matrix} A & B \\ (m \times n) & (n \times p) \end{matrix} = \begin{matrix} D \\ (m \times p) \end{matrix}$$

i.e. the number of columns of the first matrix must be equal to the number of rows of the second matrix

-The resulting matrix has size given by the number of rows of the first matrix and number of columns of the second matrix



Multiplying two matrices together-2

Notice that in general

$$AB \neq BA$$

so the order of the factors is important!!



Università degli Studi di Trieste

Multiplying two matrices together-3

The product of matrices is also called "row by column" product.

Why?

Assume for the moment that the two matrices A and B are two vectors s.t.

$$\begin{matrix} a \\ (1 \times n) \end{matrix} \begin{matrix} b \\ (n \times 1) \end{matrix} = \begin{matrix} d \\ (1 \times 1) \end{matrix}$$

Se vettore Se Matrice

$$A^T A = \sum A_i \quad A^T A = \sum_i Z_S A_S$$

so that d is a scalar.

The product ab is obtained by computing the "row by column" product as follows:

$$\left(\begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ (1 \times n) \end{matrix} \right) \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \\ (n \times 1) \end{pmatrix} = \left(\sum_{k=1}^n a_{1k} b_{k1} \right)_{(1 \times 1)} = (d_{11}) = d$$



Multiplying two matrices together-4

For the general case

$$A_{(m \times n)} B_{(n \times p)} = D_{(m \times p)}$$

$$A_{3 \times 2} = \begin{bmatrix} 3 & 0 \\ 2 & 1 \\ 5 & 1 \end{bmatrix} \quad B_{3 \times 4} = \begin{bmatrix} -1 & -1 & 2 & 3 \\ -4 & 0 & 0 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix}$$

$$AB = D_{2 \times 4} = \begin{bmatrix} -2 & -2 & 8 & 10 \\ -8 & 3 & 14 & 13 \end{bmatrix}$$

Each element d_{ij} of D is obtained by considering the $i - th$ row of A and the $j - th$ column of B and computing the "row by column" product as follows:

$$d_{ij} = \left(\begin{array}{cccc} a_{i1} & a_{i2} & \dots & a_{in} \end{array} \right)_{(1 \times n)} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}_{(n \times 1)} = \sum_{k=1}^n a_{ik} b_{kj}$$



Multiplying two matrices together-5

For example

$$\begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 5 & 3 \end{pmatrix} = \begin{pmatrix} (2 \times 1 + 3 \times 5) & (2 \times 2 + 3 \times 3) \\ (1 \times 1 + 4 \times 5) & (1 \times 2 + 4 \times 3) \\ (3 \times 1 + 1 \times 5) & (3 \times 2 + 1 \times 3) \end{pmatrix}$$



Transpose of a matrix-1

Given

$$A_{(n \times m)} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{pmatrix}$$

the transpose of matrix A , written A' or A^T :

$$A'_{(m \times n)} = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{1m} & a_{2m} & a_{3m} & \dots & a_{nm} \end{pmatrix}$$

Transpose of a matrix-2

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 3 & 1 \end{pmatrix}$$

$$A'_{(2 \times 3)} = \begin{pmatrix} 2 & 1 & 3 \\ 3 & 4 & 1 \end{pmatrix}$$



Properties of the transpose

$$(A')' = A$$

$$(A + B)' = A' + B'$$

$$(AB)' = B'A'$$

Notice that the order is reversed!



Linear combination of vectors-1

Given m column vectors of dimension $(nx1)$, a_1, a_2, \dots, a_m , and scalars $\alpha_j \in R$, for $j = 1, \dots, m$,

the **linear combination** of the m column vectors with weights α_j is the following:

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_m a_m = \begin{matrix} c \\ (nx1) \end{matrix}$$

$$\begin{matrix} A \cdot \alpha = C \\ (n \times m) (m \times 1) \quad (n \times 1) \end{matrix}$$



Linearly independent vectors-1

The m vectors are **linearly independent** if and only if the only way to obtain that

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_m a_m = \begin{matrix} c \\ (n \times 1) \end{matrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is by setting all the weights equal to zero ($\alpha_j = 0$, for $j = 1, \dots, m$).



Università degli Studi di Trieste

Linearly independent vectors-2

Notice that:

- if one of the vectors a_j is a **zero vector** then the m vectors, a_1, a_2, \dots, a_m , are not linearly independent (they are **linearly dependent**);
- if two vectors are equal or proportional to each other then the two vectors are linearly dependent (and so are all the m vectors).



Linearly dependent vectors-1

If the m vectors, a_1, a_2, \dots, a_m , are **linearly dependent** then

at least one of them can be written as a linear combination of the remaining ones.

In fact, assume that

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_m a_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

with e.g. $\alpha_1 \neq 0$ then

$$a_1 = -(\alpha_2/\alpha_1)a_2 - \dots - (\alpha_m/\alpha_1)a_m$$



Linear combination and independency-3

It worth noticing that we might have considered row vectors instead of column ones. In fact, the concepts of linear combination and linear dependency/independency of a set of vectors will apply to row vectors exactly in the same way.



Outline of the talk

- 1 Matrices
- 2 Matrix types
- 3 Operations with matrices
- 4 Matrix Rank
- 5 Inverse Matrix



Università degli Studi di Trieste

Rank of a Matrix-1

Given a rectangular ($n \times m$) matrix A , we could ask ourselves what is the maximum number of the row vectors of the matrix that are linearly independent? This number is known as the row rank of the matrix, and it is at most equal to n .

The maximum number of column vectors of the matrix A that are linearly independent is called the column rank of the matrix.

There is an important theorem that states that the row rank and the column rank of a matrix are equal.

So, it's defined *rank* of a matrix the maximum number of column (row) vectors of the matrix that are linearly independent.

From the definition it follows that

$$\text{rank}(A) \leq \min(n, m)$$



Rank of a Matrix-2

If the rank is equal to $\min(n, m)$ then the matrix is said to be of *full rank*, otherwise it is said to be *singular* (or of not full rank).

A $n \times n$ square matrix is of full rank if and only if all its column vectors and all its row vectors are linearly independent.

Some properties of the rank:

- $\text{rank}(A) = \text{rank}(A')$
- $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$
- $\text{rank}(A'A) = \text{rank}(AA') = \text{rank}(A)$



Rank of a Matrix-3

- By convention the rank of a zero matrix is equal to zero.
- A vector with at least one element different from zero has rank equal to one.
- $\text{rank} \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} = 2$
- $\text{rank} \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} = 1$



Outline of the talk

- 1 Matrices
- 2 Matrix types
- 3 Operations with matrices
- 4 Matrix Rank
- 5 Inverse Matrix



Università degli Studi di Trieste

Inverse of a matrix-1

The inverse of a matrix A , denoted A^{-1} , where defined, is that matrix which, when pre-multiplied or post-multiplied by A , will result in the identity matrix I , i.e.

$$AA^{-1} = A^{-1}A = I$$

The inverse of a matrix exists if and only if the matrix is square and it is non-singular (i.e. it is of full rank). If the two conditions are satisfied, the matrix is said to be invertible.

The inverse of a 2×2 non-singular matrix A whose elements are

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\text{will be given by } A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

det A



Inverse of a matrix-2

The calculation of the inverse of a $n \times n$ matrix for $n > 2$ is more complex and beyond the scope of the text.

Properties of the inverse of a matrix include:

- $I^{-1} = I$
- $(A^{-1})^{-1} = A$
- $(A')^{-1} = (A^{-1})'$
- $(AB)^{-1} = B^{-1}A^{-1}$ if both A and B are invertible.



Chapter 5

Classical linear regression model assumptions and diagnostics

Violation of the Assumptions of the CLRM

- Recall that we assumed of the CLRM disturbance terms:
 1. $E(u_t) = 0$
 2. $\text{var}(u_t) = \sigma^2 < \infty$
 3. $\text{cov}(u_i, u_j) = 0$
 4. The X matrix is non-stochastic or fixed in repeated samples
 $\text{cov}(u_t, x_t) = 0$
 5. $u_t \sim N(0, \sigma^2)$

Investigating Violations of the Assumptions of the CLRM

- We will now study these assumptions further, and in particular look at:
 - How we test for violations
 - Causes
 - Consequences

in general we could encounter any combination of 3 problems:

- the coefficient estimates are wrong
- the associated standard errors are wrong
- the distribution that we assumed for the test statistics will be inappropriate
- Solutions
- the assumptions are no longer violated
- we work around the problem so that we use alternative techniques which are still valid

Statistical Distributions for Diagnostic Tests

- Often, an F - and a χ^2 - version of the test are available.
- The F -test version involves estimating a restricted and an unrestricted version of a test regression and comparing the RSS.
Likelihood Maximum?
- The χ^2 - version is sometimes called an “LM” test, and only has one degree of freedom parameter: the number of restrictions being tested, m .
- Asymptotically, the 2 tests are equivalent since the χ^2 is a special case of the F -distribution:

$$\frac{\chi^2(m)}{m} \rightarrow F(m, T - k) \quad \text{as} \quad (T - k) \rightarrow \infty$$

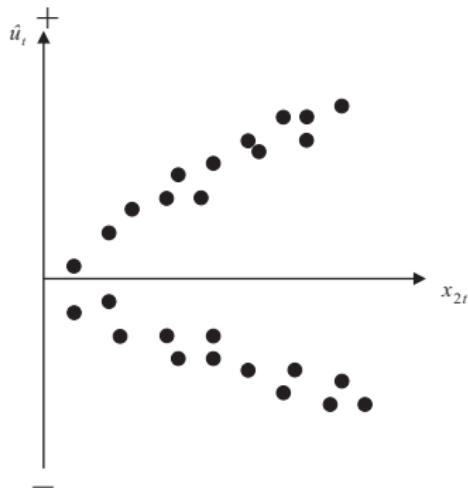
- For small samples, the F -version is preferable.

Assumption 1: $E(u_t) = 0$

- Assumption that the mean of the disturbances is zero.
- For all diagnostic tests, we cannot observe the disturbances and so perform the tests of the residuals.
termine di perturbazione \neq residuo
- The mean of the residuals will always be zero provided that there is a constant term in the regression.

Assumption 2: $\text{var}(u_t) = \sigma^2 < \infty$

- We have so far assumed that the variance of the errors is constant, σ^2 - this is known as homoscedasticity. If the errors do not have a constant variance, we say that they are heteroscedastic e.g. say we estimate a regression and calculate the residuals, \hat{u}_t .



Detection of Heteroscedasticity: The GQ Test

- Graphical methods
- Formal tests: There are many of them: we will discuss Goldfeld-Quandt test and White's test

The Goldfeld-Quandt (GQ) test is carried out as follows.

1. Split the total sample of length T into two sub-samples of length T_1 and T_2 . The regression model is estimated on each sub-sample and the two residual variances are calculated.
2. The null hypothesis is that the variances of the disturbances are equal, $H_0 : \sigma_1^2 = \sigma_2^2$

Detection of Heteroscedasticity: The GQ Test (Cont'd)

3. The test statistic, denoted GQ , is simply the ratio of the two residual variances where the larger of the two variances must be placed in the numerator.

$$GQ = \frac{s_1^2}{s_2^2} \sim F_{T_1-k, T_2-k}$$

4. The test statistic is distributed as an $F(T_1 - k, T_2 - k)$ under the null of homoscedasticity.
5. A problem with the test is that the choice of where to split the sample is usually arbitrary and may crucially affect the outcome of the test.

Detection of Heteroscedasticity using White's Test

- White's general test for heteroscedasticity is one of the best approaches because it makes few assumptions about the form of the heteroscedasticity.
- The test is carried out as follows:
 - Assume that the regression we carried out is as follows

Given $\text{rank}(X) = 3 \quad (k+1) \rightarrow k=2$

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

And we want to test $\text{Var}(u_t) = \sigma^2$. We estimate the model, obtaining the residuals, \hat{u}_t .

- Then run the auxiliary regression

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

Detection of Heteroscedasticity using White's Test (Cont'd)

3. Obtain R^2 from the auxiliary regression and multiply it by the number of observations, T . It can be shown that

$$TR^2 \sim \chi^2(m)$$

where m is the number of regressors in the auxiliary regression excluding the constant term.

$$m = k - \text{rank}(X) - 1$$

4. If the χ^2 test statistic from step 3 is greater than the corresponding value from the statistical table then reject the null hypothesis that the disturbances are homoscedastic.

Consequences of Using OLS in the Presence of Heteroscedasticity

- OLS estimation still gives unbiased coefficient estimates, but they are no longer BLUE.
- This implies that if we still use OLS in the presence of heteroscedasticity, our standard errors could be inappropriate and hence any inferences we make could be misleading.
- Whether the standard errors calculated using the usual formulae are too big or too small will depend upon the form of the heteroscedasticity.

How Do we Deal with Heteroscedasticity?

- If the form (i.e. the cause) of the heteroscedasticity is known, then we can use an estimation method which takes this into account (called generalised least squares, GLS).
- A simple illustration of GLS is as follows: Suppose that the error variance is related to another variable z_t by

$$\text{var}(u_t) = \sigma^2 z_t^2$$

- To remove the heteroscedasticity, divide the regression equation by z_t

$$\frac{y_t}{z_t} = \beta_1 \frac{1}{z_t} + \beta_2 \frac{x_{2t}}{z_t} + \beta_3 \frac{x_{3t}}{z_t} + v_t$$

How Do we Deal with Heteroscedasticity? (Cont'd)

where $v_t = \frac{u_t}{z_t}$ is an error term.

- Now $\text{var}(u_t) = \sigma^2 z_t^2$,
 $\text{var}(v_t) = \text{var}\left(\frac{u_t}{z_t}\right) = \frac{\text{var}(u_t)}{z_t^2} = \frac{\sigma^2 z_t^2}{z_t^2} = \sigma^2$ for known z_t .

Other Approaches to Dealing with Heteroscedasticity

- So the disturbances from the new regression equation will be homoscedastic.
- Other solutions include:
 1. Transforming the variables into logs or reducing by some other measure of “size”.
 2. Use White's heteroscedasticity consistent standard error estimates.

The effect of using White's correction is that in general the standard errors for the slope coefficients are increased relative to the usual OLS standard errors.

This makes us more “conservative” in hypothesis testing, so that we would need more evidence against the null hypothesis before we would reject it.

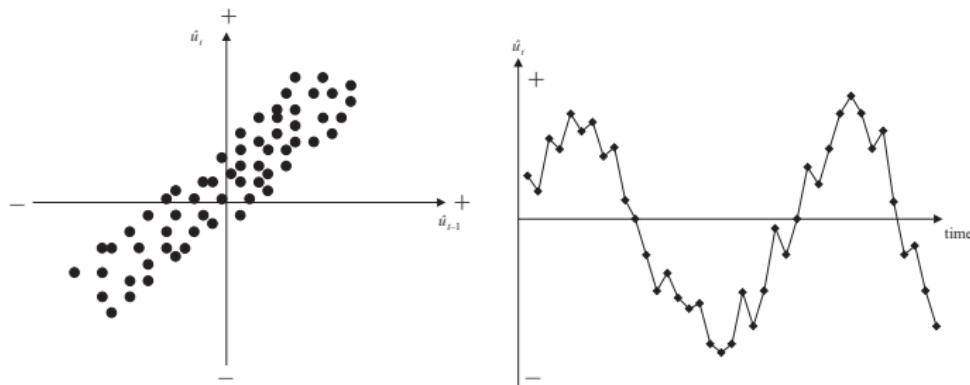
Background – The Concept of a Lagged Value

t	y_t	y_{t-1}	Δy_t
2006M09	0.8	—	—
2006M10	1.3	0.8	$(1.3 - 0.8) = 0.5$
2006M11	-0.9	1.3	$(-0.9 - 1.3) = -2.2$
2006M12	0.2	-0.9	$(0.2 - -0.9) = 1.1$
2007M01	-1.7	0.2	$(-1.7 - 0.2) = -1.9$
2007M02	2.3	-1.7	$(2.3 - -1.7) = 4.0$
2007M03	0.1	2.3	$(0.1 - 2.3) = -2.2$
2007M04	0.0	0.1	$(0.0 - 0.1) = -0.1$
.	.	.	.
.	.	.	.
.	.	.	.

Autocorrelation

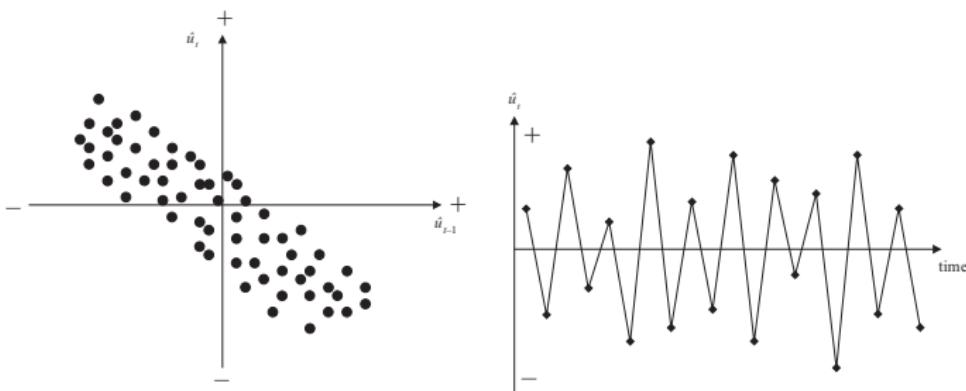
- We assumed of the CLRM's errors that $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$,
This is essentially the same as saying there is no pattern in the errors.
- Obviously we never have the actual u 's, so we use their sample counterpart, the residuals (the \hat{u}_t 's).
- If there are patterns in the residuals from a model, we say that they are autocorrelated.
- Some stereotypical patterns we may find in the residuals are given on the next 3 slides.

Positive Autocorrelation



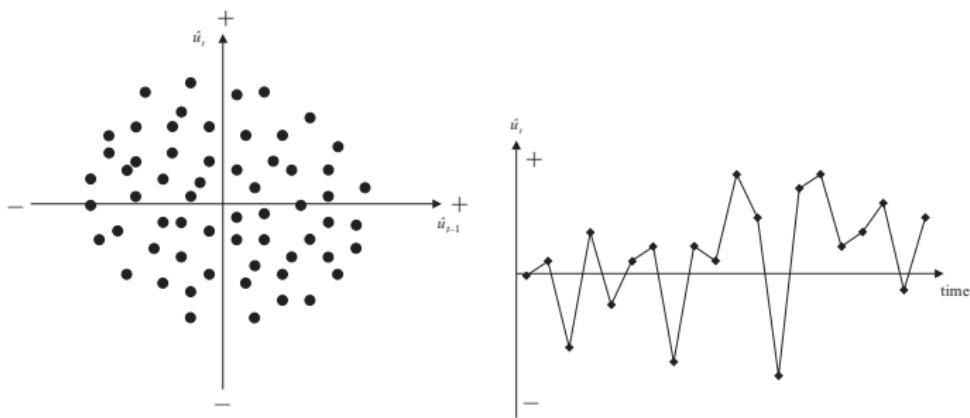
Positive Autocorrelation is indicated by a cyclical residual plot over time.

Negative Autocorrelation



Negative autocorrelation is indicated by an alternating pattern where the residuals cross the time axis more frequently than if they were distributed randomly

No pattern in residuals – No autocorrelation



No pattern in residuals at all: this is what we would like to see

Detecting Autocorrelation: The Durbin-Watson Test

- The Durbin-Watson (DW) is a test for first order autocorrelation - i.e. it assumes that the relationship is between an error and the previous one

$$u_t = \rho u_{t-1} + v_t \quad (1)$$

where $v_t \sim N(0, \sigma_v^2)$.

- The DW test statistic actually tests

$$H_0 : \rho = 0 \quad \text{and} \quad H_1 : \rho \neq 0$$

- The test statistic is calculated by

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^T \hat{u}_t^2}$$

The Durbin-Watson Test: Critical Values

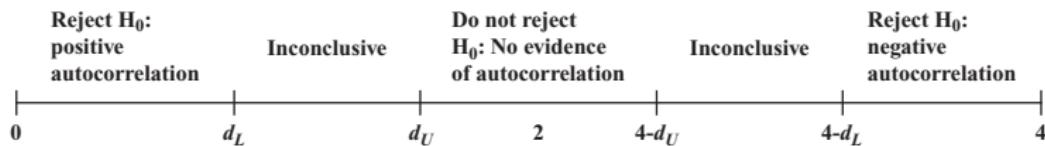
- We can also write

$$DW \approx 2(1 - \hat{\rho}) \quad (2)$$

where $\hat{\rho}$ is the estimated correlation coefficient. Since $\hat{\rho}$ is a correlation, it implies that $-1 \leq \hat{\rho} \leq 1$.

- Rearranging for DW from (2) would give $0 \leq DW \leq 4$.
- If $\hat{\rho} = 0$, $DW=2$. So roughly speaking, do not reject the null hypothesis if DW is near 2 → i.e. there is little evidence of autocorrelation
- Unfortunately, DW has 2 critical values, an upper critical value (d_U) and a lower critical value (d_L), and there is also an intermediate region where we can neither reject nor not reject H_0 .

The Durbin-Watson Test: Interpreting the Results



Conditions which Must be Fulfilled for DW to be a Valid Test

1. Constant term in regression
2. Regressors are non-stochastic
3. No lags of dependent variable

Another Test for Autocorrelation: The Breusch-Godfrey Test

- It is a more general test for r^{th} order autocorrelation:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \dots + \rho_r u_{t-r} + v_t,$$
$$v_t \sim N(0, \sigma_v^2)$$

- The null and alternative hypotheses are:

$$H_0 : \rho_1 = 0 \text{ and } \rho_2 = 0 \text{ and } \dots \text{ and } \rho_r = 0$$

$$H_1 : \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or } \dots \text{ or } \rho_r \neq 0$$

- The test is carried out as follows:

Another Test for Autocorrelation: The Breusch-Godfrey Test (Cont'd)

1. Estimate the linear regression using OLS and obtain the residuals, \hat{u}_t .
2. Regress \hat{u}_t on all of the regressors from stage 1 (the xs) plus $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-r}$;
Obtain R^2 from this regression.
3. It can be shown that

$$(T - r)R^2 \sim \chi_r^2$$

- If the test statistic exceeds the critical value from the statistical tables, reject the null hypothesis of no autocorrelation.

Consequences of Ignoring Autocorrelation if it is Present

- The coefficient estimates derived using OLS are still unbiased, but they are inefficient, i.e. they are not BLUE, even in large sample sizes.
- Thus, if the standard error estimates are inappropriate, there exists the possibility that we could make the wrong inferences.
- R^2 is likely to be inflated relative to its “correct” value for positively correlated residuals.

“Remedies” for Autocorrelation

- If the form of the autocorrelation is known, we could use a GLS procedure – i.e. an approach that allows for autocorrelated residuals e.g., Cochrane-Orcutt.
- But such procedures that “correct” for autocorrelation require assumptions about the form of the autocorrelation.
- If these assumptions are invalid, the cure would be more dangerous than the disease! - see Hendry and Mizon (1978).
- However, it is unlikely to be the case that the form of the autocorrelation is known, and a more “modern” view is that residual autocorrelation presents an opportunity to modify the regression.

Multicollinearity

- This problem occurs when the explanatory variables are very highly correlated with each other.
- Perfect multicollinearity
 - Cannot estimate all the coefficients
 - e.g. suppose $x_3 = 2x_2$
and the model is $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$
- Problems if Near Multicollinearity is Present but Ignored
 - R^2 will be high but the individual coefficients will have high standard errors.
 - The regression becomes very sensitive to small changes in the specification.
 - Thus confidence intervals for the parameters will be very wide, and significance tests might therefore give inappropriate conclusions.

Solutions to the Problem of Multicollinearity

- “Traditional” approaches, such as ridge regression or principal components. But these usually bring more problems than they solve.
- Some econometricians argue that if the model is otherwise OK, just ignore it
- The easiest ways to “cure” the problems are
 - drop one of the collinear variables
 - transform the highly correlated variables into a ratio
 - go out and collect more data e.g.
 - a longer run of data
 - switch to a higher frequency

Adopting the Wrong Functional Form

- We have previously assumed that the appropriate functional form is linear.
- This may not always be true.
- We can formally test this using Ramsey's RESET test, which is a general test for mis-specification of functional form.
- Essentially the method works by adding higher order terms of the fitted values (e.g. \hat{y}_t^2, \hat{y}_t^3 , etc.) into an auxiliary regression: Regress \hat{u}_t on powers of the fitted values:

$$\hat{u}_t = \beta_0 + \beta_1 \hat{y}_t^2 + \beta_2 \hat{y}_t^3 + \cdots + \beta_{p-1} \hat{y}_t^p + v_t$$

Obtain R^2 from this regression. The test statistic is given by TR^2 and is distributed as a $\chi^2(p - 1)$.

- So if the value of the test statistic is greater than a $\chi^2(p - 1)$ then reject the null hypothesis that the functional form was correct.

But what do we do if this is the case?

- The RESET test gives us no guide as to what a better specification might be.
- One possible cause of rejection of the test is if the true model is

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t}^2 + \beta_4 x_{2t}^3 + u_t$$

In this case the remedy is obvious.

- Another possibility is to transform the data into logarithms. This will linearise many previously multiplicative models into additive ones:

$$y_t = A x_t^\beta e^{u_t} \Leftrightarrow \ln(y_t) = \alpha + \beta \ln(x_t) + u_t$$

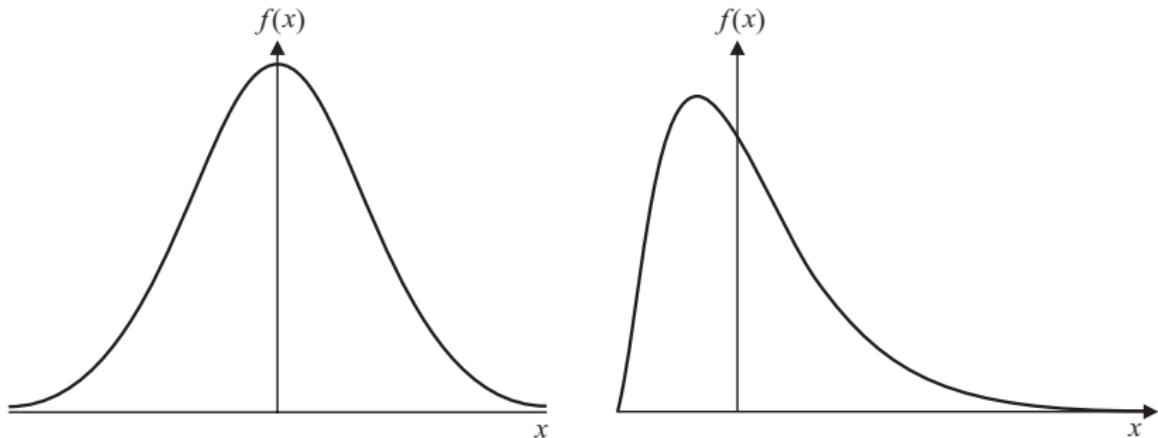
Testing the Normality Assumption

- Why did we need to assume normality for hypothesis testing?

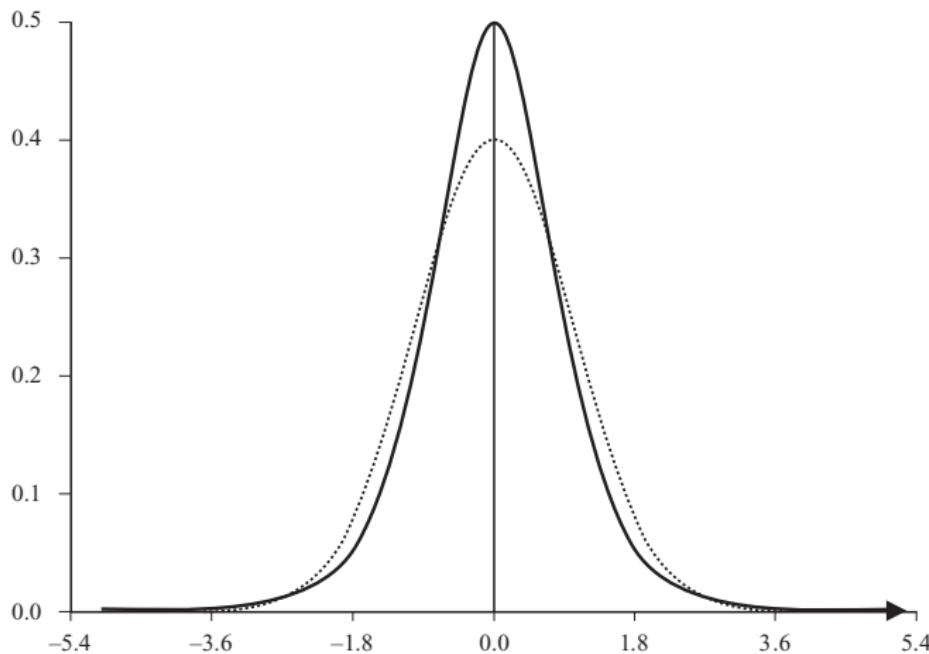
Testing for Departures from Normality

- *The Bera Jarque normality test*
- A normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3.
- The kurtosis of the normal distribution is 3 so its excess kurtosis ($b_2 - 3$) is zero.
- Skewness and kurtosis are the (standardised) third and fourth moments of a distribution.

Normal versus Skewed Distributions



Leptokurtic versus Normal Distribution



Testing for Normality

- Bera and Jarque formalise this by testing the residuals for normality by testing whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero.
- It can be proved that the coefficients of skewness and kurtosis can be expressed respectively as:

$$b_1 = \frac{E[u^3]}{(\sigma^2)^{3/2}} \quad \text{and} \quad b_2 = \frac{E[u^4]}{(\sigma^2)^2}$$

- The Bera Jarque test statistic is given by

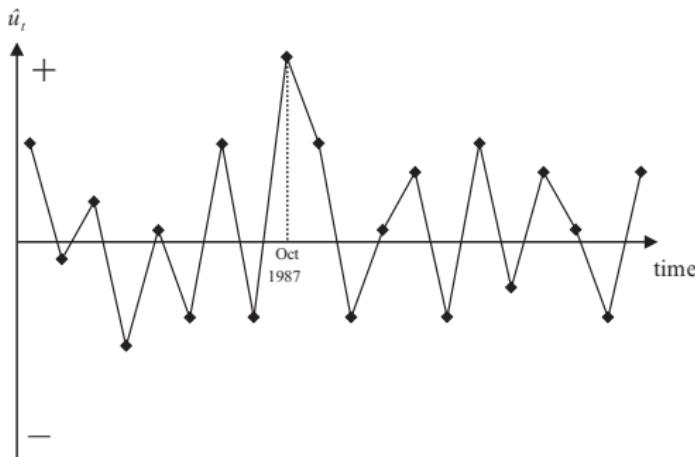
$$W = T \left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24} \right] \sim \chi^2$$

- We estimate b_1 and b_2 using the residuals from the OLS regression, .

What do we do if we find evidence of Non-Normality?

- It is not obvious what we should do!
- Could use a method which does not assume normality, but difficult and what are its properties?
- Often the case that one or two very extreme residuals causes us to reject the normality assumption.
- An alternative is to use dummy variables.
e.g. say we estimate a monthly model of asset returns from 1980-1990, and we plot the residuals, and find a particularly large outlier for October 1987:

What do we do if we find evidence of Non-Normality? (cont'd)



- Create a new variable:

$D87M10_t = 1$ during October 1987 and zero otherwise.

This effectively knocks out that observation. But we need a theoretical reason for adding dummy variables.

Omission of an Important Variable or Inclusion of an Irrelevant Variable

Omission of an Important Variable

- Consequence: The estimated coefficients on all the other variables will be biased and inconsistent unless the excluded variable is uncorrelated with all the included variables.
- Even if this condition is satisfied, the estimate of the coefficient on the constant term will be biased.
- The standard errors will also be biased.

Inclusion of an Irrelevant Variable

- Coefficient estimates will still be consistent and unbiased, but the estimators will be inefficient.

Parameter Stability Tests

- So far, we have estimated regressions such as

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

- We have implicitly assumed that the parameters (β_1 , β_2 and β_3) are constant for the entire sample period.
- We can test this implicit assumption using parameter stability tests. The idea is essentially to split the data into sub-periods and then to estimate up to three models, for each of the sub-parts and for all the data and then to “compare” the RSS of the models.
- There are two types of test we can look at:
 - Chow test (analysis of variance test)
 - Predictive failure tests

The Chow Test

- The steps involved are:
 1. Split the data into two sub-periods. Estimate the regression over the whole period and then for the two sub-periods separately (3 regressions). Obtain the RSS for each regression.
 2. The restricted regression is now the regression for the whole period while the “unrestricted regression” comes in two parts: for each of the sub-samples.
We can thus form an F-test which is the difference between the *RSS*'s.
The statistic is

$$\text{test statistic} = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2k}{k}$$

where:

RSS = RSS for whole sample

The Predictive Failure Test

- Problem with the Chow test is that we need to have enough data to do the regression on both sub-samples, i.e. $T_1 \gg k$, $T_2 \gg k$.
- An alternative formulation is the predictive failure test.
- What we do with the predictive failure test is estimate the regression over a “long” sub-period (i.e. most of the data) and then we predict values for the other period and compare the two.

To calculate the test:

- Run the regression for the whole period (the restricted regression) and obtain the RSS

The Predictive Failure Test (Cont'd)

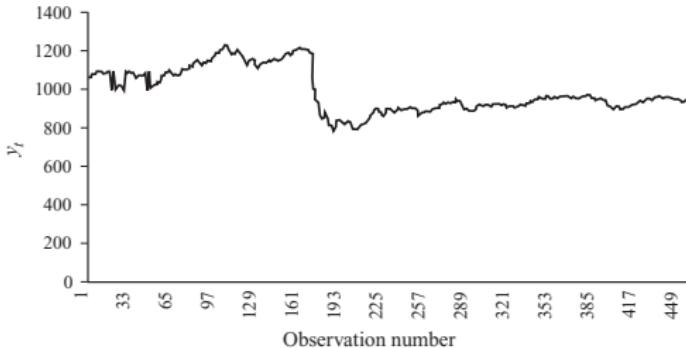
- Run the regression for the “large” sub-period and obtain the RSS (called RSS_1). Note we call the number of observations T_1 (even though it may come second).

$$\text{test statistic} = \frac{RSS - RSS_1}{RSS_1} \times \frac{T_1 - k}{T_2}$$

where T_2 = number of observations that the model is attempting to ‘predict’. The test statistic will follow an $F(T_2, T_1 - k)$.

How do we decide the sub-parts to use?

- As a rule of thumb, we could use all or some of the following
 - Plot the dependent variable over time and split the data accordingly to any obvious structural changes in the series, e.g.



- Split the data according to any known important historical events (e.g. stock market crash, new government elected)
- Use all but the last few observations and do a predictive failure test on those.

QLR test

Measurement Errors

- If there is measurement error in one or more of the explanatory variables, this will violate the assumption that the explanatory variables are non-stochastic
- Sometimes this is also known as the errors-in-variables problem
- Measurement errors can occur in a variety of circumstances, e.g.
 - Macroeconomic variables are almost always estimated quantities (GDP, inflation, and so on), as is most information contained in company accounts
 - Sometimes we cannot observe or obtain data on a variable we require and so we need to use a proxy variable – for instance, many models include expected quantities (e.g., expected inflation) but we cannot typically measure expectations.

Measurement Error in the Explanatory Variable(s)

- Suppose that we wish to estimate a model containing just one explanatory variable, x_t :

$$y_t = \beta_1 + \beta_2 x_t + u_t$$

where u_t is a disturbance term.

- Suppose further that x_t is measured with error so that instead of observing its true value, we observe a noisy version, \tilde{x}_t , that comprises the actual x_t plus some additional noise, v_t that is independent of x_t and u_t :

$$\tilde{x}_t = x_t + v_t$$

- Taking the first equation and substituting in for x_t from the second:

$$y_t = \beta_1 + \beta_2(\tilde{x}_t - v_t) + u_t$$

Measurement Error in the Explanatory Variable(s)

(Cont'd)

- We can rewrite this equation by separately expressing the composite error term, $(u_t - \beta_2 v_t)$

$$y_t = \beta_1 + \beta_2 \tilde{x}_t + (u_t - \beta_2 v_t)$$

Measurement Error in the Explanatory Variable(s)

- It should be clear from this equation and the one for the explanatory variable measured with error, \tilde{x}_t and the composite error term, $(u_t - \beta_2 v_t)$, are correlated since both depend on v_t
- Thus the requirement that the explanatory variables are non-stochastic does not hold
- This causes the parameters to be estimated inconsistently
- The size of the bias in the estimates will be a function of the variance of the noise in x_t as a proportion of the overall disturbance variance
- If β_2 is positive, the bias will be negative but if β_2 is negative, the bias will be positive
- So the parameter estimate will always be biased towards zero as a result of the measurement noise.

Measurement Error and Tests of the CAPM

- The standard approach to testing the CAPM pioneered by Fama and MacBeth (1973) comprises two stages
- Since the betas are estimated at the first stage rather than being directly observable, they will surely contain measurement error
- The effect of this has sometimes been termed attenuation bias.
- Tests of the CAPM showed that the relationship between beta and returns was smaller than expected, and this is precisely what would happen as a result of measurement error

Measurement Error and Tests of the CAPM

(Cont'd)

- Various approaches to solving this issue have been proposed, the most common of which is to use portfolio betas in place of individual betas
- An alternative approach (Shanken,1992) is to modify the standard errors in the second stage regression to adjust directly for the measurement errors.

Measurement Error in the Explained Variable

- Measurement error in the explained variable is much less serious than in the explanatory variable(s)
- This is one of the motivations for the inclusion of the disturbance term in a regression model
- When the explained variable is measured with error, the disturbance term will in effect be a composite of the usual disturbance term and another source of noise from the measurement error
- Then the parameter estimates will still be consistent and unbiased and the usual formulae for calculating standard errors will still be appropriate
- The only consequence is that the additional noise means the standard errors will be enlarged relative to the situation where there was no measurement error in y .

Capitolo 18

$$Y_i = \beta_0 X_{0i} + \dots + \beta_k X_{ki} + \nu_i \quad i=1, \dots, n$$

$$\begin{cases} Y_1 = \beta_0 X_{01} + \dots + \beta_k X_{k1} + \nu_1 \\ Y_2 = \beta_0 X_{02} + \dots + \beta_k X_{k2} + \nu_2 \\ \vdots \\ Y_n = \beta_0 X_{0n} + \dots + \beta_k X_{kn} + \nu_n \end{cases}$$

$$\underline{y} = \underline{X} \underline{\beta} + \underline{\nu}$$

$n \times 1$ $n \times (k+1)$ - $(k+1) \times 1$ $n \times 1$

$$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}$$

$$\begin{aligned} \underline{X} &= \begin{bmatrix} \underline{X}_1 & \underline{X}_2 & \dots & \underline{X}_k \end{bmatrix}_{n \times (k+1)} \\ \underline{X}_j &= \begin{bmatrix} x_{j1} \\ \vdots \\ x_{jn} \end{bmatrix}_{1 \times n} \quad j=0, \dots, k \end{aligned}$$

Prodotto interno

$$\underline{v}^T \underline{v} = c = \sum_{i=1}^n v_i^2$$

$n \times 1$ $1 \times n$ $n \times 1$

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1}$$

Minimi quadrati ordinari - HQO

$$\sum_{i=1}^n \nu_i^2 = \nu^T \nu = (\underline{y} - \underline{X} \underline{\beta})^T (\underline{y} - \underline{X} \underline{\beta})$$

$$\min_{\underline{\beta}} \underline{\nu}^T \underline{\nu}$$

$$\underline{\nu} = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix}$$

$$\min_{\{\beta_0, \dots, \beta_k\}} \sum_{i=1}^n (\underline{y_i} - \beta_0 - \beta_1 \underline{x_{i1}} - \dots - \beta_k \underline{x_{ik}})^2$$

$$\left\{ \begin{array}{l} \frac{\partial S(\cdot)}{\partial \beta_0} = \sum_{i=1}^n 2(\underline{y_i})(-1) = 0 \\ \vdots \\ \frac{\partial S(\cdot)}{\partial \beta_k} = \sum_{i=1}^n 2(\underline{y_i})(-\underline{x_{ik}}) = 0 \end{array} \right.$$

$$\begin{bmatrix} (-2) \cdot \sum \underline{y_i} \\ (-2) \sum \underline{y_i} \underline{x_{ik}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$(-2) \cdot \begin{bmatrix} \sum \underline{y_i} \cdot 1 \\ \sum \underline{y_i} \underline{x_{i1}} \\ \vdots \\ \sum \underline{y_i} \underline{x_{ik}} \end{bmatrix} = \underset{(k+1) \times 1}{0} \quad / \circ (-\frac{1}{2})$$

3/10

$$\underline{y} = \underline{y} - \underline{X}\underline{\beta} \quad X' \underline{y} = 0 \rightarrow \underline{X}'\underline{y} - (\underline{X}'\underline{X})\underline{\beta} = 0 \rightarrow \underline{\beta}^* = (\underline{X}'\underline{X})^{-1} \underline{X}'\underline{y}$$

$\underline{[(k+1) \times 1]} = \underline{[(k+1) \times (k+1)]} \times \underline{[(k+1) \times n]}$

$$\hat{\underline{y}} = \underline{X} \hat{\underline{\beta}}_{OLS} = \underbrace{[\underline{X} (\underline{X}'\underline{X})^{-1} \underline{X}']}_{M_{n \times 1}} \underline{y}$$

$$\hat{\underline{\mu}} = \underline{y} - \hat{\underline{y}} = \underline{I}_n \underline{y} - M \underline{y} = (\underline{I}_n - M) \underline{y}$$

$$\hat{\underline{\beta}}_{OLS} = \underline{\beta}^*$$

$\exists (X'X) ?$

$(X'X)$ è una matrice quadrata $(k+1) \times (k+1)$

Se $\text{range}(X'X) = (k+1) \rightarrow (X'X)$ è di range pieno \rightarrow è invertibile

T.H. $\text{range}(X'X) = \text{range}(X) = (k+1)$

$(X'X)$ è invertibile sse $\text{range}(X) = k+1$

X
 $n \times (k+1)$ se $n \geq (k+1) \rightarrow X$ è di range pieno

Assenza di collinearità perfetta \leftrightarrow le colonne di X sono lin. indip.

Lo stimatore è consistente, non distorto IAT non è efficiente

$$\hat{\beta}_{OLS} = (X'X)^{-1} X' y = (X'X)^{-1} (X'X) \beta + (X'X)^{-1} X' \nu$$

$$y = X\beta + \nu \quad = \beta + (X'X)^{-1} X' \nu$$

Errore di stima

$$(\hat{\beta}_{OLS} - \beta) = (X'X)^{-1} X' \nu$$

$$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \quad X_i \text{ sono v.v.}$$

$$E[\underline{X}] \stackrel{\text{def.}}{=} \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_k] \end{bmatrix} \quad R \times 1$$

$$E[AJ] = \begin{bmatrix} E[a_{11}] & \cdots & E[a_{1m}] \\ \vdots & & \vdots \\ E[a_{n1}] & \cdots & E[a_{nm}] \end{bmatrix}$$

Sia X ins. di v.a.

$$E[\underline{X}], \text{Var}(\underline{X}), \text{Cov}[X_i; X_j]$$

$$\text{Var}(X) \quad R \times 1$$

$$\text{Cov}(\underline{X}) = \sum_{k \times k} \begin{bmatrix} 1 & & & \\ & -\sigma_{ij} & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

Σ_x è simmetrica

$$\text{Var}(X) = E[(X - E[X])(X - E[X])']$$

$$\underline{Z} = \underline{X} - E[\underline{X}] \rightarrow E[\underline{Z}] = \underline{0}$$

$Z'Z$

Prodotto

Interno

Scalare

$Z \cdot Z'$

Prodotto

Esterno

Matrice

F vett.

$$\underline{x} \in \mathbb{R}^m \xrightarrow{f(\cdot)} \underline{y} \in \mathbb{R}$$

$$\underline{x} \in \mathbb{R}^m \xrightarrow{f(\cdot)} \underline{y} \in \mathbb{R}^n$$

$f(\cdot) =$
n × 1
 scalare

$$\begin{bmatrix} f_1(\underline{x}) \\ f_2(\underline{x}) \\ \vdots \\ f_n(\underline{x}) \end{bmatrix}$$

F vett. lin-offine

$$\underline{y}_i = c_i + \sum_{j=1}^n a_{ij} \underline{x}_j \quad i=1, \dots, n$$

$$\underline{y} = f(\underline{x}) = \underline{c} + A \cdot \underline{x}$$

lineare
affine

$$\underline{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \quad A = \begin{bmatrix} a_{ij} \end{bmatrix}_{n \times m}$$

$$E[\underline{y}] = E[\underline{c}] + \alpha E[\underline{x}] \\ = \underline{c} + \alpha \cdot E[\underline{x}]$$

$$\text{Var}[\underline{y}] = \alpha^2 \text{Var}[\underline{x}]$$

$$\underline{\underline{x}} \text{ v.e con } E[\underline{x}] = \underline{\underline{m}} \\ \text{Var}[\underline{\underline{x}}] = \underline{\underline{\Sigma_x}}$$

$$\underline{y} = \underline{c} + A \underline{x}$$

Th

$$E[\underline{y}] = \underline{c} + A E[\underline{x}] \\ \text{Var}[\underline{y}] = \text{Var}[A \underline{x}] = A \Sigma_x A'$$

Fz. di regressione vettoriale

$$E[\underline{\varphi} | \underline{x}] = g(\underline{x})$$

n × 1 n × (r+1)

$$E[\varphi | \underline{x}] = g(\underline{x})$$

1 × 1 k × 1

$$E[\varphi | X] = g(X)$$

1 × 1 n × h 1 × 1

$$E[\underline{Y} | \underline{X}] = \begin{bmatrix} E[Y_1 | \underline{X}] \\ \vdots \\ E[Y_m | \underline{X}] \end{bmatrix}$$

5/10

$$\text{Var}(\underline{\underline{X}}) = E[(\underline{\underline{X}} - E[\underline{\underline{X}}])(\underline{\underline{X}} - E[\underline{\underline{X}}])'] = \Sigma_{n \times n}$$

$$\text{Cov}(\underline{\underline{X}}, \underline{\underline{Y}}) = E[(\underline{\underline{X}} - E[\underline{\underline{X}}])(\underline{\underline{Y}} - E[\underline{\underline{Y}}])'] = C_{n \times m}$$

$$\text{Cov}(\underline{\underline{Y}}, \underline{\underline{X}}) = \tilde{C} = C'$$

Non vale la simmetria in generale

Proprietà Cov

- 1) $\text{Cov}(\underline{\alpha}, \underline{\underline{X}}) = \underline{\alpha}' \Sigma$
- 2) $\text{Cov}(\underline{\underline{X}}, \underline{\underline{X}}) = \text{Var}(\underline{\underline{X}})$
- 3) $\text{Cov}(A\underline{\underline{X}}, B\underline{\underline{Y}}) = A \text{Cov}(\underline{\underline{X}}, \underline{\underline{Y}}) B'$
- 4) $\text{Cov}(\underline{\alpha} + A\underline{\underline{X}} + C\underline{\underline{Z}}, B\underline{\underline{Y}}) = A \cdot \text{Cov}(\underline{\underline{X}}, \underline{\underline{Y}}) B' + C \cdot \text{Cov}(\underline{\underline{Z}}, \underline{\underline{Y}}) B'$

Se $\exists \underline{v} \in \mathbb{R}^n : \underline{v}' Q \underline{v} > 0 \rightarrow Q$ è def. positive

Se $\exists \underline{v} \in \mathbb{R}^n : \underline{v}' Q \underline{v} \geq 0 \rightarrow Q$ è semi def. positive

$\text{Var}(\underline{\underline{X}}) = \Sigma_{\underline{\underline{X}}} \quad \text{Se } X_1, \dots, X_n \text{ sono lin. indip.} \rightarrow \Sigma_{\underline{\underline{X}}} \text{ è def. positive}$

TH: se $\Sigma_{\underline{\underline{X}}}$ def. pos. ($\Sigma_{\underline{\underline{X}}} > 0$) $\rightarrow \exists! \Sigma_{\underline{\underline{X}}}^{-1}$ ed $\Sigma_{\underline{\underline{X}}}^{-1}$ è invertibile

se $\Sigma_{\underline{\underline{X}}}$ è semidef. pos. ($\Sigma_{\underline{\underline{X}}} \geq 0$) $\Sigma_{\underline{\underline{X}}}$ è singolare e quindi non è invertibile

G/10

$$\mathbb{E}[\underline{\gamma}_i | \underline{x}_{-i}, \underline{x}_i] = \underline{\beta}' \underline{x}_i = \mathbb{E}[\underline{\gamma}_i | \underline{x}]$$

$\underline{x}_i = \begin{bmatrix} x_{0i} \\ \vdots \\ x_{ki} \end{bmatrix}$

$$\gamma_i \perp x_{i,j}, \dots, x_{k,j} \rightarrow i \neq j$$

Condizioni necessarie

1) $\mathbb{E}[\underline{\mu} | \underline{x}] = \underline{\Omega} \rightarrow \mathbb{E}[\underline{\gamma} | \underline{x}] = \underline{x} \cdot \underline{\beta}$

2) $\text{Var}(\underline{\mu}_i | \underline{X}) = h(x_i) = \sigma_i^2$

2.1) $\text{Cov}(\mu_i, \mu_j | \underline{X}) = 0$

$\left. \begin{array}{l} \text{Var}(\underline{\mu} | \underline{X}) = \Sigma_{\mu} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \end{array} \right\}$

INVECE $\mu_i \sim \text{iid. } (0, \sigma_i^2)$

$\text{Var}(\underline{\mu}) = \sigma^2 I_k$

3) $0 < \mathbb{E}[\gamma^2] < +\infty$

Anche per i regressori:

$\gamma \in$ spazio lineare delle colonne d - x

$\gamma \in L(x)$

4) $\text{range}(x) = k+1$

$\hat{\beta}_{\text{OLS}} = (x'x)^{-1} x' \underline{\gamma}$

$\hat{\gamma} = x \hat{\beta} = \hat{\beta}_0 \tilde{x}_0 + \dots + \hat{\beta}_k \tilde{x}_k$

OLS - Ordinary Least Square

$$\min_{\beta} \sum_{i=1}^n (\underline{Y}_i - \underline{\beta})' (\underline{Y}_i - \underline{\beta})$$

$$\underline{Z} \in \mathbb{R}^{n \times 1}$$

$$d(\underline{z}, \underline{w}) = \sqrt{(\underline{z} - \underline{w})' (\underline{z} - \underline{w})}$$
$$= \sqrt{\sum_{i=1}^n (z_i - w_i)^2}$$

Minimizza la distanza tra \underline{Y} e $\hat{\underline{Y}}$.

Correttamente stima $\underline{\beta}$?

Penso farlo su: singoli coeff. $E[\hat{\beta}_j] = \beta_j$ $j = 0, \dots, k$

O posso farlo sul vettore

$$\hat{\underline{\beta}}_{OLS} = \underline{\beta} + (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{y}$$
$$A = -\underline{X}$$

Per la legge del valore atteso iterato

Se $E[\hat{\beta}_{OLS}|X] = \underline{\beta}$ allora $E[E[\hat{\beta}_{OLS}|X]] = E[\hat{\beta}_{OLS}]$

→ F. matriciale di X → condizione a X essere costante

$$E[\hat{\beta}|X] = E[\underline{\beta}|X] + E[A(X) \cdot \underline{\nu}|X]$$

$\underbrace{A(X)}_{\text{costante}} \cdot \underbrace{E[\underline{\nu}|X]}_{=0} = 0$

→ Dimostrazione
Correttezza

Se $E[\nu_i|X_i] = 0$ ma $E[\nu_i|X] \neq 0$ allora β_{OLS} è distorto.

$E[\nu|X] = 0 \leftrightarrow E[\nu_i|X_i] = 0$:

- I) $E[\nu_i|X_i] = 0$
- II) $(Y_i, X_i) \sim iid$

Problema consistenza

$\hat{\beta} \xrightarrow{n \rightarrow +\infty} \beta$ → Convergenza in probabilità $\nabla \delta > 0$

$n\text{-lim } \hat{\beta}_n = \beta \quad \lim_{n \rightarrow +\infty} P(|\hat{\beta}_n - \beta| \geq \delta) = 0 \quad \text{sse} \quad \lim_{n \rightarrow +\infty} P(|\hat{\beta}_n - \beta| < \delta) = 1$

Condizioni sufficienti

1) $\hat{\beta}$ asintoticamente corretto $\lim_{n \rightarrow +\infty} E[\hat{\beta}] = \beta$

2) $\lim_{n \rightarrow +\infty} \text{Var}(\hat{\beta}) = 0$

E' più facile dimostrare la consistenza che la correttezza

$n\text{-lim}_{n \rightarrow +\infty} \hat{\beta} = \beta \Leftrightarrow n\text{-lim}_{n \rightarrow +\infty} \hat{\beta}_j = \beta_j \quad \forall j \quad \hat{\beta} \xrightarrow{c} \beta$

$n\text{-lim} \hat{\beta} = n\text{-lim} (\beta + (X'X)^{-1}X'\underline{\nu}) = n\text{-lim} \beta + n\text{-lim} ((X'X)^{-1}X'\underline{\nu})$

Possiamo farlo sui vettori perché vale sui singoli elementi

$$\underline{\nu} = 0$$

$$\begin{cases} \text{Cov}(\nu_i, \chi_j) = 0 \\ E[\nu_i] = 0 \end{cases}$$

$(X'X)^{-1} \frac{1}{n} X' \underline{\nu} = (\frac{1}{n} X' X)^{-1} (\frac{1}{n} X' \underline{\nu}) \rightarrow \text{Momento campionario}$

$$(CA)^{-1} = A^{-1} C^{-1} = \frac{1}{C} A^{-1} \quad C \neq 0$$

$$\text{Cov}(X, \underline{\nu}) = \text{Cov}(X, Y) - \text{Cov}(X, X) = \frac{\text{Cov}(X, Y) - \beta \text{Var}(X)}{\text{Var}(X)} = \frac{S_{xy} - \beta}{S_{xx}} \xrightarrow{n \rightarrow +\infty} 0 \rightarrow S_{xy} = \beta$$

$$\frac{1}{n} \underline{\underline{X}}' \underline{\underline{X}} = \frac{1}{n} \sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\text{Prod esterno}} =$$

$\frac{1}{n} \sum_i^n 1$	$\frac{1}{n} \sum 1 \cdot \underline{\underline{X}}_{ii}$	Media campionaria = = Primo momento $= E[\underline{\underline{X}}_{ii}] = \mu_x$ $\underline{\underline{X}}_{ii} \sim \text{iid. } \forall i$
$\frac{1}{n} \sum \underline{\underline{X}}_{ii} \cdot 1$	$\frac{1}{n} \sum \underline{\underline{X}}_{ii}^2$	

$\frac{1}{n} \sum \underline{\underline{X}}_{ii}^2 \rightarrow E[w_i] = E[\underline{\underline{X}}_{ii}^2]$ $\rightarrow E[\underline{\underline{X}}_{ii}^4] < +\infty$

$$(j, l) \rightarrow \frac{1}{n} \sum_{i=1}^n \underline{\underline{X}}_{ji} \underline{\underline{X}}_{li} \rightarrow E[\underline{\underline{X}}_{ji}, \underline{\underline{X}}_{li}]$$

$$(\frac{1}{n} \underline{\underline{X}}' \underline{\underline{X}}) \xrightarrow{\rho} Q_x = E[\underline{\underline{X}}, \underline{\underline{X}}']$$

$$\begin{cases} \text{Cov}(\mu_i, \underline{\underline{X}}_{ji}) = 0 \\ j=1, \dots, k \\ E[\mu_i] = 0 \end{cases}$$

$$\rho\text{-lim}(\hat{\beta}) = (\underbrace{\rho\text{-lim} \frac{1}{n}(\underline{\underline{X}}' \underline{\underline{X}})}_{Q_x})^{-1} \cdot \rho\text{-lim}(\frac{1}{n} \underline{\underline{X}}' \underline{\underline{\mu}}) =$$

$$\rho\text{-lim}(g(\hat{\beta})) = g(\rho\text{-lim}(\hat{\beta}))$$

$$= \beta + (Q_x)^{-1} \rho\text{-lim}(\frac{1}{n} \underline{\underline{X}}' \underline{\underline{\mu}})$$

9/10

$$\rho\text{-lim}(\frac{1}{n} \underline{\underline{X}}' \underline{\underline{\mu}}) = \rho\text{-lim}(\frac{1}{n} \sum \underline{\underline{X}}_i \mu_i) = E[\underline{\underline{X}}_i \cdot \mu_i] \xrightarrow{\text{momento terzico}} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \rightarrow \text{posizione} = \text{cov}(\underline{\underline{X}}_i, \mu_i)$$

$$\underline{\underline{X}}_i = \begin{pmatrix} \underline{\underline{X}}_{j1} \\ \underline{\underline{X}}_{j2} \\ \vdots \\ \underline{\underline{X}}_{jk} \end{pmatrix}$$

$$\frac{1}{n} \sum_{i=1}^n \underline{\underline{X}}_{ji} \mu_i \xrightarrow{(k+1) \times 1} E[\underline{\underline{X}}_{ji} \cdot \mu_i]$$

\rightarrow Il momento seconds deve essere finito e non nullo
 \rightarrow Momento componev. o

9/10

$$E[Cov(X_i, \nu_i)] = Cov(X_i, \nu_i)$$

2) $(Y_i; X_{i1}, \dots, X_{ik}) \sim i.i.d.$ sse $(\nu_i; X_{i1}, \dots, X_{ki}) \sim i.i.d.$

$$\begin{aligned} 0 &< E[Y_i^4] < +\infty \\ 0 &< E[X_{ij}^4] < +\infty \end{aligned}$$

Secondo criterio congruente
 $\ln \left[\sum X_{ji}^2 / \nu_i^2 \right]$

Dis. di Cauchy $|E[XY]| \leq E[X]E[Y]$

$$\rho\text{-lim } (\hat{\beta} - \beta) = \underline{\beta}$$

$$(x'x)^{-1} \frac{1}{n} x' \nu$$

Al crescere di n va a convergere in una normale multivariata ($x_i \sim N(0, I)$)

$$\underbrace{(\rho\text{-lim } \frac{1}{n} x'x)^{-1}}_{Q_x} \quad \underbrace{\rho\text{-lim } \frac{1}{n} x' \nu}_{\underline{\beta}}$$

Parte sulla normale multivariate

si modifica regine di libro vice

NORMALE UNIVARIATA

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

i.d. $f(\underline{x}; \mu_1, \dots, \mu_n, \sigma_1^2, \dots, \sigma_n^2) = \prod f(x_i; \mu_i, \sigma_i^2) = (2\pi)^{-\frac{n}{2}} (\sigma_1^2, \dots, \sigma_n^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right\}$

NORMALE MULTIVARIATA

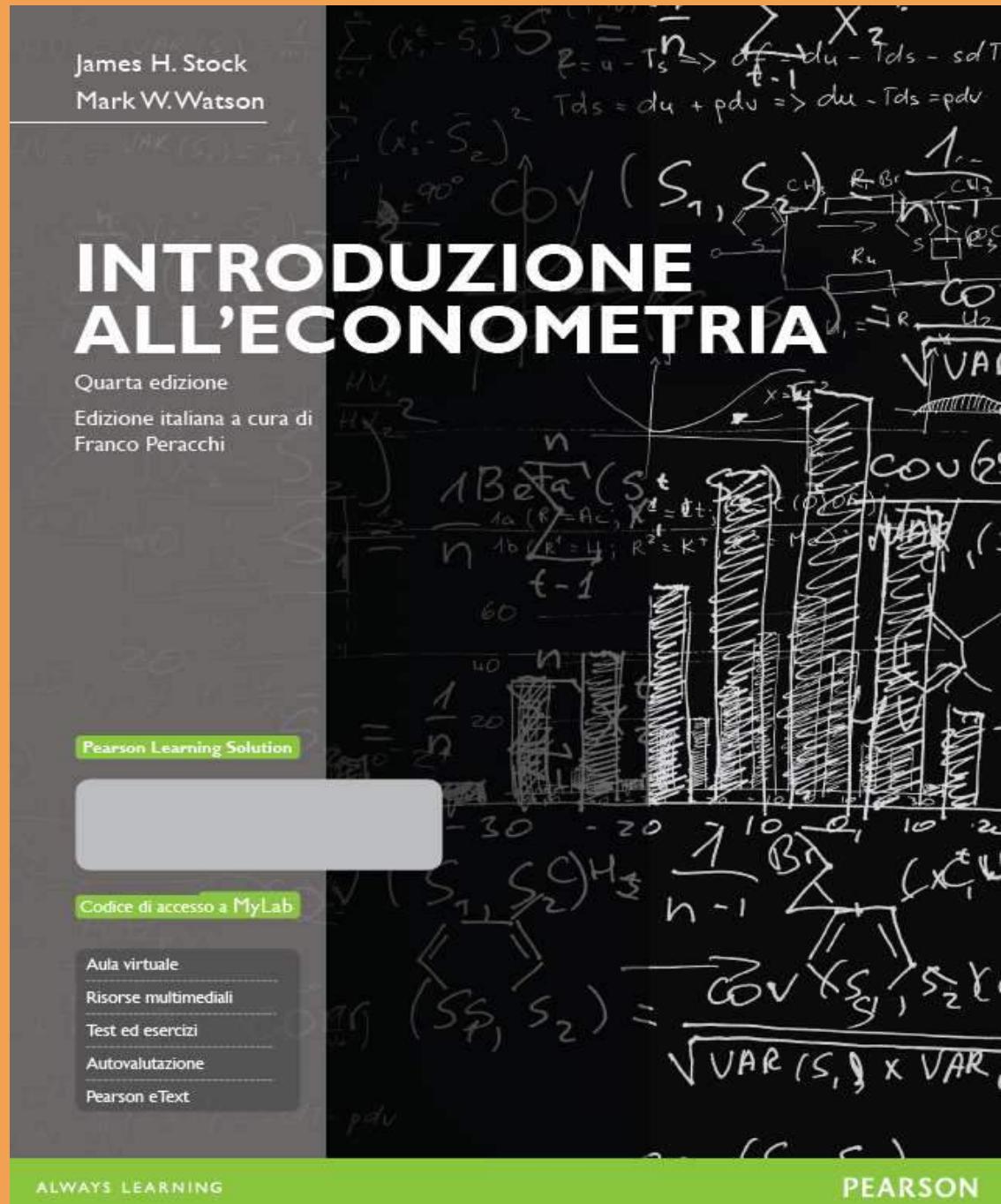
Formule generali

i.i.d. $f(\underline{x}; \mu, \Sigma) = \prod f(x_i; \mu_i, \sigma_i^2) = (2\pi)^{-\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})\right\}$

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

Capitolo 9

Valutazione di studi basati sulla regressione multipla



Quadro di riferimento per la valutazione di studi statistici: validità interna ed esterna (Paragrafo 9.1)

- **Validità interna**: le inferenze statistiche sugli effetti causali sono valide per la popolazione studiata.
- **Validità esterna**: le inferenze statistiche possono essere generalizzate dalla popolazione e dal contesto studiati ad altre popolazioni e altri contesti, dove il “contesto” fa riferimento all’ambiente legale, istituzionale e fisico e alle caratteristiche salienti.

Minacce alla validità interna dell'analisi di regressione multipla (Paragrafo 9.2)

Vedi anche slides Chap. 5 (messe in precedente)

Validità interna: le inferenze statistiche sugli effetti causali sono valide per la popolazione studiata.

Cinque minacce alla validità interna degli studi a regressione:

- Distorsione da variabili omesse → Alte instabilità delle stime e distorsione
- Forma funzionale incorretta → Test RESET di Ramsey
- Distorsione da errori nelle variabili → Errori di misurazione nelle variabili
- Distorsione da selezione campionaria → Componente Casuale Simplice
- Distorsione da causalità simultanea → Alte Correlazioni? → Test Durbin-Watson
Test Breusch-Godfrey

Tutte queste implicano che $E(u_i | X_{1i}, \dots, X_{ki}) \neq 0$

(o che non vale l'indipendenza in media condizionata) – nel qual caso lo stimatore OLS è distorto e inconsistente.

1. Distorsione da variabili omesse

La distorsione da variabili omesse nasce quando una variabile omessa è **sia:**

- I. una determinante di Y e
- II. correlata con almeno un regressore incluso.

- È stata esaminata in precedenza la distorsione da variabili omesse con una singola X . La distorsione nasce nelle regressioni multiple se la variabile omessa soddisfa le condizioni (i) e (ii) date in precedenza.
- Se la regressione multipla comprende variabili di controllo, occorre chiedersi se vi siano dei fattori omessi per i quali non esista un adeguato controllo, cioè se il termine di errore sia correlato con la variabile di interesse anche dopo che siano state inserite le variabili di controllo.

Soluzioni alla distorsione da variabili omesse

1. Se è possibile misurare la variabile causale omessa, inserirla come regressore aggiuntivo nella regressione multipla;
2. Se si possiedono dati su uno o più controlli e questi sono adeguati (nel senso del mantenimento della plausibilità dell'indipendenza in media condizionata), inserire le variabili di controllo;
3. Se possibile, usare *dati panel* nei quali ciascuna unità (individuo) venga osservata più di una volta;
4. Se non è possibile misurare la variabile omessa, usare la regressione con variabili strumentali;
5. Condurre un esperimento controllato casualizzato.
 - *Perché funziona?* Si ricordi: se X viene assegnata casualmente, allora X sarà necessariamente distribuita indipendentemente da u ; perciò $E(u|X = x) = 0$.

A. Errore di misura classico

Il modello di errore di misura classico presume che

$$\tilde{X}_i = X_i + v_i,$$

dove v_i è rumore casuale a media zero con $\text{corr}(X_i, v_i) = 0$ e $\text{corr}(u_i, v_i) = 0$.

Con il modello di errore di misura classico, $\hat{\beta}$ è distorto verso zero. Questa è l'idea: si supponga di prendere la variabile vera e quindi aggiungere una grande quantità di rumore casuale – numeri casuali generati dal computer. Entro il limite del "solo rumore", \tilde{X}_i sarà incorrelata a Y_i (e a qualsiasi altra cosa), quindi il coefficiente di regressione avrà valore atteso zero. Se X_i contiene del rumore ma \tilde{X}_i non è "solo rumore" allora la relazione tra $\hat{\beta}_Y$ e Y_i sarà attenuata, per cui \tilde{X}_i è distorto verso zero.

Errore di misura classico: i calcoli

$\tilde{X}_i = X_i + v_i$, dove $\text{corr}(X_i, v_i) = 0$ e $\text{corr}(u_i, v_i) = 0$.

Quindi $\text{var}(\tilde{X}_i) = \sigma_X^2 + \sigma_v^2$

$$\text{cov}(\tilde{X}_i, X_i - \tilde{X}) = \text{cov}(X_i + v_i, -v_i) = -\sigma_v^2$$

così

$$\text{cov}(\tilde{X}_i, \tilde{u}_i) = -\beta_1 \sigma_v^2$$

così $\hat{\beta}_1 \xrightarrow{p} \beta_1 - \beta_1 \frac{\sigma_v^2}{\sigma_{\tilde{X}}^2} = \left(1 - \frac{\sigma_v^2}{\sigma_{\tilde{X}}^2}\right) \beta_1$

$$= \left(\frac{\sigma_{\tilde{X}}^2 - \sigma_v^2}{\sigma_{\tilde{X}}^2}\right) \beta_1 - \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_v^2}\right) \beta_1$$

Quindi $\hat{\beta}_1$ è distorto verso zero.

Il modello di errore di misura classico è speciale perché assume che $\text{corr}(X_i, v_i) = 0$.

B. Errore di misura “migliore ipotesi”

Si supponga che l'intervistato non ricordi X_i , ma faccia un'ipotesi del tipo $\tilde{X}_i = E(X_i|W_i)$, dove $E(u_i|W_i) = 0$. Allora,

$$\begin{aligned}\text{cov}(\tilde{X}_i, \tilde{u}_i) &= \text{cov}(\tilde{X}_i, \beta_1(X_i - \tilde{X}_i) + u_i) \\ &= \beta_1 \text{cov}(\tilde{X}_i, X_i - \tilde{X}_i) + \text{cov}(\tilde{X}_i, u_i)\end{aligned}$$

• $\text{cov}(\tilde{X}_i, X_i - \tilde{X}_i) = 0$ perché $\tilde{X}_i = E(X_i|W_i)$ (siccome è la migliore ipotesi, l'errore $\tilde{X}_i - X_i$ è incorrelato con \tilde{X}_i).

• $\text{Cov}(\tilde{X}_i, u_i) = 0$ perché $E(u_i|W_i) = 0$ (u_i è funzione di W_i e $E(u_i|W_i) = 0$).

• Così $\text{cov}(\tilde{X}_i, \tilde{u}_i) = 0$, quindi β_1 è non distorto.

Errore di misura “migliore ipotesi” (continua)

- Con il modello “migliore ipotesi”, l’errore di misura è ancora presente – non si osserva il vero valore di X_i – ma qui questo errore di misura non introduce distorsione in $\hat{\beta}_1$!
- Il modello “migliore ipotesi” è estremo – non è sufficiente fare una buona ipotesi, è necessaria la “migliore” ipotesi $\tilde{X}_i \equiv E(X_i|W_i)$, cioè il valore atteso condizionato di X data W , dove $E(u_i|W_i) = 0$.

Insegnamenti dai modelli classico e «migliore ipotesi»:

- Il livello di distorsione in $\hat{\beta}_1$ dipende dalla natura dell'errore di misura – questi due modelli sono casi speciali.
- Se a X_i viene aggiunto rumore puro, allora $\hat{\beta}_1$ è distorto verso zero. \rightarrow Solo la varianza cambia
- Il modello “migliore ipotesi” è estremo. In generale, se si pensa che vi sia un errore di misura, ci si dovrebbe preoccupare della distorsione da errore di misura.
- L’importanza potenziale della distorsione da errore di misura dipende dal modo in cui i dati vengono raccolti.
 - Spesso alcuni dati amministrativi (per esempio il numero di insegnanti in un distretto scolastico) sono molto accurati.
 - Spesso i sondaggi su argomenti sensibili (quanto guadagna?) presentano notevoli errori di misura.

Soluzioni alla distorsione da errori nelle variabili

1. Ottener dati migliori (spesso più facile a dirsi che a farsi).
2. Sviluppare un modello specifico del processo degli errori di misura. Questo è possibile solo se si sa molto sulla natura dell'errore di misura – per esempio, un sottocampione dei dati viene sottoposto a controlli incrociati usando dati amministrativi e le discrepanze vengono analizzate e modellizzate.
(Altamente specialistico; non ce ne occuperemo qui)
3. Regressione con variabili strumentali.

4. Distorsione da dati mancanti e selezione campionaria

Spesso i dati mancano. A volte i dati mancanti introducono distorsione, ma a volte no. È utile considerare tre casi:

1. I dati sono mancanti a caso.

2. I dati sono mancanti in base al valore di una o più X

3. I dati sono mancanti in parte in base al valore di Y o u

I casi 1 e 2 non introducono distorsione: gli errori standard sono più grandi di come sarebbero se i dati non fossero mancanti, ma $\hat{\beta}_1$ è non distorto

$$\hat{\beta}_1$$

Il caso 3 introduce la distorsione da “selezione campionaria”.

5. Distorsione da causalità simultanea

Finora si è ipotizzato che X causasse Y .
E se anche Y causa X ?

Esempio: effetto delle dimensioni delle classi

- Un basso STR porta a migliori punteggi nei test
- Ma si supponga che ai distretti con bassi risultati nei test vengano fornite risorse ulteriori: come risultato di un processo politico anch'essi avranno un basso STR
- Che significato ha tutto ciò per una regressione di $TestScore$ su STR ?

Distorsione da causalità simultanea: in equazioni

(a) Effetto causale su Y di X : $Y_i = \beta_0 + \beta_1 X_i + u_i$

(b) Effetto causale su X di Y : $X_i = \gamma_0 + \gamma_1 Y_i + v_i$

- Un grande u_i significa un grande Y_i , il che implica un grande X_i (se $\gamma_1 > 0$)
- Quindi $\text{corr}(X_i, u_i) \neq 0$
- Quindi $\hat{\beta}_1$ è distorto e inconsistente.
- *Esempio*: un distretto con risultati particolarmente negativi dato STR (u_i negativo) riceve risorse aggiuntive, che abbassano il suo STR ; quindi STR_i e u_i sono correlati

Soluzioni alla distorsione da causalità simultanea

1. Eseguire un esperimento casualizzato controllato. Siccome X_i viene scelto a caso dallo sperimentatore, non vi è feedback dalla variabile risultante su Y_i (ipotizzando una perfetta corrispondenza).
2. Sviluppare e stimare un modello completo di entrambe le direzioni di causalità. È l'idea alla base di molti macromodelli (per esempio la Federal Reserve Bank-USA). Questo *nella pratica è estremamente difficile*.
3. Usare regressione a variabili strumentali per stimare l'effetto causale di interesse (effetto di X su Y , ignorando l'effetto di Y su X).

Validità interna ed esterna quando la regressione è usata per le previsioni (Paragrafo 9.3)

- Previsione e stima degli effetti causali sono obiettivi piuttosto diversi.
- Per la previsione,
 - \bar{R}^2 importante (molto!)
 - La distorsione da variabili omesse non è un problema!
 - L'interpretazione dei coefficienti nei modelli di previsione non è importante – ciò che conta sono un buon adattamento e un modello che si possa ritenere “affidabile” per la propria applicazione
 - La validità esterna è fondamentale: il modello stimato con dati storici deve mantenersi valido per il futuro (immediato)
 - La previsione verrà trattata in seguito con i dati da serie storiche

10/11

Dis. di Cauchy-Schwarz

$$|E[X \cdot Y]| \leq \sqrt{E[X^2] \cdot E[Y^2]}$$

$$|E[X^2 \cdot Y^2]| \leq \sqrt{E[X^4] \cdot E[Y^4]}$$

Congiuntamente normali \rightarrow le ipotesi

$$(Y, X) \sim N^{\omega}(\quad)$$

$$Y|X \sim N(a + bX; h(x)) \rightarrow$$
 Non ipotesi le normalità congiunte

$$E[Y|X] = a + bX$$

$$\sqrt{V[Y|X]} = \text{costante}$$

$$\underline{X}_1, \underline{X}_2 \sim N(\underline{\mu}_{112}, \underline{\Sigma}_{112})$$

OBIETTIVO: $E[X_1 | X_2] = ?$

$$\underline{\mu}_1 = E[\underline{X}_1] = E_{\underline{X}_2}[E[\underline{X}_1 | \underline{X}_2]]$$

$$\underline{\Sigma}_{11} = \text{Var}[\underline{X}_1] = E_{\underline{X}_2}[\underbrace{\text{Var}[\underline{X}_1 | \underline{X}_2]}_{\underline{\Sigma}_{112}} + \underbrace{\text{Var}[E[\underline{X}_1 | \underline{X}_2]]}_{\underline{\mu}_{112}}]$$

$$E[X_1 | X_2] = \underline{\alpha} + A X_2 \rightarrow E[X_1 | X_2 = x] = \underline{\alpha} + A x$$

Modello multivariato normale

$$\underline{X}_1 = \underline{\alpha} + A \underline{X}_2 + \underline{v}$$

$$\text{con } E[\underline{v} | \underline{X}_2] = \underline{0} \quad (\underline{v} | \underline{X}_2) \sim N(\underline{0}; \Sigma_{vv})$$

$$Y_i = \beta_0 + \beta_1 X_i + \nu_i \quad \text{Var}(Y_i | X_i) = \text{Var}(\nu_i | X_i)$$

$$X_1 | X_2 \sim N(\underbrace{\alpha + b X_2}_{\bar{X}_1}; \sigma_{12}^2) \quad \text{con } v | X_2 \sim N(0; \sigma_{vv}^2)$$

$$\hat{X}_1 = \alpha + b X_2 \rightarrow \text{Prestazione}$$

Errore quadratico medio di prestazione

(Se v_i : condizioni ad altre variabili r. da le varianze)

$$\sigma_{112}^2 = \sigma_v^2 (1 - \beta^2) \leq \sigma_v^2 \left\{ \begin{array}{l} \text{Var}(X_1) \\ \beta = \text{Cov}(X_1, X_2) \end{array} \right.$$

$$\sigma_{112}^2 = \sigma_v^2 (1 - \beta^2) < \sigma_v^2 \text{ se } \beta \neq 0 \quad \beta \neq \pm 1$$

$$X_1 \sim N(\mu_1, \sigma_1^2) \quad \hat{X}_1 = E[X_1] = \mu_1$$

$$E[(X_1 - \hat{X}_1)^2] = E_{X_2} [E[(X_1 - \hat{X}_1)^2 | X_2]]$$

$$= E_{X_2} [\sigma_{112}^2] = \sigma_{112}^2$$

$$E[(X_1 - g(X_2))^2]$$

con $g(X_2) = E[X_1 | X_2]$ minima l'errore quadratico medio

$$\sigma_{112}^2 = \sigma_v^2 \rightarrow S_{X_1 X_2} = 0 \rightarrow X_1 \perp X_2 \text{ ortogonalità}$$

Lo Costante

$\neq E[X_1]$

Distribuzione Asintotica dello Stimatore

Convergenza in dist. S & W 18.2 IV ed.

$S_1, S_2, \dots, S_n, \dots$
 \rightarrow Succ. di variabili casuali
 $F_1(S), F_2(S) \dots F_n(S)$
 \rightarrow variabile reale
 $F_n(S) = \text{Prob}\{S_n \leq s\}$
 \hookrightarrow F_n di ripartizione associata

$$S_n \xrightarrow[n \rightarrow \infty]{d} S \sim F_{CS})$$

$$\lim_{n \rightarrow \infty} F_n(s) = F_{CS}) \quad \text{f.s.t.c. } F_{CS} \text{ cont.}$$

TLC multivariata

$$\underline{z}_i \sim \text{iid} \quad E[\underline{z}_i] = \underline{\Omega} \rightarrow E[\underline{z}_i \cdot \underline{z}_i^T] = \Sigma > 0$$

\downarrow def. pos. e finite

$$\bar{z}_n = \left(\frac{1}{n} \sum \underline{z}_i \right) \xrightarrow{P} \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad \begin{matrix} \text{Variabile degenera} \\ \rightarrow S \end{matrix}$$

\downarrow $\bar{z}_n \xrightarrow{d} S \sim N(F_{CS})$

\downarrow v.a. degenera

$$\text{Var}(\bar{z}_i) = \sigma_z^2 > 0 \text{ e finite}$$

$$F_n(S) \rightarrow F(S)$$

$$\sqrt{n} \left(\frac{1}{n} \sum \underline{z}_i \right) \xrightarrow[n \rightarrow \infty]{d} S \sim N(\underline{\Omega}; \Sigma)$$

\downarrow S_n

$$\Sigma = \text{AVAR}(S_n)$$

\hookrightarrow Varianza asintotica

$$\text{AVAR} \left(\frac{1}{n} \sum \underline{z}_i \right) = \left(\frac{1}{n} \right) \text{AVAR}(S_n) = \frac{1}{n} \Sigma$$

$$\left(\frac{1}{n} \sum \underline{z}_i \right) \underset{n \rightarrow \infty}{\sim} N(\underline{\Omega}; \frac{1}{n} \Sigma)$$

Applicazione del TLC al nostro caso: Asintotica normalità

$$\widehat{\beta} - \beta = (X'X)^{-1} \cdot \frac{n}{n} (X' \underline{\mu}) = (\frac{1}{n} X'X)^{-1} \cdot \frac{1}{n} I_n v \rightsquigarrow S_n$$

$\left(\frac{1}{n} \sum_{i=1}^n x_i x_i' \right)$ $\hookrightarrow I_n^t = I_n$

$\underline{v}_i = \underline{x}_i \cdot \underline{u}_i \sim \text{iid}$

$E[\underline{v}_i] = E[\underline{x}_i \cdot \mu_i] = \text{Cov}(\underline{x}_i, \mu_i) = 0$	$\text{Var}[\underline{v}_i] = E[\underline{v}_i \cdot \underline{v}_i'] = \sum_v$ è finito e assumiamo
	che sia def. pos.

$$S_n = \sqrt{n} \left(\frac{1}{n} \sum \underline{v}_i \right) \xrightarrow{d} S \sim N(\underline{0}, \Sigma_{\underline{v}})$$

$$\sqrt{n}(\widehat{\beta} - \beta) = \sqrt{n} (X'X)^{-1} \cdot \frac{n}{n} (X' \underline{\mu}) = (\frac{1}{n} X'X)^{-1} S_n$$

A_n

$$\rho\text{-lim } \frac{1}{n} X'X \stackrel{\rho\text{-lim}}{=} \left(\frac{1}{n} \sum \underline{x}_i \underline{x}_i' \right) = Q_X = E[X_i \cdot X_i']?$$

$$\beta = \widehat{\beta} + \frac{(\frac{1}{n} X'X)^{-1} S_n}{\sqrt{n}}$$

$$\rho\text{-lim } \frac{1}{n} X'X \stackrel{\rho\text{-lim}}{=} \left(\frac{1}{n} \sum \underline{x}_i \underline{x}_i' \right) = Q_X = E[X_i \cdot X_i']?$$

TH. DI SLUTSKY applicato (non è il caso generale)

Sia $S_n \xrightarrow{d} S \sim N(\underline{0}, \sigma^2)$ $a_n \xrightarrow{a}$ $a_n \cdot S_n \xrightarrow{d} a \cdot S \sim N(\underline{0}, a^2 \sigma^2)$

$$\rho\text{-lim } (\frac{1}{n} X'X)^{-1} = Q_X^{-1}$$

$A_n \xrightarrow{d} Q_X^{-1}$
 $S_n \xrightarrow{d} S \sim N(\underline{0}, \Sigma_{\underline{v}})$

$\left\{ \begin{array}{l} \widehat{\beta} - \beta \xrightarrow{d} Q_X^{-1} \cdot S \\ A_n \cdot S_n \xrightarrow{d} Q_X^{-1} \cdot S \sim N(\underline{0}, Q_X^{-1} \sum_v Q_X^{-1}) \end{array} \right.$

Asintotica distribuzione
dello stimatore

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{d} N(0; Q_x^{-1} \Sigma_{\varepsilon\varepsilon} Q_x^{-1})$$

$$\hat{\beta} \approx N(\beta; \frac{1}{n} Q_x^{-1} \Sigma_{\varepsilon\varepsilon} Q_x^{-1})$$

AVAR($\hat{\beta}$) $\rightarrow \frac{1}{n} Q_x^{-1} \Sigma_{\varepsilon\varepsilon} Q_x^{-1} \frac{1}{n}$

II/10

Distribuzione asintotica dello stimatore

X TLC

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{d} S \sim N(0; Q_x^{-1} \Sigma_{\varepsilon\varepsilon} Q_x^{-1})$$

NOTAZIONE ALTERNATIVA $\rightarrow \text{AVAR}(\sqrt{n}(\hat{\beta} - \beta)) = \sum \sqrt{n}(\hat{\beta} - \beta) \rightarrow \sum \hat{\beta} - (\frac{1}{n}) \sum \sqrt{n}(\hat{\beta} - \beta)$

Notazione per complessi grandi ma finiti $\rightarrow \hat{\beta} \underset{n \gg \infty}{\approx} N^{(k+1)}(\beta, \Sigma_{\hat{\beta}})$

Stimatore di White

$$\hat{Q}_x^{-1} \Sigma_{\varepsilon\varepsilon} \hat{Q}_x^{-1}$$

è lo stimatore consistente di $Q_x^{-1} \Sigma_{\varepsilon\varepsilon} Q_x^{-1}$ (STIMATORE SANDWICH)

$$\frac{1}{n} X'X \xrightarrow{P} Q_x \text{ def. pos. finita} \rightarrow Q_x^{-1}$$

$$\hat{Q}_x^{-1} \xrightarrow{P} Q_x^{-1}$$

$$\underline{v}_i = \underline{x}_i \cdot \mu_i$$

$$\sum \underline{v} = E[\underline{v}_i \cdot \underline{v}_i'] = E[\underline{x}_i \cdot \underline{x}_i' \mu_i^2]$$

$$\frac{1}{n} \sum \underline{x}_i \underline{x}_i' \mu_i^2 \rightarrow \sum \underline{v}$$

Non sono noti

\downarrow E uno stimatore consistente

$$\sum \hat{\underline{v}} = \frac{1}{n} \sum \underline{x}_i \underline{x}_i' \hat{\mu}_i^2 \rightarrow \sum \underline{v}$$

$$\hat{\sum}_{\underline{n}} (\hat{\beta} - \beta) \rightarrow \sum_{\underline{n}} (\hat{\beta} - \beta)$$

$$\hat{Q}_x^{-1} \sum \hat{\underline{v}} \hat{Q}_x^{-1}$$

$$\hat{\sum}_{\hat{\beta}} = \frac{1}{n} \sum_{\underline{n}} (\hat{\beta} - \beta)$$

Stima d. White

$$\hat{\sigma}_{ij, \hat{\beta}} = \hat{\sigma}_{\hat{\beta}_j}^2 \rightarrow SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}_{\hat{\beta}_j}^2}$$

Questi stimatori sono robusti sia se
c'è che se non c'è eteroschedasticità

$$\frac{1}{n} \sum \underline{x}_i \underline{x}_i' \hat{\mu}_i^2$$

$\xrightarrow{\text{ingret}}$ HCO

$$\left(\frac{1}{n - \text{ChH}} \right) \sum \underline{x}_i \underline{x}_i' \hat{\mu}_i^2$$

$\xrightarrow{\text{ingret}}$ HC1

} heteroschedasticity

Test T

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0; \Sigma_{\sqrt{n}}(\hat{\beta} - \beta))$$

$$\hat{\beta} = \begin{bmatrix} S_n & \hat{\beta}_0 \\ & \vdots \\ & \hat{\beta}_k \end{bmatrix}$$

Se voglio estrarre la seconda componente ($\hat{\beta}_1$) $\rightarrow \sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0; [\Sigma_{\sqrt{n}}(\hat{\beta} - \beta)]_{2,2})$

righe - colonne

$$\alpha = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \underbrace{\alpha' S_n}_{g(S_n) \text{ lineare}} \xrightarrow{d} \alpha' \cdot S \sim N(\alpha' 0; \alpha' \Sigma \alpha)$$

Sono equivalenti

Ese.

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0; \sigma_1^2)$$

$$\left(\frac{1}{\sigma_1^2}\right) \xrightarrow{P} \left(\frac{1}{\sigma_1^2}\right)$$

Trasformazione
lineare

Il secondo è uno stimatore

$$\sqrt{\frac{1}{\sigma_1^2}} \sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \left(\frac{1}{\sigma_1^2}\right) \cdot S \sim N(0, 1)$$

$$\left(\frac{1}{\sqrt{\sigma_1^2}}\right) \cdot S_n \xrightarrow{P} \left(\frac{1}{\sigma_1^2}\right) \cdot S \sim N(0, 1)$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{S.E.(\hat{\beta}_1)}$$

$$\frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\sigma_1}{\sqrt{n}}\right)}$$

$$S.E.(\hat{\beta}_1) = \sum \hat{\beta} = \frac{1}{n} \cdot \sum \sqrt{n}(\hat{\beta} - \beta)$$

Test d'ipotesi:

$$\begin{cases} H_0: \beta_1 = \beta_{1,0} \\ H_1: \beta_1 \neq \beta_{1,0} \end{cases}$$

$$t_{H_0}^{\text{act}} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\text{SE}(\hat{\beta}_1)} \stackrel{n \gg 0}{\sim} N(0, 1)$$

Test χ^2 chi-quadrato

$$\chi^2(m) \stackrel{\text{def}}{=} \sum_{i=1}^m z_i^2$$

$$z_i \sim N(0, 1)$$

$$z_i \perp z_j$$

In forme matriciali:

$$\chi^2(m) \stackrel{\text{def}}{=} \underline{z}' \underline{z} = \underline{z}' \cdot \mathbf{I}_m \cdot \underline{z}$$

$$\underline{z} \sim N(0; \mathbf{I}_m)$$

Th:

Se $\underline{x} \sim N(\mu; \Sigma) \stackrel{\text{def. pos.}}{=} \underline{z}$

Allora $(\underline{x} - \mu)' \Sigma^{-1} (\underline{x} - \mu) = \underline{z}' \underline{z} \stackrel{\text{m gradi di libertà}}{\sim} \chi^2(m)$

Con questa trasformazione faccio 2 passaggi:

I) trasformo le x_i in z_i std.

II) faccio il prodotto interno (somma dei quadrati)

Ottengendo così una chi quadro con m gradi di libertà

$$\text{Es. } \sqrt{n} (\hat{\beta} - \beta) \xrightarrow{d} S \sim N(0; \Sigma) \quad \underbrace{\Sigma}_{S_n (k+1) \times 1} \quad \underbrace{\Sigma}_{F_S(S)} \xrightarrow{(k+1) \times (k+1)} g(S) \sim F_{g(S)}$$

$$g(S_n) \text{ e-continua} \quad S' \Sigma^{-1} S \sim \chi^2_{(k+1)}$$

$$\text{Se } \hat{\Sigma} \xrightarrow{d} \Sigma \rightarrow \hat{\Sigma}' \xrightarrow{d} \Sigma'$$

$$S_n' \hat{\Sigma}'^{-1} S_n \xrightarrow{d} S' \Sigma'^{-1} S \sim \chi^2_{(k+1)}$$

F Test

$$\text{Es. } H_0: \begin{cases} \beta_0 = 0 \\ \beta_1 = 0 \end{cases} \quad H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_0: \beta_1 + \beta_2 = 1 \quad H_0: R\beta = r$$

$$R\beta = r \quad R\beta \xrightarrow[\text{metrice allo stimatore}]{} R\hat{\beta}$$

$$q \times (R+1) \quad (R+1) \times 1$$

$$q \leq (k+1)$$

$$\text{con } \text{range}(R) = q$$

$$R \left[\sqrt{n} (\hat{\beta} - \beta) \right] \xrightarrow{d} R - S \sim N(0; R\Sigma R')$$

$$\tilde{S}_n' (R\Sigma R')^{-1} \tilde{S}_n \xrightarrow{d} \chi^2_q$$

$$\frac{\tilde{S}_n' (R\hat{\Sigma} R')^{-1} \tilde{S}_n}{q} \xrightarrow{d} \frac{\chi^2_q}{q} \sim F(q, \infty)$$

$$S_n \xrightarrow{d} N(0; \Sigma)$$

$$(RS_n)' (R\hat{\Sigma} R') (RS_n) \xrightarrow{d} \sum_{q \times 1} \sim \chi^2_q \xrightarrow{\text{regolare}}$$

$$\sqrt{n} (\hat{\beta} - \beta)^T (R^T R)^{-1} (R^T \hat{\beta})$$

$$\hat{\Sigma} = \hat{\Sigma}_{\hat{\beta}} = (R \hat{\beta} - R \beta)^T (R \sum_{\beta} R^T)^{-1} (R \hat{\beta} - R \beta)$$

$$\hat{\beta} \sim N(\beta, \Sigma_{\hat{\beta}}) \rightarrow R \hat{\beta} \sim N(R\beta, R \sum_{\hat{\beta}} R^T)$$

$$(R \hat{\beta} - R \beta)^T (R \sum_{\hat{\beta}} R^T)^{-1} (R \hat{\beta} - R \beta) \sim \chi^2(q)$$

$$F_{q,m} = \frac{\chi^2(q)/q}{\chi^2(m)/m} = \frac{\chi^2(q)}{\chi^2(m)}$$

$$\text{Se } m \rightarrow +\infty \quad \frac{\chi^2(m)}{m} \xrightarrow[m \rightarrow +\infty]{\text{c.d.}} 1$$

$$F_{q,\infty} = \chi^2(q)/q$$

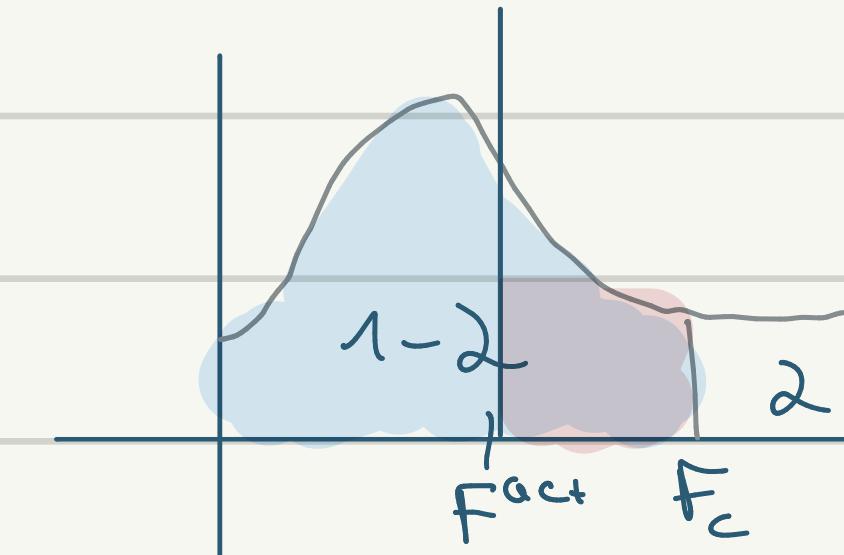
$$\text{La statistica } F \text{ esatta } \leftarrow F_{q,n-(k+1)} \xrightarrow[n \rightarrow +\infty]{\text{d.f.}} F_{q,\infty}$$

Non imponendo l'omoschedasticità e la normalità non abbiamo statistiche esatte ma solo assintotiche

$$t_{n-(k+1)} \xrightarrow[n \rightarrow +\infty]{\text{d.f.}} Z \sim N(0,1)$$

$$\frac{W}{\sigma^2} \xrightarrow{d} \frac{1}{\sigma^2} S \sim F_{q,\infty}$$

Sostituisco $\underline{\beta}$ in $R\underline{\beta}$ ($H_0: R\underline{\beta} = \underline{r}$)



$$\rightarrow F_0 = F^{\text{act}} = \left(\frac{1}{\sigma}\right) (R\hat{\beta} - \underline{r})' (R\hat{\Sigma}_{\beta} R')^{-1} (R\hat{\beta} - \underline{r}) \overset{n \gg 0}{\sim} F_{q,\infty}$$

p-value = $\text{prob}_{H_0} \{ F_{q,\infty} > F^{\text{act}} \}$ p-value < 2 → Rif. H_0

$$t_m^2 = F_{1,m} \quad \chi^2 = F_{1,\infty} = \chi^2(1)$$

APPROCCIO DI WAID: Stimo il modello non ristretto

APPROCCIO DEI MOLTIPLICATORI DI LAGRANGE: Stimo il modello sotto H_0 (ristretto)

APPROCCIO DEL RAPPORTO DELLE VEROSIMIGLIANZE: Stimo tutti e 2 modelli

Modello di reg. lin. $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \nu_i$

$$1) E[\nu_i | X_{1i}, \dots, X_{ki}] = 0 \rightarrow \text{Cov}[X_{1i}, X_{ji}] = 0 \quad \forall j$$

$\hat{\beta}_{j,OLS}$ sono interpretabili come effetti causali parziali

↪ ipotesi meno forte

$$1*) \text{ Indip. in media condizionata } Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \nu_i$$

$$E[\nu_i | X_{1i}, X_{2i}] \neq 0$$

$$E[\nu_i | X_{1i}, X_{2i}] = E[\nu_i | X_{2i}] \neq 0 \rightarrow \text{perché meglio stimare } \beta_2:$$

↪ Dipende solo da X_{2i} e non da X_{1i}

$$g(X_{2i}) = \gamma_0 + \gamma_1 \cdot X_{2i}$$

p.v. lineare

$$\nu_i = \gamma_0 + \gamma_1 X_{2i} + \nu_i$$

errori

$$\text{Se } \nu_i = \mu_i - E[\nu_i | X_{1i}, X_{2i}] \rightarrow E[\nu_i | X_{2i}] = 0 \quad \Leftrightarrow E[\nu_i | X_{1i}, X_{2i}] = 0$$

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \gamma_0 + \gamma_1 X_{2i} + \nu_i && \text{con } E[\nu_i | X_{1i}, X_{2i}] = 0 \\ Y_i &= (\underbrace{\beta_0 + \gamma_0}_{\delta_0}) + \underbrace{\beta_1 X_{1i}}_{S_1} + (\underbrace{\beta_2 + \gamma_1}_{S_2}) X_{2i} + \nu_i && \text{errori degli errori} \end{aligned}$$

$(\hat{\delta}_0, \hat{\beta}_1, \hat{\delta}_2)$ sono corrette consist. e asint. normali

$$E[\nu_i | X_{1i}, X_{2i}] = 0$$

$$\begin{aligned} \hat{\delta}_2 &\xrightarrow{c} \delta_2 = \beta_2 + \gamma_1 \\ \hat{\beta}_1 &\xrightarrow{P} \beta_1 \neq \beta_2 \end{aligned} \quad (\hat{\delta}_0, \hat{\beta}_1, \hat{\delta}_2) \text{ sono corrette consistenti e asintot. normali}$$

Test per le verifiche sui modelli

Chap. 5 - Brooks

Test sull'eteroschedesticità

$$H_0: \nu_i \text{ omoschedestici} \rightarrow \text{Var}(\nu_i | X_i) = \sigma_{\nu_i}^2, \forall i$$

$$\text{Var}(\nu_i | X_i) = \text{Var}(\nu_i)$$

$$E[\nu_i^2 | X_i] \quad E[\nu_i^2] = \sigma_{\nu_i}^2$$

$$E[\nu_i^2] = E[E[\nu_i^2 | X_i]] = k = \sigma_{\nu_i}^2$$

TEST DI WHITE

È più generale per quando non sappiamo come varia l'eteroschedesticità

$$E[\nu_i^2 | X_i] \rightarrow \nu_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \nu_i$$

$$H_1: \text{Var}(\nu_i | X_i) = h(X_i) = \lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{1i}^2 + \lambda_4 X_{2i}^2 + \lambda_5 X_{1i} X_{2i}$$

$$\text{Sotto } H_0: \begin{cases} \lambda_1 = \lambda_2 = \dots = \lambda_5 = 0 \\ \lambda_0 = \sigma_{\nu_i}^2 \end{cases}$$

Se non lo metto la varianza sarebbe nulla

$$\nu_i^2 = h(X_i) + v_i \text{ t.c. } E[v_i | X_i] = 0$$

$$\hat{\mu}_i \rightarrow \hat{\mu}_i^2 \quad \text{OLS } \hat{\mu}_i^2 = h(X_i) + \tilde{\nu}_i \rightarrow \text{Regressione Auxiliaria} \\ (n \cdot R^2) \approx \chi^2(5)$$

mi serve solo per il calcolo del test

16/10

Test di Breush-Pagan

Meno generale del Test di White

$$\text{Var}(\hat{\mu}_i | X_i) = \sigma^2 h(X_i' \cdot \underline{\lambda})$$

\rightarrow Fz. auxiliaria. Non è necessario conoscerla
 $h(\underline{\lambda}) = \ell$

$H_0: \underline{\lambda} = 0$ E' basato sul metodo dei moltiplicatori di Lagrange

$$\hat{\mu}_{\text{scaduti}}^2 = \lambda_0 + X_i' \cdot \underline{\lambda} + \varepsilon_i$$

$$\text{B-P} \xrightarrow{H_0} \chi^2_q$$

Test di White

Si estendono le ipotesi.

$$\hat{\mu}_i^2 = \lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{3i}^2 + \lambda_4 X_{4i}^2 + \lambda_5 X_{1i} X_{2i} + \varepsilon_i$$

$$H_0: \lambda_1 = \dots = \lambda_5 = 0$$

\rightarrow Al crescere del numero dei regressori, aumenta esponenzialmente il numero di gradi vincolati, riducendo la precisione delle stime.

$$n \cdot R^2 \xrightarrow{H_0} \chi^2_5$$

Dimensione campionaria ✓ $n \cdot R^2$ del modello

In GRETL ci sono versioni ristrette del Test, con maggiori ipotesi e più potenti

Se ai modelli OLS (con le prime 4 ip.) aggiungiamo anche un'ulteriore ipotesi:
 OMOSCHEDASTICITÀ 5) $\text{Var}(\underline{\nu}|X) = \sigma^2_n I_n \rightarrow \text{Var}(\underline{\nu}|X) = \sigma^2_n I_n$

Quindi se valgono le ip. 1)-5) \rightarrow vale condiz. αX , il th. di GAUSS-MARKOV:

TH. di GAUSS-MARKOV

Se:

- 1) $E[\underline{\nu}|X] = 0$
- 2) $\text{Var}[\underline{\nu}|X] = \sigma^2_n I_n$
- 3) $\text{range}(X) = K+1$

Allora:

Subordinatamente a X , $\hat{\beta}_{OLS}$ è BLUE ossia nella classe degli stimatori lineari e corretti, lo stimatore OLS è il più efficiente

Best Linear Unbiased Estimator

$$\hat{\beta}_{OLS}|X \quad f_{\underline{\nu}} \text{ lineare vettoriale di } \underline{\nu} \quad e \quad E[\hat{\beta}_{OLS}|X] = \underline{\beta}$$

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'\underline{\nu} = A\underline{\nu} \quad \text{con } A = A(X)$$

Corretto

Quindi se $\tilde{\beta}$ un altro stimatore di $\underline{\nu}$ t.c. $\tilde{\beta} = B \cdot \underline{\nu}$ con $B = B(X) \neq A(X)$ e

$$E[\tilde{\beta}|X] = \underline{\beta} \rightarrow E[\tilde{\beta}] = \underline{\beta}$$

$$\text{Allora } \text{Var}(\hat{\beta}_{OLS}) \leq \text{Var}(\tilde{\beta})$$

$$\text{Var}(\hat{\beta}_{OLS}|X) = \Sigma \rightarrow (\sigma^2_n (X'X)^{-1})$$

$$\text{Var}(\tilde{\beta}|X) = \tilde{\Sigma}$$

$$(\tilde{\Sigma} - \Sigma) \geq 0$$

semi-def. positive

Dico dimostrato

$$\tilde{\Sigma} = \text{Var}(\underline{\nu}|X)$$

$$A \cdot \text{Var}(\underline{\nu}|X) \cdot A' = A \cdot \sigma^2_n I_n \cdot A' = \sigma^2_n \cdot A A' = \sigma^2_n (X'X)^{-1} (X'X) X'$$

$$= \sigma^2_n (X'X)^{-1} (X'X) \cdot (X'X)^{-1} = \sigma^2_n (X'X)^{-1}$$

I

$$\mathbf{D} = \tilde{\Sigma} - \Sigma$$

$\underline{c}' \mathbf{D} \underline{c} \geq 0 \rightarrow \mathbf{D}$ è semi-def. positive

$$\underline{c}' \tilde{\beta} \\ \text{Var}(\underline{c}' \tilde{\beta} | \mathbf{x}) = \underline{c}' \tilde{\Sigma} \underline{c}$$

$$\underline{c}' \hat{\beta}_{OLS} \\ \text{Var}(\underline{c}' \hat{\beta}_{OLS}) = \underline{c}' \Sigma_{\underline{c}}$$

$$+ c \neq 0, \quad \underline{c}' \tilde{\Sigma} \underline{c} - \underline{c}' \Sigma \underline{c} = \underline{c}' \mathbf{D} \underline{c} \geq 0 ? \quad \begin{cases} > \\ = \end{cases} \quad \begin{array}{l} \text{per } \tilde{\beta} \neq \beta \\ \text{per } \tilde{\beta} = \beta \end{array}$$

$$f_j \quad \text{Var}(\hat{\beta}_{j, OLS}) < \text{Var}(\tilde{\beta}_j)$$

Importante semplificazione

Se c'è omosched.

$$1) \rightarrow \sqrt{n} (\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(\underline{0}, Q_x^{-1} \Sigma_v Q_x^{-1}) = N(\underline{0}, \sigma_\nu^2 Q_x^{-1})$$

$$5) \quad \text{Var}(\mu_i | \mathbf{x}_i) = \sigma_\nu^2 = E[\nu_i^2 | \mathbf{x}_i]$$

$$\sum_v = E[\nu_i - \bar{\nu}_i]^2 = E[\mathbf{x}_i' \mathbf{x}_i \cdot \nu_i^2] = E[\mathbf{x}_i' E[\mathbf{x}_i \cdot \nu_i^2 | \mathbf{x}_i]] = E[(\mathbf{x}_i' \mathbf{x}_i) \cdot E[\nu_i^2 | \mathbf{x}_i]] = \sigma_\nu^2 E_{\mathbf{x}_i}[\mathbf{x}_i' \mathbf{x}_i]$$

$$\bar{\nu}_i = \mathbf{x}_i' \nu_i$$

$$Q_x^{-1} \Sigma_v Q_x^{-1} = Q_x^{-1} (\sigma_\nu^2 Q_x) - Q_x^{-1} = \sigma_\nu^2 \cdot Q_x^{-1}$$

Devo stimare in modo consistente $\sigma_\nu^2 \cdot (Q_x)^{-1}$

$$\left(n - k - 1 \right) \sum \hat{\nu}_i^2 = SER^2 \xrightarrow{d} \sigma_\nu^2$$

$$\frac{1}{n} \mathbf{x}' \mathbf{x} \xrightarrow{d} Q_x$$

$$(\frac{1}{n} \mathbf{x}' \mathbf{x})^{-1} \xrightarrow{d} Q_x^{-1}$$

$$SER^2 \left(\frac{1}{n} \mathbf{x}' \mathbf{x} \right)^{-1} \xrightarrow{d} \sigma_\nu^2 Q_x^{-1} = \text{AUXAR}(\sqrt{n} (\hat{\beta} - \beta))$$

$$\hat{\beta} \approx N(\beta, \underbrace{\frac{1}{n} \sigma_\nu^2 \cdot Q_x^{-1}}_{\Sigma \hat{\beta}})$$

$$\hat{\Sigma}_{\beta} = \left(\frac{1}{n}\right) \cdot SER^2 \left(\frac{1}{n} X' X\right)^{-1} = SER^2 (X' X)^{-1}$$

Stimatore consistente di Σ_{β} (solo nel caso di omoschedasticità)

$$Var(\hat{\beta}|X) = \sigma_{\beta}^2 (X' X)^{-1}$$

\hookrightarrow Se è omosch.
 $\hookrightarrow SER^2$

Le stime consist. di $Var(\hat{\beta}|X)$ e $\hat{\Sigma}_{\beta} = SER^2 \cdot (X' X)^{-1}$

Test di Normalità di SARAUE - BERG

H_0 : Normalità $\rightarrow \begin{cases} S=0 \\ K-3=0 \end{cases} \rightarrow$ indice di simmetria con ipotesi: $Y_i \sim iid N(\mu; \sigma^2)$

$$\hat{S} = \frac{1}{n} \sum \frac{(Y_i - \bar{Y})^3}{\hat{\sigma}_u^3} \rightarrow 0$$

$$\sqrt{n} \cdot \begin{bmatrix} \hat{S} \\ (\hat{K}-3) \end{bmatrix} \xrightarrow[H_0]{n \rightarrow +\infty} N([0], [6 \ 0 \ 0 \ 24])$$

$$\text{Asint. indip.} \quad \hat{S} \xrightarrow{\alpha} (K-3)$$

$$(\hat{K}-3) \xrightarrow{\alpha} 0$$

\rightarrow Somme di quadrati di normali

$$\text{Statistica test} \quad \hat{S}' \begin{bmatrix} 1/6 & 0 \\ 0 & 1/24 \end{bmatrix} \hat{S} \xrightarrow[H_0]{n \rightarrow +\infty} \chi^2_2$$

$\underbrace{\text{normalized la varianza}}$

$$\frac{1}{6} (\sqrt{n} (\hat{S} - 0))^2 + \frac{1}{24} (\sqrt{n} (\hat{K}-3))^2 = n \left(\frac{\hat{S}}{\sqrt{6}} \right)^2 + n \left(\frac{(\hat{K}-3)}{\sqrt{24}} \right)^2$$

$\xrightarrow{d} N(0, 1) \quad \xrightarrow{d} N(0, 1)$

$$J-B = n \left[\left(\frac{\hat{S}}{\sqrt{6}} \right)^2 + \left(\frac{(\hat{K}-3)}{\sqrt{24}} \right)^2 \right] \xrightarrow{H_0} \chi^2_2$$

17/10

GRET

18/10

Assunzioni del modello

 y

1) - 4) classiche

5) omoschedasticità

$$E[y|X] = X\beta$$

$$\text{Var}[y|X] = \sigma^2_n I_n$$

$$6) \nu | X \sim N(0; \sigma^2_n I_n) \rightarrow y | X \sim N(X\beta; \sigma^2_n I_n)$$

$$\hookrightarrow \nu \sim N(0; \sigma^2_n I_n)$$

$$\hat{\beta}_{OLS} | X \sim N(\beta; \sigma^2_n (X'X)^{-1})$$

 $(k+1) \times 1$

$$\hat{\beta}_{ML} = \hat{\beta}_{OLS}$$

$$t^{act} \xrightarrow{\text{f. t.}} t_{n-k-1} \xrightarrow{n \rightarrow +\infty} z \sim N(0, 1)$$

$$F^{act} \xrightarrow{\text{f. F.}} F_{q, n-k-1} \quad q \text{ vincoli}$$

 $\hat{\beta}_{ML}$: Maximum Likelihood

$$L = f(y|X; \beta, \sigma^2_n) = N(X\beta; \sigma^2_n I_n) \quad \forall y \in \mathbb{R}^n$$

 \hookrightarrow f.z. di verosimiglianza

dat.

dati del campione

$$L(\beta, \sigma^2_n; y, X) \stackrel{f}{=} f(y|X; \beta, \sigma^2_n)$$

Cerco il massimo valore della f.z. L attraverso i parametri:

$$\max_{\{\beta; \sigma^2_n\}} L(\beta, \sigma^2_n; y, X) = \exp \left\{ -\frac{1}{2\sigma^2_n} (y - X\beta)' (y - X\beta) \right\}$$

$$\begin{cases} \hat{\beta}_{ML} = \hat{\beta}_{OLS} \\ \hat{\sigma}^2_{ML} = \frac{1}{n} \sum_{i=1}^n \hat{\nu}_{OLS}^2 \end{cases}$$

 $\hookrightarrow E$ distretto me consistente

$$X \sim \text{I.I.D. } \mathcal{N}(\underline{\beta}, \Sigma) \quad \hat{\beta} = \underline{\beta} + \Sigma^{-1} X \quad \hat{\beta} \sim N(\hat{\beta}, \Sigma^{-1})$$

Metodo 1

$$\underline{\phi} = g(\underline{\beta})$$

$\underline{\phi}$ vincoli non lineari:

$$g(\hat{\beta}) \rightarrow g(\beta) = \underline{\phi} \quad \text{e } g(\cdot) \text{ continua, derivabile e con derivate continue}$$

Matrice Jacobiana

$$J = \begin{bmatrix} \frac{\partial g_1(\cdot)}{\partial \beta_1} & \dots & \frac{\partial g_1(\cdot)}{\partial \beta_k} \\ \vdots & & \vdots \\ \frac{\partial g_q(\cdot)}{\partial \beta_1} & \dots & \frac{\partial g_q(\cdot)}{\partial \beta_k} \end{bmatrix}_{q \times k}$$

Esegui l'approssimazione di Taylor del I ordine in un intorno $\hat{\beta} = \beta$

$$\hat{\phi} = g(\hat{\beta}) = g(\beta) + J(\beta) \cdot (\hat{\beta} - \beta) + \varepsilon$$

$$\sqrt{n}(\hat{\phi} - \phi) = J(\beta) \cdot \sqrt{n}(\hat{\beta} - \beta) + \sqrt{n} \cdot \varepsilon \xrightarrow[n \rightarrow \infty]{P=0} 0$$

$\varepsilon \sim N(0, \Sigma)$
converge in prob. e
in distrib.

$$\hat{\phi} \underset{n \rightarrow \infty}{\sim} N(\phi; \frac{1}{n} \cdot \sum_{\hat{\phi}} (\hat{\phi} - \phi)^2)$$

$$\sum_{\hat{\phi}} = \text{AVAR}(\sqrt{n}(\hat{\phi} - \phi)) = J(\beta) \cdot \Sigma \cdot J(\beta)^T$$

$$\hat{\Sigma}_{\hat{\phi}} = \left(\frac{1}{n} \right) J(\beta) \sum \hat{\beta} \hat{\beta}^T$$

$$\begin{aligned} \hat{\phi} &= \hat{\beta}_1^2 \\ g'(\beta_1) &= 2 \cdot \beta_1 \\ g'(\hat{\beta}_1) &= 2 \cdot \hat{\beta}_1 \end{aligned}$$

$$\begin{aligned} \hat{\phi} &\sim N(\phi; 2\hat{\beta}_1 \cdot \text{Var}(\hat{\beta}_1) \cdot 2\hat{\beta}_1) \\ &\approx N(\phi, 4\hat{\beta}_1^2 \text{Var}(\hat{\beta}_1)) \end{aligned}$$

$$\text{Var}(\hat{\phi})$$

$$\text{Var}(\hat{\phi}) = 4\hat{\beta}_1^2 \text{Var}(\hat{\beta}_1)$$

$$S.E.(\hat{\phi}) = \sqrt{Var(\hat{\phi})} - 2|\hat{\beta}_1| S.E.(\hat{\beta}_1)$$

$$t = \frac{\hat{\phi} - 6}{S.E.(\hat{\phi})}$$

Modello di Solow (1956)

Modello di crescita economica esogeno

Ci sono 2 tipi di fasi \rightarrow di transizione

\rightarrow Pro-capite

\hookrightarrow di equilibrio

Il PIL è determinabile, al netto dell'avanzamento tecnologico, del tasso di accumulazione del capitale e del tasso di crescita della popolazione.

Modello pensato nel continuo:

$$P/L \rightarrow Y(t) = K(t)^{\alpha} (A(t)L(t))^{1-\alpha} \quad 0 < \alpha < 1$$

$$\ln \left[\frac{Y(t)}{L(t)} \right] = \ln A(0) + g \cdot t + \frac{\alpha}{1-\alpha} \ln (C_S) - \frac{\alpha}{1-\alpha} \ln (n + g + \delta)$$

\hookrightarrow PIL pro capite \hookrightarrow Livello delle tecnologie + un errore specifico (shock) di agi-peise $\ln A(0) = \alpha + \varepsilon$

$$\ln \left(\frac{Y}{L} \right) = \underbrace{\alpha}_{\beta_0} + \underbrace{\frac{\alpha}{1-\alpha}}_{\beta_1} \ln (C_S) - \underbrace{\frac{\alpha}{1-\alpha}}_{\beta_2} \ln (n + g + \delta) + \underbrace{\varepsilon}_{\nu}$$

20110

Metodo delle variabili strumentali Cap. 3 del libro

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Problema di endogenicità

I residui OLS sono non correlati con le var. indip. → $\frac{1}{n} \sum \hat{\mu}_j X_{ij} = 0$
ortogonali

↳ Non sempre è così, però: → $\text{Cov}(X_{1i}, \mu_i) \neq 0$ X_{1i} è un regressore endogeno

Simultanea causalità

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + u_i \\ X_i = \gamma_0 + \gamma_1 Y_i + v_i \end{cases} \quad \text{Reverse causality}$$

Modello a equazioni simultanee
(modello strutturale)

Modello di regressione lineare

$\text{Cov}(u_i, v_i) \neq 0 \rightarrow$ Anche se è 0 il problema dovrebbe persistere

Se $\text{Cov}(u_i, v_i) = 0$

$$\begin{aligned} \text{Cov}(x_i, u_i) &= \text{Cov}(\gamma_0 + \gamma_1 Y_i + v_i, u_i) = 0 + \gamma_1 \text{Cov}(Y_i, u_i) + \text{Cov}(v_i, u_i) \underset{\neq 0}{=} 0 \\ &\underset{|}{=} \gamma_1 \text{Cov}(\beta_0 + \beta_1 X_{1i} + u_i, u_i) = \gamma_1 (\beta_1 \text{Cov}(X_{1i}, u_i) + \sigma_u^2) \\ &\underset{|}{=} \gamma_1 \beta_1 \text{Cov}(X_{1i}, u_i) + \gamma_1 \sigma_u^2 = \frac{\gamma_1 \sigma_u^2}{1 - \gamma_1 \beta_1} \neq 0 \end{aligned}$$

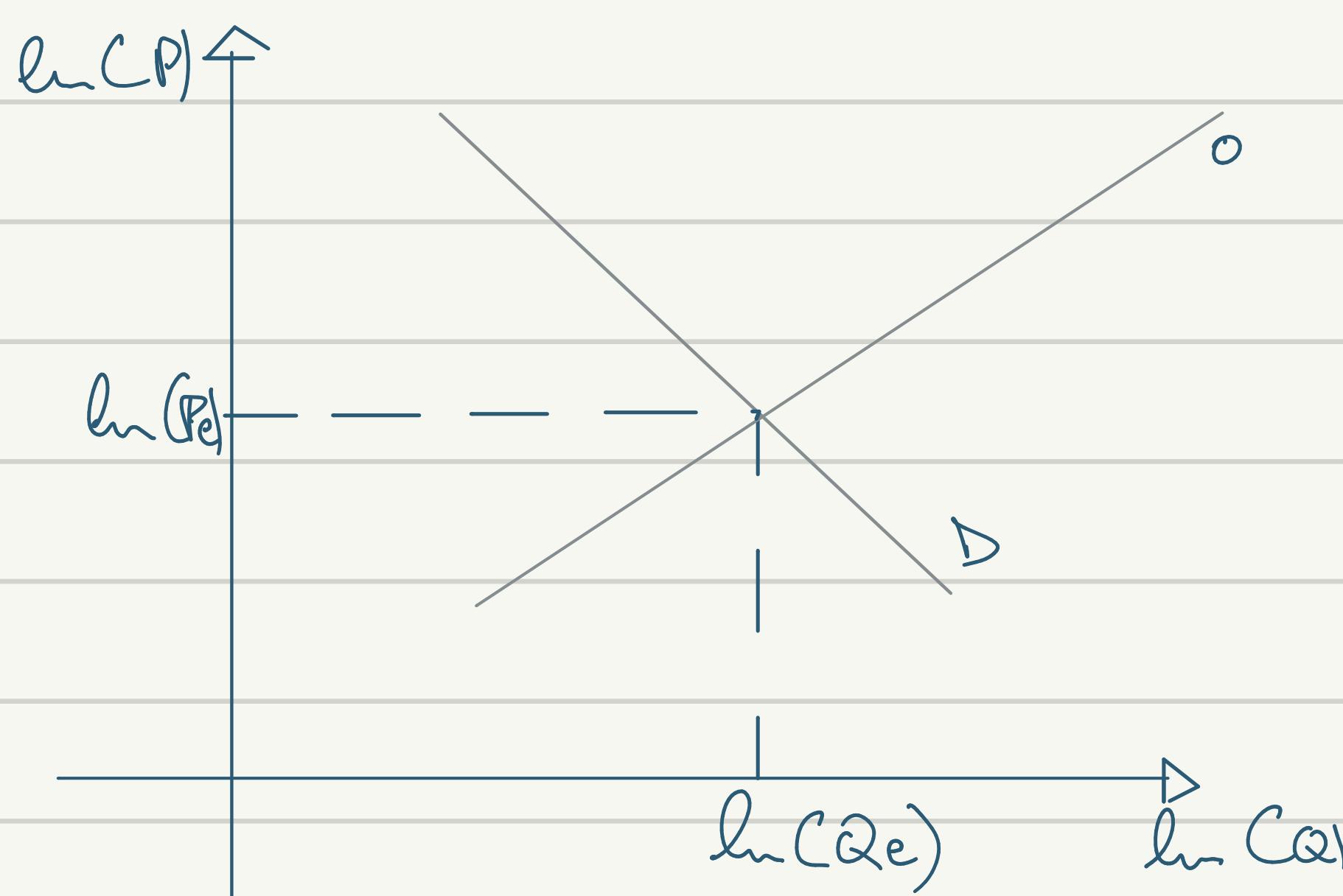
$$Q_D = A_0 P^{\beta_1} \exp(u) \rightarrow \begin{cases} \ln(Q_D) = \beta_0 + \beta_1 \ln(P) + u \\ \ln(Q_D) = \delta_0 + \delta_1 \ln(P) + v \\ Q_D = Q_D \end{cases}$$

$$Q_D = B_0 P^{\delta_1} \exp(v) \rightarrow$$

C'è simultanea causalità?

$$\ln(P) = \left(-\frac{\delta_0}{\delta_1} \right) + \frac{1}{\delta_1} \ln(Q_D) + \left(-\frac{v}{\delta_1} \right)$$

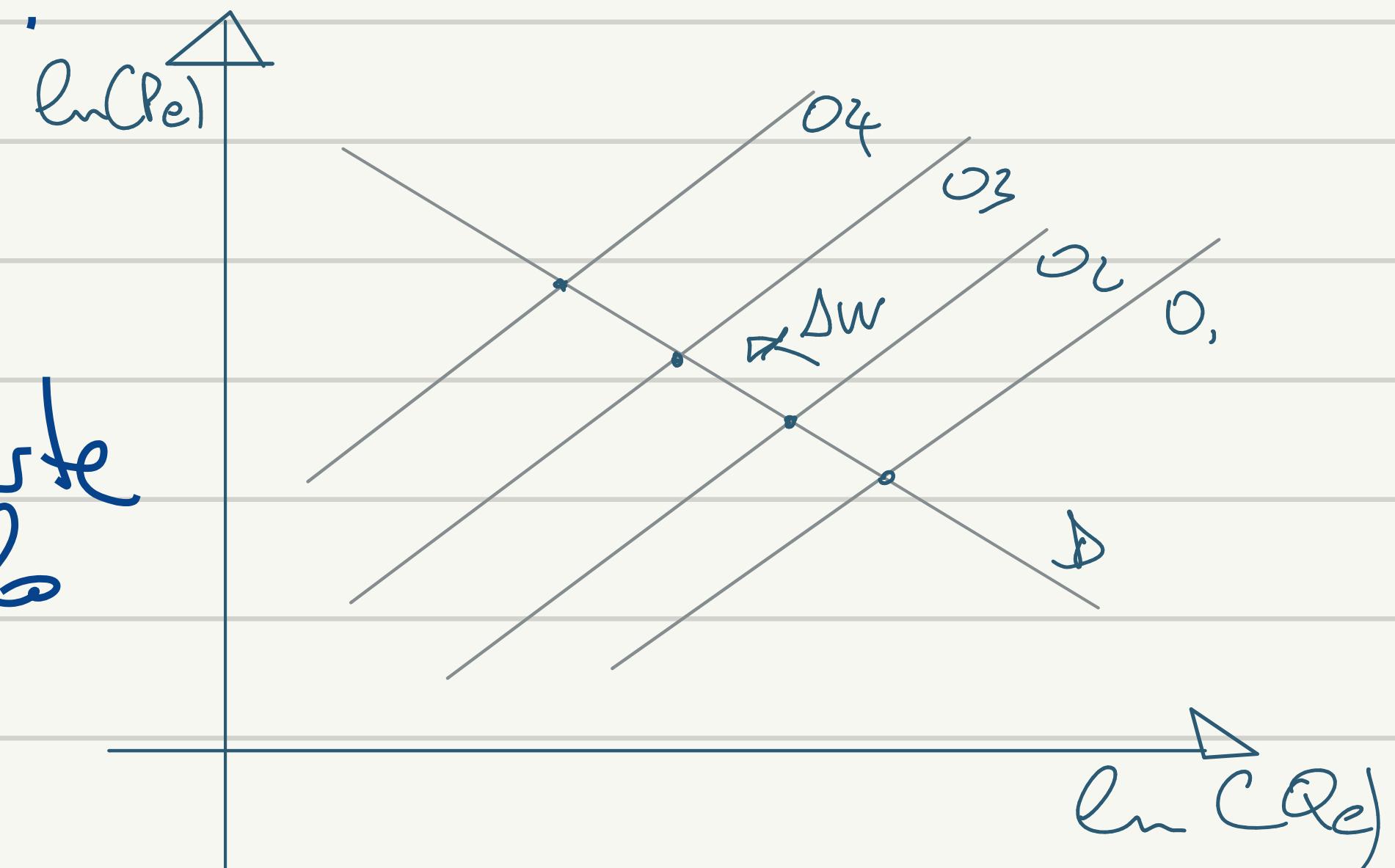
$$\ln(P) = \delta_0 + \delta_1 \ln(Q_D) + w$$



Nel mercato reale avrò più offerte
e dovrò trovare quelli variabili: muovo solo
l'offerta e non la domanda

La curva di domanda cambia quando
c'è una variazione nell'errore dell'eq.

β_1 e δ_1 identificabili. parametri strutturali, non sono



$\Delta w \leftrightarrow \delta z$

\hookrightarrow Stimolo

$\Delta x \leftrightarrow \Delta z \quad \text{cov}(x, z) \neq 0$

$\text{cov}(z, v) = 0 \rightarrow z \text{ è } v\text{-esogene}$

// Confusione da rivedere
C.a. 20/10 10:20

Stimolo 2:

(1) esogeno, ossia $\text{cov}(z, v) = 0$

(2) rilevato $\rightarrow \text{cov}(x, z) \neq 0$

$\Delta z \nrightarrow \Delta v$

$\Delta z \rightarrow \Delta x = \Delta \ln(P_e) \rightarrow$ fa variare il prezzo indipendentemente

X è correlato con v : $\text{cov}(x, v) \neq 0$ perciò x non è correlato

con v

$\Delta x = \Delta x^* + \Delta w$

endogene esogene endogene

$X = \underbrace{\pi_0 + \pi_1 z}_\text{Parte esogene di X} + w$

\rightarrow Perche' z è esogene

$\text{cov}(\pi_0 + \pi_1 z, v) = 0$

$\hat{X} = \hat{\pi}_0 + \hat{\pi}_1 z + w$

Sostituisce la parte endogene con l'esogene

Metodo delle variabili strumentali (instrumental variables)

Sottocaso del metodo dei minimi quadrati al secondo (TSLS o 2SLS)

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

Z_i :
(1) esogeno rispetto a u_i :
(2) rilevante per X_i

1° Studio

OLS

$$X_i = \bar{\Pi}_0 + \bar{\Pi}_1 Z_i + w_i$$

$$\rightarrow \text{Cov}(\hat{X}_i, u_i) = 0$$

$$\hat{X}_i = \hat{\bar{\Pi}}_0 + \hat{\bar{\Pi}}_1 Z_i$$

2° Studio

OLS

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i \quad \text{Posso farlo perché ora è esogeno}$$

$$\hat{\beta}_{1, 2SLS} \xrightarrow{P} \beta_1 \quad \text{distorto ma consistente}$$

23/10

Capitolo 12

Regressione con variabili strumentali

James H. Stock
Mark W. Watson

INTRODUZIONE ALL'ECONOMETRIA

Quarta edizione
Edizione italiana a cura di
Franco Peracchi

Pearson Learning Solution

Codice di accesso a MyLab

Aula virtuale

Risorse multimediali

Test ed esercizi

Autovalutazione

Pearson eText

ALWAYS LEARNING

PEARSON

Regressione IV: perché?

Tre importanti minacce alla validità interna sono:

- Distorsione da variabili omesse per una variabile correlata con X ma inosservata (perciò non può essere inclusa nella regressione) e per cui vi sono variabili di controllo inadeguate;
- Distorsione da causalità simultanea (X causa Y , Y causa X);
- Distorsione da errori nelle variabili (X è misurata con errore)

Tutti e tre i problemi comportano $E(u|X) \neq 0$.

- La regressione con variabili strumentali può eliminare la distorsione quando $E(u|X) \neq 0$ – usando una *variabile strumentale* (IV), Z .

Lo stimatore IV con un singolo regressore e un singolo strumento (Paragrafo 12.1)

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

- La regressione IV divide X in due parti: una che potrebbe essere correlata con u , e una che non lo è. Isolando la parte che non è correlata con u , è possibile stimare β_1 .
- Per fare questo si utilizza una **variabile strumentale**, Z_i , che è correlata con X_i ma incorrelata con u_i .

Terminologia: endogeneità ed esogeneità

Una variabile **endogena** è una variabile correlata con u

Una variabile **esogena** è una variabile incorrelata con u

Nella regressione IV ci concentriamo sul caso in cui X è endogena ed esiste uno strumento, Z , esogeno.

Digressione sulla terminologia: “endogeno” significa letteralmente “determinato all’interno del sistema”. Se X è congiuntamente determinata con Y , allora una regressione di Y su X è soggetta a distorsione da causalità simultanea. Ma questa definizione di endogeneità è troppo stretta perché sia possibile usare la regressione IV per risolvere i problemi di distorsione da variabili omesse e da errori nelle variabili, quindi usiamo la definizione più ampia fornita sopra.

Due condizioni per avere uno strumento valido

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

Perché una variabile strumentale (uno “**strumento**”) Z sia valida, deve soddisfare due condizioni:

- 1. Rilevanza:** $\text{corr}(Z_i, X_i) \neq 0$
- 2. Esogeneità:** $\text{corr}(Z_i, u_i) = 0$

Supponiamo per ora di avere un tale Z_i (vedremo più avanti come trovare variabili strumentali); come possiamo usarlo per stimare β_1 ?

Lo stimatore IV con una X e una Z

Spiegazione 1: minimi quadrati in due stadi (TSLS)

Ci sono due stadi – due regressioni:

- (1) Si isola la parte di X che non è correlata con u mediante la regressione di X su Z usando gli OLS:

$$X_i = \pi_0 + \pi_1 Z_i + v_i \quad (1)$$

- Poiché Z_i non è correlato con u_i , $\pi_0 + \pi_1 Z_i$ non è correlato con u_i . Non conosciamo π_0 o π_1 ma li abbiamo stimati, perciò...
- Si calcolano i valori predetti di X_i , \hat{X}_i , dove $\hat{X}_i = + \hat{\pi}_1 Z_i$, $i = 1, \dots, n$.

Minimi quadrati in due stadi (continua)

(2) Si sostituisce X_i con \hat{X}_i nella regressione di interesse:
si esegue la regressione di Y su \hat{X}_i usando gli OLS:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i \quad (2)$$

- Poiché \hat{X}_i è incorrelato con u_i , la prima assunzione dei minimi quadrati vale per la regressione (2). (Ciò richiede che n sia grande in modo che π_0 e π_1 siano stimati con precisione)
- Quindi, in grandi campioni, β_1 può essere stimato con gli OLS usando la regressione (2)
- Lo stimatore risultante è detto *stimatore dei minimi quadrati in due stadi (TSLS)*, $\hat{\beta}_1^{TSLS}$.

Minimi quadrati in due stadi: riepilogo

Supponiamo che Z_i , soddisfi le due condizioni per uno strumento valido:

- 1. Rilevanza:** $\text{corr}(Z_i, X_i) \neq 0$
- 2. Esogeneità:** $\text{corr}(Z_i, u_i) = 0$

Minimi quadrati in due stadi:

Stadio 1: Regressione di X_i su Z_i (inclusa intercetta), ottenendo i valori predetti \hat{X}_i

Stadio 2: Regressione di Y_i su \hat{X}_i (inclusa intercetta); il coefficiente di \hat{X}_i è lo stimatore TSLS, $\hat{\beta}_1^{\text{TSLS}}$.

$\hat{\beta}_1^{\text{TSLS}}$ è uno stimatore consistente di β_1 .

Lo stimatore IV, una X e una Z (continua)

Spiegazione 2: derivazione algebrica diretta

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

Allora

$$\begin{aligned}\text{cov}(Y_i, Z_i) &= \text{cov}(\beta_0 + \beta_1 X_i + u_i, Z_i) \\ &= \text{cov}(\beta_0, Z_i) + \text{cov}(\beta_1 X_i, Z_i) + \text{cov}(u_i, Z_i) \\ &= 0 + \text{cov}(\beta_1 X_i, Z_i) + 0 \\ &= \beta_1 \text{cov}(X_i, Z_i)\end{aligned}$$

dove $\text{cov}(u_i, Z_i) = 0$ per l'esogeneità dello strumento; quindi

$$\beta_1 = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(X_i, Z_i)}$$

Lo stimatore IV, una X e una Z (continua)

$$\beta_1 = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(X_i, Z_i)}$$

Lo stimatore IV sostituisce queste covarianze della popolazione con covarianze campionarie:

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}},$$

s_{YZ} e s_{XZ} sono covarianze campionarie. Questo è lo stimatore TSLS – con una derivazione diversa!

Lo stimatore IV, una X e una Z (continua)

Spiegazione 3: derivazione dalla “forma ridotta”

La “forma ridotta” mette in relazione Y a Z e X a Z :

$$\begin{cases} X_i = \pi_0 + \pi_1 Z_i + v_i \\ Y_i = \gamma_0 + \gamma_1 Z_i + w_i \end{cases}$$

dove w_i è un termine d'errore. Poiché Z è esogena, è incorrelata con v_i e con w_i .

L'idea: una variazione unitaria in Z_i comporta una variazione in X_i di π_1 e una variazione in Y_i di γ_1 . Poiché tale variazione in X_i nasce dalla variazione esogena in Z_i , tale variazione in X_i è esogena.

Quindi una variazione esogena in X_i di π_1 unità è associata a una variazione in Y_i di γ_1 unità – perciò l'effetto su Y di una variazione esogena in X è $\beta_1 = \gamma_1 / \pi_1$ unità.

Lo stimatore IV dalla forma ridotta (continua)

I calcoli:

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

$$Y_i = \gamma_0 + \gamma_1 Z_i + w_i$$

Risolviamo l'equazione di X in Z :

$$Z_i = -\pi_0/\pi_1 + (1/\pi_1)X_i - (1/\pi_1)v_i$$

Sostituiamo nell'equazione di Y e raccogliamo i termini:

$$\begin{aligned} Y_i &= \gamma_0 + \gamma_1 Z_i + w_i \\ &= \gamma_0 + \gamma_1[-\pi_0/\pi_1 + (1/\pi_1)X_i - (1/\pi_1)v_i] + w_i \\ &= [\gamma_0 - \pi_0\gamma_1/\pi_1] + (\gamma_1/\pi_1)X_i + [w_i - (\gamma_1/\pi_1)v_i] \\ &= \beta_0 + \beta_1 X_i + u_i, \end{aligned}$$

dove $\beta_0 = \gamma_0 - \pi_0\gamma_1/\pi_1$, $\beta_1 = \gamma_1/\pi_1$, e $u_i = w_i - (\gamma_1/\pi_1)v_i$.

Lo stimatore IV dalla forma ridotta (continua)

$$X_i = \pi_0 + \pi_1 Z_i + v_i$$

$$Y_i = \gamma_0 + \gamma_1 Z_i + w_i$$

quindi

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

dove

$$\beta_1 = \gamma_1 / \pi_1$$

Interpretazione: una variazione esogena in X_i di π_1 unità è associata a una variazione in Y_i di γ_1 unità – perciò l'effetto su Y di una variazione unitaria esogena in X è $\beta_1 = \gamma_1 / \pi_1$.

Consistenza dello stimatore TSLS

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}}$$

Le covarianze campionarie sono consistenti:

$s_{YZ} \xrightarrow{p} \text{cov}(Y, Z)$ e $s_{XZ} \xrightarrow{p} \text{cov}(X, Z)$. Quindi

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}} \xrightarrow{p} \frac{\text{cov}(Y, Z)}{\text{cov}(X, Z)} = \beta_1$$

- La condizione di rilevanza dello strumento, $\text{cov}(X, Z) \neq 0$, assicura che non si esegua una divisione per zero.

Inferenza con TSLS

- In grandi campioni, la distribuzione campionaria dello stimatore TSLS è normale
- L'inferenza (verifiche di ipotesi, intervalli di confidenza) procede nel modo consueto, ovvero $\pm 1,96SE$
- Il concetto alla base della distribuzione normale in grandi campioni dello stimatore TSLS è che – come tutti gli altri estimatori che abbiamo considerato – comporta variabili casuali i.i.d. con media nulla, a cui possiamo applicare il TLC.
- Di seguito riportiamo i calcoli abbozzati (si veda l'Appendice 12.3 per i dettagli)...

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}} = \frac{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Z_i - \bar{Z})}$$

$$= \frac{\sum_{i=1}^n Y_i (Z_i - \bar{Z})}{\sum_{i=1}^n X_i (Z_i - \bar{Z})}$$

Sostituiamo in $Y_i = \beta_0 + \beta_1 X_i + u_i$ e semplifichiamo:

$$\hat{\beta}_1^{TSLS} = \frac{\beta_1 \sum_{i=1}^n X_i (Z_i - \bar{Z}) + \sum_{i=1}^n u_i (Z_i - \bar{Z})}{\sum_{i=1}^n X_i (Z_i - \bar{Z})}$$

quindi...

$$\hat{\beta}_1^{TSLS} = \beta_1 + \frac{\sum_{i=1}^n u_i (Z_i - \bar{Z})}{\sum_{i=1}^n X_i (Z_i - \bar{Z})}$$

Quindi $\hat{\beta}_1^{TSLS} - \beta_1 =$

$$\frac{\sum_{i=1}^n u_i (Z_i - \bar{Z})}{\sum_{i=1}^n X_i (Z_i - \bar{Z})}$$

Moltiplicando entrambi i membri per \sqrt{n}

$$\sqrt{n}(\hat{\beta}_1^{TSLS} - \beta_1) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \bar{Z}) u_i}{\frac{1}{n} \sum_{i=1}^n X_i (Z_i - \bar{Z})}$$


$$\sqrt{n} (\hat{\beta}_1^{TSLS} - \beta_1) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \bar{Z}) u_i}{\frac{1}{n} \sum_{i=1}^n X_i (Z_i - \bar{Z})}$$

$$\frac{1}{n} \sum_{i=1}^n X_i (Z_i - \bar{Z}) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Z_i - \bar{Z}) \xrightarrow{p} \text{cov}(X, Z) \neq 0$$

$\frac{1}{\sqrt{n}} \sum_{i=1}^n (Z_i - \bar{Z}) u_i$ ha distribuzione $N(0, \text{var}[(Z - \mu_Z) u])$ (TLC)

quindi: $\hat{\beta}_1^{TSLS}$ ha distribuzione appr. $N(\beta_1, \sigma_{\hat{\beta}_1^{TSLS}}^2)$

dove $\sigma_{\hat{\beta}_1^{TSLS}}^2 = \frac{1}{n} \frac{\text{var}[(Z_i - \mu_Z) u_i]}{[\text{cov}(Z_i, X_i)]^2}$

dove $\text{cov}(X, Z) \neq 0$ perché lo strumento è rilevante

Inferenza con TSLS (continua)

$\hat{\beta}_1^{TSLS}$ ha distribuzione appr. $N(\beta_1, \sigma_{\hat{\beta}_1^{TSLS}}^2)$

- L'inferenza statistica procede nel modo consueto.
- La giustificazione è (come di consueto) basata su grandi campioni
- Tutto questo assume che gli strumenti siano validi – vedremo tra breve che cosa accade se non lo sono.
- ***Nota importante sugli errori standard:***
 - Gli errori standard OLS dalla regressione del secondo stadio non sono corretti – non tengono conto della stima al primo stadio (\hat{X} è stimata).
 - Si utilizza invece un singolo comando apposito che calcola lo stimatore TSLS e gli errori standard corretti.
 - Come di consueto, si usano errori standard robusti all'eteroschedasticità

Quali sono le conseguenze di strumenti deboli?

Se gli strumenti sono deboli, la distribuzione campionaria del TSLS e della sua statistica t non è normale, anche con n grande.

Consideriamo il caso più semplice:

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

$$X_i = \pi_0 + \pi_1 Z_i + u_i$$

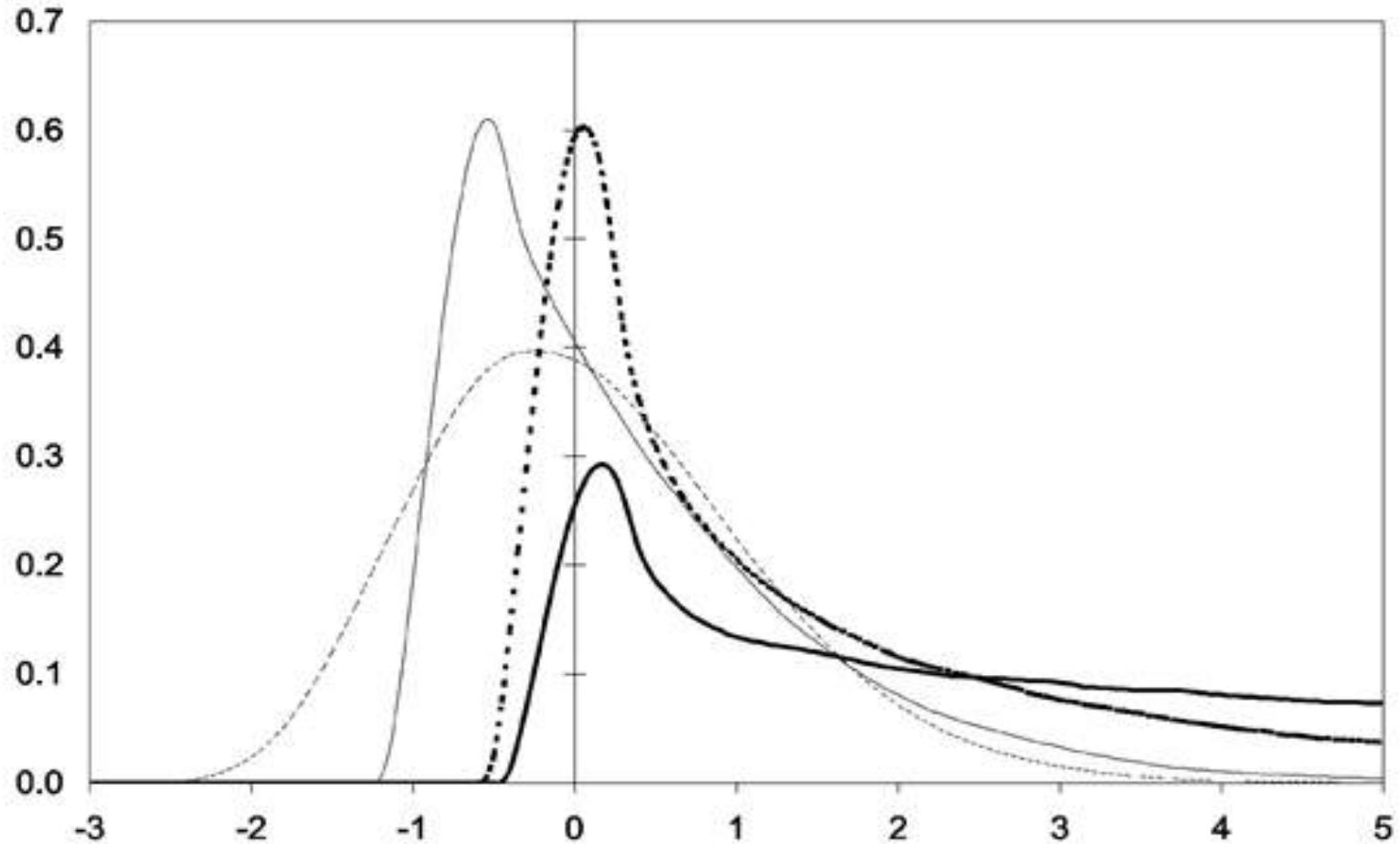
- Lo stimatore IV è

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}}$$

- Se $\text{cov}(X, Z)$ è zero o minore, allora s_{XZ} sarà piccolo: con strumenti deboli, il denominatore è quasi zero.

- In questo caso, la distribuzione campionaria di $\hat{\beta}_1^{TSLS}$ (e la sua statistica t) non è ben approssimata dall'approssimazione normale per n grande...

Esempio: la distribuzione campionaria della statistica t del TSLS con strumenti deboli



Linea scura = strumenti non rilevanti

Linea chiara tratteggiata = strumenti forti

Perché la nostra approssimazione normale ci tradisce?

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}}$$

- Se $\text{cov}(X, Z)$ è piccola, piccole variazioni in s_{XZ} (da un campione al successivo) può indurre grandi variazioni in $\hat{\beta}_1^{TSLS}$
- Supponiamo di calcolare in un campione $s_{XZ} = 0,00001\dots$
- Allora l'approssimazione normale per n grande non è una buona approssimazione della distribuzione campionaria di $\hat{\beta}_1^{TSLS}$
- Un'approssimazione migliore è quella di $\hat{\beta}_1^{TSLS}$ come il rapporto di due variabili casuali normali correlate (si veda l'Appendice 12.4)
CUPOLE
- Se gli strumenti sono deboli, i consueti metodi di inferenza sono inaffidabili – potenzialmente molto inaffidabili.

Misurazione della forza degli strumenti in pratica: la statistica F del primo stadio

- La regressione del primo stadio (una sola X):
- Regressione di X su $Z_1, \dots, Z_m, W_1, \dots, W_k$.
- Strumenti totalmente irrilevanti \leftrightarrow tutti i coefficienti di Z_1, \dots, Z_m sono zero.
- La **statistica F del primo stadio** verifica l'ipotesi che Z_1, \dots, Z_m non entrino nella regressione del primo stadio.
- Strumenti deboli implicano un valore basso della statistica F del primo stadio.

Verifica di strumenti deboli con una singola X

- Si calcola la statistica F del primo stadio.

Regola empirica: se la statistica F del primo stadio è minore di 10, allora l'insieme di strumenti è debole.

- In questo caso, lo stimatore TSLS sarà distorto, e le inferenze statistiche (errori standard, verifiche di ipotesi, intervalli di confidenza) possono essere fuorvianti.

Verifica di strumenti deboli con una singola X (continua)

- Perché confrontare la statistica F del primo stadio con 10?
- Non è sufficiente respingere l'ipotesi nulla che i coefficienti delle Z siano zero – serve un contenuto predittivo sostanziale per una buona approssimazione normale.
- Il confronto della statistica F del primo stadio con 10 verifica se la distorsione del TSLS, rispetto all'OLS, è minore del 10%. Se la F è minore di 10, la distorsione relativa è superiore al 10%, cioè il TSLS può avere una distorsione sostanziale (si veda l'Appendice 12.5).

Che cosa fare se si hanno strumenti deboli

- Procurarsi strumenti migliori (più facile a dirsi che a farsi!)
- Se si hanno molti strumenti, alcuni sono probabilmente più deboli di altri ed è una buona idea scartare i più deboli (scartando uno strumento irrilevante si aumenta la statistica F del primo stadio)
- Se si hanno pochi strumenti, e sono tutti deboli, allora occorre eseguire un'analisi IV al di là del TSLS...
 - Separare il problema della stima di β_1 e della costruzione di intervalli di confidenza
 - Sembra strano, ma se il TSLS non ha distribuzione normale, ha senso (davvero?)

Intervalli di confidenza con strumenti deboli

- Con strumenti deboli, gli intervalli di confidenza TSLS non sono validi, ma altri intervalli di confidenza lo sono. Riportiamo due modi per calcolare intervalli di confidenza validi in grandi campioni anche se gli strumenti sono deboli:

1. L'intervallo di confidenza di Anderson-Rubin

- L'intervallo di confidenza di Anderson-Rubin si basa sulla statistica **test di Anderson-Rubin** che verifica $\beta_1 = \beta_{1,0}$:
 - Si calcola $= Y_i - \beta_{1,0}X_i$
 - Si esegue la regressione su $W_{1i}, \dots, W_{ri}, Z_{1i}, \dots, Z_{mi}$
 - Il test AR è la statistica F su Z_{1i}, \dots, Z_{mi}
- Ora si inverte il test: l'intervallo di confidenza AR al 95% è l'insieme di β_1 non rifiutati al livello del 5% dal test AR.
- Calcolo: si usa software specialistico.

Intervalli di confidenza con strumenti deboli (continua)

2. Intervallo di confidenza del rapporto di verosimiglianza condizionato di Moreira

- L'intervallo di confidenza del rapporto di verosimiglianza condizionato è basato sull'inversione del test del rapporto di verosimiglianza condizionato di Moreira. Per calcolare questo test, il suo valore critico e l'intervallo di confidenza del rapporto di verosimiglianza condizionato, è richiesto un software specialistico.
- Questo intervallo di confidenza tende a essere più ristretto di quello di Anderson-Rubin, soprattutto quando vi sono molti strumenti.
- Se si dispone di un software che produce questo intervallo, è il caso di usarlo.

Stima con strumenti deboli

Non ci sono stimatori non distorti se gli strumenti sono deboli o irrilevanti. Tuttavia, alcuni stimatori hanno una distribuzione più centrata su β_1 del TSLS.

- Uno di questi stimatori è quello di massima verosimiglianza con informazione limitata (LIML)
- Lo stimatore LIML
 - può essere derivato come stimatore di massima verosimiglianza
 - è il valore di β_1 che minimizza il valore- p del test AR (!)
- Per approfondire stimatori, verifiche e intervalli di confidenza nel caso di strumenti deboli, si veda l'Appendice 12.5

Verifica dell'assunzione 2: esogeneità dello strumento

- Esogeneità dello strumento: **Tutti** gli strumenti sono correlati con il termine d'errore:
 $\text{corr}(Z_{1i}, u_i) = 0, \dots, \text{corr}(Z_{mi}, u_i) = 0$
- Se gli strumenti sono correlati con il termine d'errore, il primo stadio del TSLS non può isolare una componente di X incorrelata con il termine d'errore, perciò \hat{X} è correlata con u e il TSLS è inconsistente.
- Se ci sono più strumenti che regressori endogeni, è possibile verificare – *parzialmente* – l'esogeneità dello strumento.

Verifica di restrizioni di sovraidentificazione

Consideriamo il caso più semplice:

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

- Supponiamo che vi siano due strumenti validi: Z_{1i}, Z_{2i}
- Allora potremmo calcolare due stime TSLS separate.
- Intuitivamente, se queste due stime TSLS sono molto diverse tra loro, ci dev'essere qualcosa di sbagliato: uno strumento o l'altro (o entrambi) devono essere non validi.
- Il test J di restrizioni sovraidentificanti esegue questo confronto in un modo statisticamente preciso.
- Si può fare soltanto se il numero di Z è maggiore del numero di X (sovraidentificazione).

Il test J di restrizioni di sovraidentificazione

Supponiamo che il numero di strumenti = $m >$ numero di $X = k$
(sovraidentificazione)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \dots + \beta_{k+r} W_{ri} + u_i$$

Il test J è il test di Anderson-Rubin, usando lo stimatore TSLS al posto del valore ipotizzato $\beta_{1,0}$. Procedura:

1. Prima si stima l'equazione di interesse usando TSLS e tutti gli m strumenti; si calcolano i valori predetti \hat{Y}_i , usando le X effettive (non le usate per stimare il secondo stadio)
2. Si calcolano i residui $\hat{u}_i = Y_i - \hat{Y}_i$
3. Si esegue la regressione rispetto a $Z_{1i}, \dots, Z_{mi}, W_{1i}, \dots, W_{ri}$
4. Si calcola la statistica F che verifica l'ipotesi che i coefficienti di Z_{1i}, \dots, Z_{mi} siano tutti zero;
5. La **statistica J** è $J = mF$

Il test J (continua)

$J = mF$, dove F = la statistica F che verifica i coefficienti di Z_{1i}, \dots, Z_{mi} in una regressione dei residui TSLS rispetto a $Z_{1i}, \dots, Z_{mi}, W_{1i}, \dots, W_{ri}$.

Distribuzione della statistica J

- Sotto l'ipotesi nulla che tutti gli strumenti siano esogeni, J ha una distribuzione chi-quadro con $m-k$ gradi di libertà
- Se $m = k$, $J = 0$ (*ha senso?*)
- Se alcuni strumenti sono esogeni e altri sono endogeni, la statistica J sarà grande, e l'ipotesi nulla che tutti gli strumenti sono esogeni sarà respinta.

Verifica della validità degli strumenti: riepilogo

Questo riepilogo considera il caso di una singola X . I due requisiti per la validità degli strumenti sono:

1. *Rilevanza*

- Almeno uno strumento deve entrare nella controparte della regressione del primo stadio.
- Se gli strumenti sono deboli, allora lo stimatore TSLS è distorto e la statistica t ha una distribuzione non normale
- Per verificare strumenti deboli con un singolo regressore endogeno incluso, si verifica la statistica F del primo stadio
 - Se $F > 10$, gli strumenti sono forti – si usa il TSLS
 - Se $F < 10$, gli strumenti sono deboli – si fa qualcosa.

2. Esogeneità

- **Tutti** gli strumenti devono essere incorrelati con il termine d'errore: $\text{corr}(Z_{1i}, u_i) = 0, \dots, \text{corr}(Z_{mi}, u_i) = 0$
- Possiamo eseguire una verifica parziale di esogeneità: se $m > 1$, possiamo verificare l'ipotesi nulla che tutti gli strumenti siano esogeni contro l'alternativa che almeno $m-1$ siano endogeni (correlati con u)
- Si usa il test J , realizzato usando i residui TSLS.
- Se il J respinge l'ipotesi, allora almeno alcuni degli strumenti sono endogeni, perciò occorre prendere una decisione difficile e scartare alcuni (o tutti) gli strumenti.

Chapter 6

Univariate time series modelling and forecasting

Univariate Time Series Models

- Where we attempt to predict returns using only information contained in their past values.

Some Notation and Concepts

- A Strictly Stationary Process

A strictly stationary process is one where

$$P\{y_{t_1} \leq b_1, \dots, y_{t_n} \leq b_n\} = P\{y_{t_1+m} \leq b_1, \dots, y_{t_n+m} \leq b_n\}$$

- A Weakly Stationary Process

↳ Debolmente

Univariate Time Series Models (Cont'd)

If a series satisfies the next three equations, it is said to be weakly or covariance stationary

$$(1) \ E(y_t) = \mu \quad t = 1, 2, \dots, \infty$$

$$(2) \ E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty$$

$$(3) \ E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1} \quad \forall t_1, t_2$$

- So if the process is covariance stationary, all the variances are the same and all the covariances depend on the difference between t_1 and t_2 . The moments

$$E((y_t - E(y_t))(y_{t-s} - E(y_{t-s}))) = \gamma_s, s = 0, 1, 2, \dots$$

are known as the covariance function.

- The covariances, γ_s , are known as autocovariances.

Univariate Time Series Models (Cont'd)

- However, the value of the autocovariances depend on the units of measurement of y_t .
- It is thus more convenient to use the autocorrelations which are the autocovariances normalised by dividing by the variance:

$$\rho_s = \frac{\gamma_s}{\gamma_0} \quad \tau_s = \frac{\gamma_s}{\gamma_0}, \quad s = 0, 1, 2, \dots$$

- If we plot τ_s against $s=0,1,2,\dots$ then we obtain the autocorrelation function or correlogram.

A White Noise Process

- A white noise process is one with (virtually) no discernible structure. A definition of a white noise process is

$$E(y_t) = \mu$$

$$\text{var}(y_t) = \sigma^2$$

$$\gamma_{t-r} = \begin{cases} \sigma^2 & \text{if } t = r \\ 0 & \text{otherwise} \end{cases}$$

- Thus the autocorrelation function will be zero apart from a single peak of 1 at $s=0$. $\hat{\tau}_s \sim \text{approx. } N(0, 1/T)$ where T = sample size
- We can use this to do significance tests for the autocorrelation coefficients by constructing a confidence interval.

Joint Hypothesis Tests

- We can also test the joint hypothesis that all m of the τ_k correlation coefficients are simultaneously equal to zero using the Q -statistic developed by Box and Pierce:

$$Q = T \sum_{k=1}^m \hat{\tau}_k^2$$

where T =sample size, m =maximum lag length

- The Q -statistic is asymptotically distributed as a χ_m^2 .
- However, the Box Pierce test has poor small sample properties, so a variant has been developed, called the Ljung-Box statistic:

$$Q^* = T(T+2) \sum_{k=1}^m \frac{\hat{\tau}_k^2}{T-k} \sim \chi_m^2$$

- This statistic is very useful as a portmanteau (general) test of linear dependence in time series.

Moving Average Processes

$$v_t = \epsilon_t$$

- Let u_t ($t = 1, 2, 3, \dots$) be a sequence of independently and identically distributed (iid) random variables with $E(u_t) = 0$ and $\text{var}(u_t) = \sigma^2$, then

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$$

is a q th order moving average model MA(q).

- Its properties are

$$E(y_t) = \mu$$

$$\text{var}(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2) \sigma^2$$

Covariances

$$\gamma_s = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s}) \sigma^2 & \text{for } s = 1, \dots, q \\ 0 & \text{for } s > q \end{cases}$$

Example of an MA Problem

Esercizio, utile farlo

1. Consider the following MA(2) process:

$$y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$$

where u_t is a zero mean white noise process with variance σ^2 .

- ✓ i. Calculate the mean and variance of X_t
- / ii. Derive the autocorrelation function for this process (i.e. express the autocorrelations, τ_1 , τ_2 , ...as functions of the parameters θ_1 and θ_2).
- ✓ iii. If $\theta_1 = -0.5$ and $\theta_2 = 0.25$, sketch the acf of X_t .

Autoregressive Processes

- An autoregressive model of order p , an AR(p) can be expressed as

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t$$

- Or using the lag operator notation:

$$Ly_t = y_{t-1} \quad L^i y_t = y_{t-i}$$

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + u_t$$

- or

$$y_t = \mu + \sum_{i=1}^p \phi_i L^i y_t + u_t$$

- or

$$\phi(L)y_t = \mu + u_t \quad \text{where} \quad \phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p).$$

The Stationary Condition for an AR Model

- The condition for stationarity of a general AR(p) model is that the roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ all lie outside the unit circle.



- A stationary AR(p) model is required for it to have an MA(∞) representation.

- Example 1: Is $y_t = y_{t-1} + u_t$ stationary?

The characteristic root is 1, so it is a unit root process (so non-stationary)

- Example 2: Is $y_t = 3y_{t-1} - 2.75y_{t-2} + 0.75y_{t-3} + u_t$ stationary?

The characteristic roots are 1, 2/3, and 2. Since only one of these lies outside the unit circle, the process is non-stationary.

Wold's Decomposition Theorem

- States that any stationary series can be decomposed into the sum of two unrelated processes, a purely deterministic part and a purely stochastic part, which will be an $\text{MA}(\infty)$.
- For the $\text{AR}(p)$ model, $\phi(L)y_t = u_t$, ignoring the intercept, the Wold decomposition is

$$y_t = \psi(L)u_t$$

where,

$$\psi(L) = \phi(L)^{-1} = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)^{-1}$$

The Moments of an Autoregressive Process

- The moments of an autoregressive process are as follows. The mean is given by

$$E(y_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

$\mathbb{P}(L)Y_t = \mu + \epsilon_t$
 $E[\phi(L)Y_t] = E[\mu] + E[\epsilon_t]$
 $E[\epsilon_t] = \frac{\nu}{\phi(L)}$

- The autocovariances and autocorrelation functions can be obtained by solving what are known as the Yule-Walker equations:

$$\tau_1 = \phi_1 + \tau_1 \phi_2 + \cdots + \tau_{p-1} \phi_p$$

$$\tau_2 = \tau_1 \phi_1 + \phi_2 + \cdots + \tau_{p-2} \phi_p$$

⋮ ⋮ ⋮

$$\tau_p = \tau_{p-1} \phi_1 + \tau_{p-2} \phi_2 + \cdots + \phi_p$$

- If the AR model is stationary, the autocorrelation function will decay exponentially to zero.

Sample AR Problem

- Consider the following simple AR(1) model

$$y_t = \mu + \phi_1 y_{t-1} + u_t$$

- i. Calculate the (unconditional) mean of y_t .
 - For the remainder of the question, set $\mu = 0$ for simplicity.
 - ii. Calculate the (unconditional) variance of y_t .
 - iii. Derive the autocorrelation function for y_t .

Solution

i. Unconditional mean:

$$\begin{aligned} E(y_t) &= E(\mu + \phi_1 y_{t-1}) \\ E(y_t) &= \mu + \phi_1 E(y_{t-1}) \end{aligned}$$

But also

So

$$\begin{aligned} E(y_t) &= \mu + \phi_1(\mu + \phi_1 E(y_{t-2})) \\ &= \mu + \phi_1\mu + \phi_1^2 E(y_{t-2}) \\ &= \mu + \phi_1\mu + \phi_1^2(\mu + \phi_1 E(y_{t-3})) \\ E(y_t) &= \mu + \phi_1\mu + \phi_1^2\mu + \phi_1^3 E(y_{t-3}) \end{aligned}$$

Solution (Cont'd)

An infinite number of such substitutions would give

$$E(y_t) = \mu(1 + \phi_1 + \phi_1^2 + \dots) + \phi_1^\infty y_0$$

So long as the model is stationary, i.e. $|\phi_1| < 1$, then $\phi_1^\infty = 0$.

So

$$E(y_t) = \mu(1 + \phi_1 + \phi_1^2 + \dots) = \frac{\mu}{1 - \phi_1}$$

- ii. Calculating the variance of y_t : $y_t = \phi_1 y_{t-1} + u_t$

Solution (Cont'd)

From Wold's decomposition theorem:

$$y_t(1 - \phi_1 L) = u_t$$

$$y_t = (1 - \phi_1 L)^{-1} u_t$$

$$y_t = (1 + \phi_1 L + \phi_1^2 L^2 + \dots) u_t$$

So long as, $|\phi_1| < 1$, this will converge.

$$\text{var}(y_t) = E[y_t - E(y_t)][y_t - E(y_t)]$$

Solution (Cont'd)

but $E(y_t) = 0$, since μ is set to zero.

$$\begin{aligned}\text{var}(y_t) &= E[(y_t)(y_t)] \\ &= E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \dots)(u_t + \phi_1 u_{t-1} \\ &\quad + \phi_1^2 u_{t-2} + \dots)] \\ &= E[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \dots + \text{cross-products}] \\ &= E[u_t^2 + \phi_1^2 u_{t-1}^2 + \phi_1^4 u_{t-2}^2 + \dots] \\ &= \sigma_u^2 + \phi_1^2 \sigma_u^2 + \phi_1^4 \sigma_u^2 + \dots \\ &= \sigma_u^2 (1 + \phi_1^2 + \phi_1^4 + \dots) \\ &= \frac{\sigma_u^2}{(1 - \phi_1^2)}\end{aligned}$$

Solution (Cont'd)

- iii. Turning now to calculating the acf, first calculate the autocovariances:

$$\gamma_1 = \text{cov}(y_t, y_{t-1}) = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$$

Since a_0 has been set to zero, $E(y_t) = 0$ and $E(y_{t-1}) = 0$, so

$$\gamma_1 = E[y_t y_{t-1}]$$

Solution (Cont'd)

under the result above that $E(y_t) = E(y_{t-1}) = 0$. Thus

$$\begin{aligned}\gamma_1 &= E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \dots)(u_{t-1} + \phi_1 u_{t-2} \\&\quad + \phi_1^2 u_{t-3} + \dots)] \\ \gamma_1 &= E[\phi_1 u_{t-1}^2 + \phi_1^3 u_{t-2}^2 + \dots + \text{cross-products}] \\ \gamma_1 &= \phi_1 \sigma^2 + \phi_1^3 \sigma^2 + \phi_1^5 \sigma^2 + \dots \\ \gamma_1 &= \frac{\phi_1 \sigma^2}{(1 - \phi_1^2)}\end{aligned}$$

For the second autocorrelation coefficient,

$$\gamma_2 = \text{cov}(y_t, y_{t-2}) = E[y_t - E(y_t)][y_{t-2} - E(y_{t-2})]$$

Solution (Cont'd)

Using the same rules as applied above for the lag 1 covariance

$$\gamma_2 = E[y_t y_{t-2}]$$

$$\begin{aligned}\gamma_2 &= E[(u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \dots)(u_{t-2} + \phi_1 u_{t-3} \\ &\quad + \phi_1^2 u_{t-4} + \dots)]\end{aligned}$$

$$\gamma_2 = E[\phi_1^2 u_{t-2}^2 + \phi_1^4 u_{t-3}^2 + \dots + \text{cross-products}]$$

$$\gamma_2 = \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + \dots$$

$$\gamma_2 = \phi_1^2 \sigma^2 (1 + \phi_1^2 + \phi_1^4 + \dots)$$

$$\gamma_2 = \frac{\phi_1^2 \sigma^2}{(1 - \phi_1^2)}$$

Solution (Cont'd)

If these steps were repeated for γ_3 , the following expression would be obtained

$$\gamma_3 = \frac{\phi_1^3 \sigma^2}{(1 - \phi_1^2)}$$

and for any lag s , the autocovariance would be given by

$$\gamma_s = \frac{\phi_1^s \sigma^2}{(1 - \phi_1^2)}$$

Solution (Cont'd)

The acf can now be obtained by dividing the covariances by the variance:

$$\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{\left(\frac{\phi_1 \sigma^2}{(1 - \phi_1^2)} \right)}{\left(\frac{\sigma^2}{(1 - \phi_1^2)} \right)} = \phi_1$$

$$\tau_2 = \frac{\gamma_2}{\gamma_0} = \frac{\left(\frac{\phi_1^2 \sigma^2}{(1 - \phi_1^2)} \right)}{\left(\frac{\sigma^2}{(1 - \phi_1^2)} \right)} = \phi_1^2$$

$$\tau_3 = \phi_1^3$$

$$\tau_s = \phi_1^s$$

The Partial Autocorrelation Function (denoted τ_{kk})

- Measures the correlation between an observation k periods ago and the current observation, after controlling for observations at intermediate lags (i.e. all lags $< k$).
- So τ_{kk} measures the correlation between y_t and y_{t-k} after removing the effects of $y_{t-k+1}, y_{t-k+2}, \dots, y_{t-1}$
- At lag 1, the acf = pacf always
- At lag 2,

$$\tau_{22} = (\tau_2 - \tau_1^2) / (1 - \tau_1^2)$$

- For lags 3+, the formulae are more complex.

The Partial Autocorrelation Function (denoted τ_{kk}) (Cont'd)

- The pacf is useful for telling the difference between an AR process and an ARMA process.
- In the case of an $AR(p)$, there are direct connections between y_t and y_{t-s} only for $s \leq p$.
- So for an $AR(p)$, the theoretical pacf will be zero after lag p .
- In the case of an $MA(q)$, this can be written as an $AR(\infty)$, so there are direct connections between y_t and all its previous values.
- For an $MA(q)$, the theoretical pacf will be geometrically declining.

ARMA Processes

- By combining the AR(p) and MA(q) models, we can obtain an ARMA(p,q) model:

$$\phi(L)y_t = \mu + \theta(L)u_t$$

where

$$\phi(L) = 1 - \phi_1L - \phi_2L^2 - \cdots - \phi_pL^p \quad \text{and}$$

$$\theta(L) = 1 + \theta_1L + \theta_2L^2 + \cdots + \theta_qL^q$$

or

$$y_t = \mu + \phi_1y_{t-1} + \phi_2y_{t-2} + \cdots + \phi_py_{t-p} + \theta_1u_{t-1}$$

$$+ \theta_2u_{t-2} + \cdots + \theta_qu_{t-q} + u_t$$

with

$$E(u_t) = 0; E(u_t^2) = \sigma^2; E(u_t u_s) = 0, t \neq s$$

The Invertibility Condition

- Similar to the stationarity condition, we typically require the MA(q) part of the model to have roots of $\theta(z) = 0$ greater than one in absolute value.
$$\begin{cases} \theta(z)=0 \\ |z| > 1 \end{cases}$$
- The mean of an ARMA series is given by

$$E(y_t) = \frac{\mu}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

- The autocorrelation function for an ARMA process will display combinations of behaviour derived from the AR and MA parts, but for lags beyond q , the acf will simply be identical to the individual AR(p) model.

Summary of the Behaviour of the acf for AR and MA Processes

An autoregressive process has

- a geometrically decaying acf
- number of spikes of pacf = AR order

A moving average process has

- Number of spikes of acf = MA order
- a geometrically decaying pacf

Building ARMA Models - The Box Jenkins Approach

- Box and Jenkins (1970) were the first to approach the task of estimating an ARMA model in a systematic manner. There are 3 steps to their approach:
 1. Identification
 2. Estimation
 3. Model diagnostic checking

Step 1:

- Involves determining the order of the model.
- Use of graphical procedures
- A better procedure is now available

Step 2:

Building ARMA Models - The Box Jenkins Approach (Cont'd)

- Estimation of the parameters
- Can be done using least squares or maximum likelihood depending on the model.

Step 3:

- Model checking

Box and Jenkins suggest 2 methods:

- deliberate overfitting
- residual diagnostics

Some More Recent Developments in ARMA Modelling

- Identification would typically not be done using acf's.
- We want to form a parsimonious model.
- Reasons:
 - variance of estimators is inversely proportional to the number of degrees of freedom.
 - models which are profligate might be inclined to fit to data specific features
- This gives motivation for using information criteria, which embody 2 factors
 - a term which is a function of the RSS
 - some penalty for adding extra parameters
- The object is to choose the number of parameters which minimises the information criterion.

Information Criteria for Model Selection

- The information criteria vary according to how stiff the penalty term is.
- The three most popular criteria are Akaike's (1974) information criterion (AIC), Schwarz's (1978) Bayesian information criterion (SBIC), and the Hannan-Quinn criterion (HQIC).

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T}$$

$$SBIC = \ln(\hat{\sigma}^2) + \frac{k}{T} \ln T$$

$$HQIC = \ln(\hat{\sigma}^2) + \frac{2k}{T} \ln(\ln(T))$$

Information Criteria for Model Selection (Cont'd)

where $k = p + q + 1$, T = sample size. So we min. IC s.t.
 $p \leq \bar{p}$, $q \leq \bar{q}$

$SBIC$ embodies a stiffer penalty term than AIC .

- Which IC should be preferred if they suggest different model orders?
 - $SBIC$ is strongly consistent but (inefficient).
 - AIC is not consistent, and will typically pick “bigger” models.

ARIMA Models

- As distinct from ARMA models. The I stands for integrated.
- An integrated autoregressive process is one with a characteristic root on the unit circle.
- Typically researchers difference the variable as necessary and then build an ARMA model on those differenced variables.
- An ARMA(p,q) model in the variable differenced d times is equivalent to an ARIMA(p,d,q) model on the original data.

SERIE STORICHE



OBBIETTIVO: Dati i valori y_1, \dots, y_T cerco di prevedere i valori al tempo $T+1$

$$\hat{Y}_{T+1} = f(Y_t, Y_{T-1}, \dots, Y_1, \dots)$$

↳ Nei modelli teorici posso considerare l'infinito
passato

Realizzazione di un processo stocastico

$$\dots, Y_1, Y_2, \dots, Y_T, \dots$$

$(Y_t)_{\substack{t \in T \\ t \in N}}$ è una succ. dr v.c. Y_t ordinate (con l'ordine)

PROBLEMA Come faccio a riassumere infinite v.c.?

Th. di Kolmogorov

Si è un distrib. congi. finita dimensionale $F_{(Y_1, \dots, Y_n)}(y_1, \dots, y_n)$ delle v.c. (Y_1, \dots, Y_n)

finiti e t_1, \dots, t_n distinti

L'insieme di tutte queste distrib. $\mathcal{L} = \{F_{(Y_{t_1}, \dots, Y_{t_n})}(y_1, \dots, y_n)\}$ si chiama legge del processo stocastico e queste è l'informazione massima

Quindi: non è necessario sapere la distribuzione infinita ma bastano le possibili distribuzioni finite

Analisi del II ordine

Ci concentriamo sui primi 2 momenti

$$E[Y_t] = \mu_t$$

$$V[Y_t] = \sigma_t^2$$

$$\text{Cov}(Y_{t_i}, Y_{t_j}) = \gamma(t_i, t_j) \xrightarrow{\text{con } i \neq j} f_t \text{- di autocovarianza}$$

$$\text{Cov}(Y_t, Y_{t+1}) \neq 0 \rightarrow Y_{t+1} = \beta_0 + Y_t + \varepsilon_t$$

Posto $t_i = t$ e $t_j = t - k \rightarrow$ Cambio la parametrizzazione

Allora $\text{Cov}(Y_t, Y_{t-k}) = \gamma(t, k) \quad \forall t \neq k = \pm 1, \pm 2, \dots$

Normalmente abbiamo una sola realizzazione (y_1, \dots, y_T) di Y_1, \dots, Y_T e con queste realizzazioni
dobbiamo calcolare i momenti $E[Y_t] = \mu_t \rightarrow$ Non ha senso calcolarlo perché ho solo una realizzazione
 $= 2 + \beta \cdot t$

Aggiungo ipotesi per ridurre l'eterogeneità nel tempo dei parametri:

Processi che hanno memoria infinita $\gamma(t, k) \neq 0$ fix. $t \neq k$

Se la memoria non è infinita c'è correzione asintotica e si può modellizzare stimando molti meno
parametri di μ_t

Il processo è stationario se è omogeneo nel tempo

Devo ridurre l'eterogeneità nel tempo

Devo verificare che abbia memoria corta il processo

PROCESSO STOCASTICO STAZIONARIO

In senso forte

$(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}) \stackrel{d}{=} (Y_{t+h}, Y_{t+2h}, \dots, Y_{t+nh}) \rightarrow$ traslazione rigida dei periodi
Le distribuzioni congiunte sono uguali: \Rightarrow può anche essere negl.

Con $n=1$

$$F_{Y_t}(y_1) = F_{Y_{t+h}}(y_1) \rightarrow \text{id} \rightarrow \text{condizione di stazionarietà in senso forte}$$

$$E[Y_t] = \mu$$

$$V[Y_t] = \sigma_y^2$$

$$\text{Cov}(Y_t, Y_{t-k}) = \gamma(t, k) = \gamma(|k|) = \gamma(k)$$

in generale no iid

$$F_{(Y_{t_1}, \dots, Y_{t_n})}(y_1, \dots, y_n) = F_{(Y_{t+h}, \dots, Y_{t+nh})}(y_1, \dots, y_n) \quad t_1 \geq t, t_1, \dots, t_n \neq h$$

\rightarrow Costanti ma non obbligatoriamente finite

$\rightarrow \text{Cor } K \gg$

Stazionario ≠ costante

stazionario = legge costante nel tempo

Tutti i momenti sono costanti perché sono iid

In senso debole (in covarianza)

Condizioni:

- 1) $E[Y_t] = \mu$
- 2) $V[Y_t] = \sigma_y^2$
- 3) $\text{Cov}(Y_t, Y_{t-k}) = \gamma(k)$

$\left. \begin{array}{l} \\ \\ \end{array} \right\}$ sono finite

Non necessariamente iid

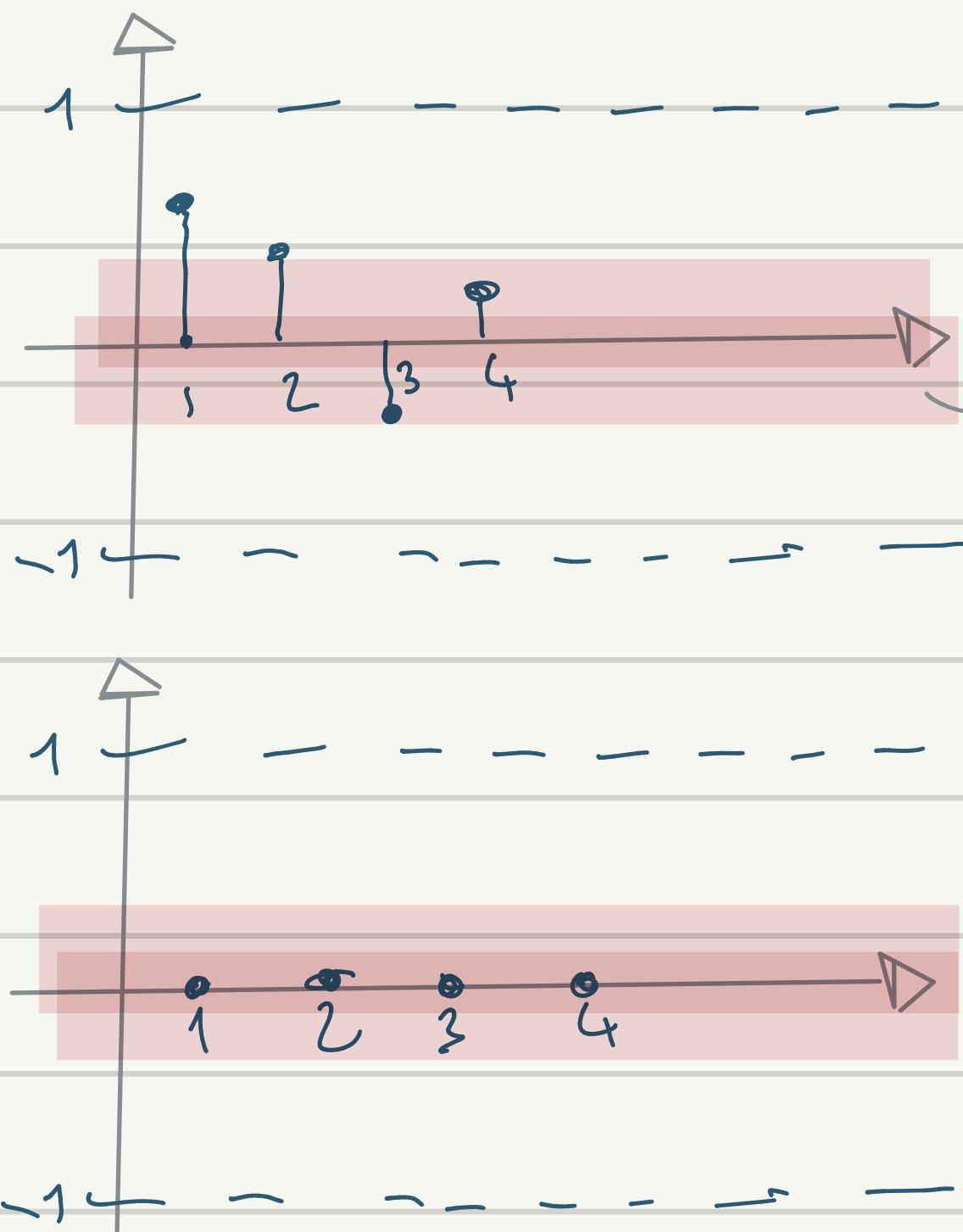
Staz. in senso debole

$$S(k) = \text{corr}(Y_t, Y_{t-k}) = \frac{\gamma(k)}{\sigma_y^2} = \frac{\gamma(k)}{\gamma(0)}$$

$$\text{Corr}(Y_t, Y_{t-k}) = S(t, k) = \frac{\gamma(t, k)}{\sqrt{\sigma_y^2 - \sigma_{t-k}^2}}$$

In senso debole non c'è più la condizione dell'identica distribuzione per le distrib. congiunte

Fz. di autocorrelazione



Correlogramma

→ Bande di confidenze

Processo con assenza di memoria
(White noise)

Processo normale (ϵ_t)^{id}

Le leggi di evoluzione nel tempo è data da distribuzioni finite dimensionali normali
Un processo normale è stazionario in senso forte \Leftrightarrow è stazionario in senso debole
solo in questo caso

→ In questo caso c'è anche l'indipendenza

30/10

Processo White Noise

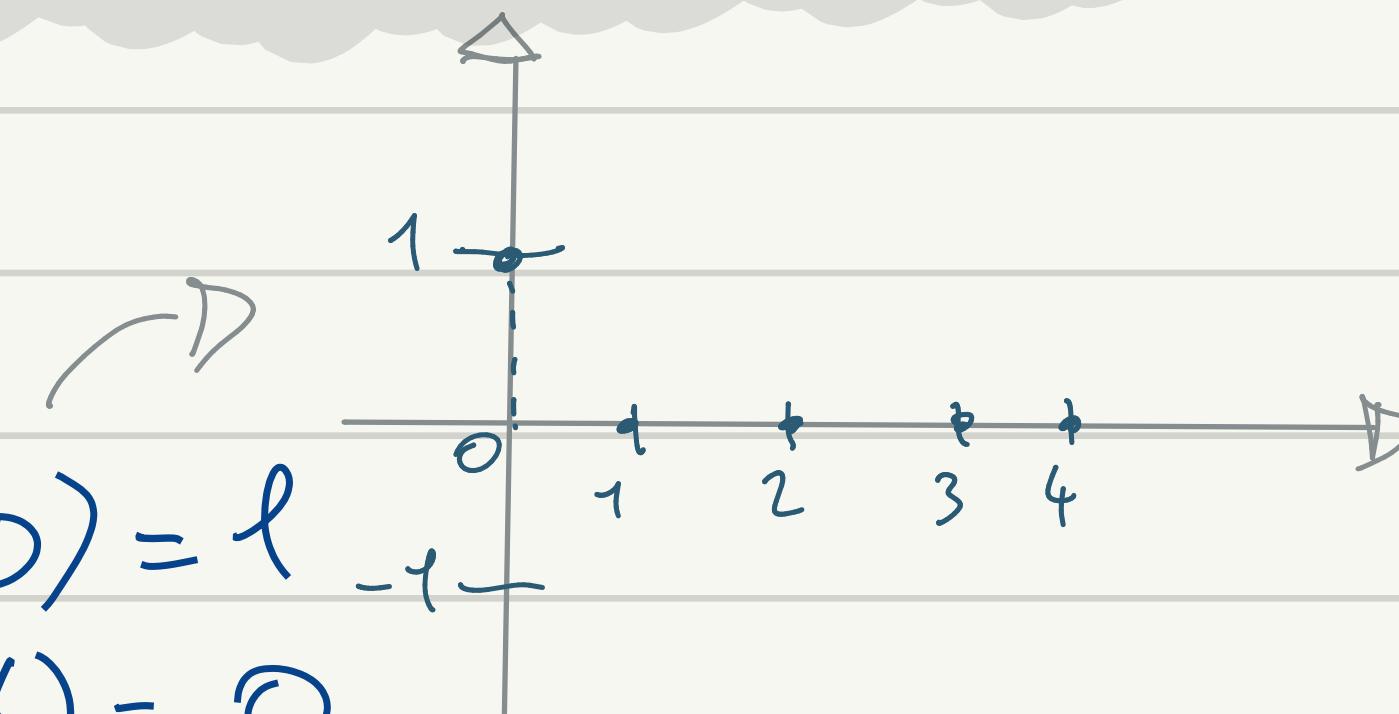
$$\sum_t \sim \text{iid}(0, \sigma_\sum^2) \rightarrow \sum_t \perp \sum_j \quad \text{con } t \neq j$$

$$\text{Con } \left\{ \begin{array}{l} E[\sum_t] = 0 \\ \text{Var}[\sum_t] = \sigma_\sum^2 < +\infty \end{array} \right.$$

$$\left. \begin{array}{l} \text{e} \\ \text{Cov}[\sum_t, \sum_{t-k}] = 0, \quad \forall k \neq 0 \end{array} \right\} \Leftrightarrow \gamma(k) = 0$$

Stationarietà in
senso forte e
debole

! $\sum_t \rightarrow E_t$ sarebbe l'errore



ché per ipotesi sono id

$$\begin{aligned} \gamma(0) &= 1 \\ \gamma(k) &= 0 \end{aligned}$$

Proprietà White Noise

$$\mu_t \sim WN(0, \sigma^2)$$

con \rightarrow Possono essere dipendenti

$$\begin{cases} \text{Cov}(\mu_t, \mu_{t-k}) = 0 & \forall k \neq 0 \\ E[\mu_t] = 0, \quad \forall t \\ \text{Var}(\mu_t) = \sigma_\mu^2 < \infty, \quad \forall t \end{cases}$$

Non sappiamo se c'è indipendenza, non sono iol

$$\begin{aligned} \varepsilon_t &\sim WN(0; \sigma_\varepsilon^2) \rightarrow \varepsilon_t \stackrel{iid}{\sim} N(0; \sigma_\varepsilon^2) \\ (\varepsilon_t)_t &\sim P. \text{NORMALE} \rightarrow \varepsilon_t \sim iid \end{aligned}$$

?

$$Y_t = \beta_0 + \varepsilon_t \quad \rightarrow \quad Y_t \perp Y_{t-k} \quad t=1, \dots, T$$

$$\varepsilon_t \sim i.i.d. (0; \sigma_\varepsilon^2)$$

$$E[Y_t] = \beta_0$$

$$\text{Var}[Y_t] = \sigma_\varepsilon^2$$

$$\text{Cov}(Y_t, Y_{t+k}) = 0$$

$$k \neq 0$$

$$\hat{\mu}_Y = \left(\frac{1}{T} \sum_{t=1}^T Y_t \right) \xrightarrow{d} E[Y_t] = \beta_0 = \mu_Y \quad \text{Cov}(Y_t, Y_{t-k}) = 0$$

$$\text{Var}[Y_t] = \sigma_\varepsilon^2$$

$$\sqrt{T}(\hat{\mu}_Y - \mu_Y) \xrightarrow{d} N(0; \sigma_{\hat{\mu}_Y}^2)$$

$$\hat{\mu}_Y \underset{T \gg 0}{\approx} N(\mu_Y; \frac{\sigma_Y^2}{T})$$

$$\text{Se voglio stimare la varianza: } S_Y^2 = \left(\frac{1}{T-1} \right) \sum_{t=1}^T (Y_t - \hat{\mu}_Y)^2$$

$\xrightarrow{t \rightarrow \infty} \sigma_Y^2$

$$\hat{\sigma}_{Y_{ML}}^2 = \frac{T-1}{T} S_Y^2 \rightarrow \sigma_Y^2$$

$$\hat{\mu}_y \xrightarrow{P} E[\hat{Y}_t] = \mu_y$$

↳ momento teorico o spaziale
↳ momento campionario o temporale

In teorie avendo più realizzazioni potrei stimare $E[\hat{Y}_t] = \frac{1}{N} \sum_{j=1}^N Y_{jt}$

Ci può essere correlazione nel tempo e il momento campionario può non essere un stimatore affidabile del momento teorico

In caso ci sia correlazione:

$$Var(\hat{\mu}_y) = \frac{1}{n} (Var(Y_1) + Var(Y_2) + Cov(Y_1, Y_2))$$

→ 1. E' finita?
2. Le sue varianze convergono a 0? } → Consistenza

Ergodicità

L'ergodicità del processo garantisce la consistenza degli stimatori dei momenti campionari.

$$E[g(Y_t)] \rightarrow \frac{1}{T} \sum_{t=1}^T g(Y_t)$$

↳ momenti marginali

Un processo si dice ergotico per il momento r -esimo finito sse $\frac{1}{T} \sum_{t=1}^T Y_t^r \xrightarrow[T \rightarrow \infty]{P} E[Y_t^r]$

L'ergodicità ha a che fare con le memorie del processo

$S(k) \xrightarrow[k \rightarrow \infty]{P} 0 \longrightarrow$ asintotica incorrelazione

E consistente

$$\frac{1}{T} \sum_{t=1}^T Y_t^r \xrightarrow[T \rightarrow \infty]{P} E[Y_t^r]$$

Processo asintotico indipendente

Stock e Watson garantiscono oltre all'asintotica incovariante anche l'asintotica indipendenza
 → Vale anche a livello vettoriale ($\underline{Y}_t, \underline{Y}_{t+k}$)

$$\underline{Y}_t \quad \underline{Y}_{t+k} \quad \begin{array}{l} \text{Con } K \text{ piccolo assumiamo} \\ \text{Con } K \rightarrow +\infty \end{array} \lim_{k \rightarrow +\infty} \left[\frac{F_{Y_t, Y_{t+k}}(y_1, y_2) - F_{Y_t}(y_1) \cdot F_{Y_{t+k}}(y_2)}{F_{Y_t}(y_1) \cdot F_{Y_{t+k}}(y_2)} \right] = 0 \quad h(y_1, y_2)$$

errore, differente di prob.

TH Si è $(Y_t)_t$ staz. in senso forte e asintotic. indipendente con $E[Y_t^n]$ finiti $n=1, \dots, r$
 → $(Y_t)_t$ è ergotico per il momento $E[Y_t^n]$ $n=1, \dots, r$

$$\hat{\gamma}(k) = \frac{1}{T} \sum_{t=k+1}^T (Y_t - \hat{\mu}_y)(Y_{t-k} - \hat{\mu}_y)$$

↳ Stime affidabili per k piccolo, meno affidabili per $K \rightarrow T$
 Più è grande K e più osservazioni si perdono

Se c'è stazionarietà in covarianze

$$Var(\underline{Y}) = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{bmatrix} = \Sigma \quad \text{con } \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

↳ def. pos.

$\hat{\Sigma} = [\hat{\gamma}(k)]$ è def. positiva

$$\hat{S}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} \quad \hookrightarrow S(k)$$

↳ Dovrò fare inferenze sugli stimatori

Modelli AR(1)

Modelli moving average

MA(1)

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

La media mobile è una combinazione lineare di osservazioni temporali vicine \rightarrow sono indip.

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1} \rightarrow \text{Var}[Y_t] = E[(\varepsilon_t + \theta \varepsilon_{t-1}) - E[\varepsilon_t + \theta \varepsilon_{t-1}]]^2 = E[(1+\theta) \cdot \varepsilon_t]^2 = (1+\theta)^2 \sigma_\varepsilon^2$$

\hookrightarrow Il processo è correlato nel tempo però solo con le variabili contigue, già con quelle successive non lo sarà. Utile per regolarizzare la serie, come nel caso delle stagionalità

$$P_q(k) = \begin{cases} 1 & k=0 \\ \frac{\theta}{(1+\theta)^2} & k=1 \\ 0 & k>1 \end{cases}$$

\hookleftarrow Ho creato una correlazione che non c'era, però se nel corologramma vedo questa struttura posso inferire che il processo è MA(1)

$$(\varepsilon_t)_t \rightarrow MA(1) \rightarrow (Y_t)_t$$

INPUT
 $WN(0, \sigma_\varepsilon^2)$

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

OUTPUT

$$E[Y_t] = 0$$

$$\text{Var}[Y_t] = \sigma_\varepsilon^2 + \theta^2 \sigma_\varepsilon^2 = (1+\theta^2) \sigma_\varepsilon^2$$

$$P(2) = \frac{P(1)}{P(0)} = \frac{\text{Cov}(Y_t, Y_{t-1})}{\text{Cov}(Y_0, Y_t)} = \frac{\theta \sigma_\varepsilon^2}{(1+\theta^2) \sigma_\varepsilon^2} \rightarrow \text{Var}(Y_t)$$

$$Y_t(t,k) = S_q(|k|) = \begin{cases} \sigma_\varepsilon^2 (1+\theta^2) & k=0 \\ \sigma_\varepsilon^2 \theta & k=1 \\ 0 & k>1 \end{cases}$$

$$\begin{cases} Y_t = \varepsilon_t + \theta \varepsilon_{t-1} \\ Y_{t-k} = \varepsilon_{t-k} + \theta \varepsilon_{t-k-1} \end{cases}$$

$$\text{Cov}(Y_t, Y_{t-k}) = \begin{cases} \sigma_\varepsilon^2 (1+\theta^2) & \text{se } k=0 \\ \sigma_\varepsilon^2 \theta & \text{se } k=1 \\ 0 & \text{se } k>1 \end{cases}$$

31/10

$Y_t \sim MA(1)$ è sempre stat. in cov.

$$\text{Cov}(Y_t, Y_{t-1}) = E[\varepsilon_t + \theta \varepsilon_{t-1}, \varepsilon_{t-1} + \theta \varepsilon_{t-2}] = E[\varepsilon_t \varepsilon_{t-1}] + \theta E[\varepsilon_{t-1} \varepsilon_{t-2}] + \theta E[\varepsilon_t \varepsilon_{t-2}]$$

con $E[\varepsilon_t] = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$

MA(2)

$$\psi_t \sim MA(2)$$

$$\psi_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

$$E[\psi_t] = 0$$

$$Var[\psi_t] = \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2)$$

$$\begin{cases} \psi_q(1) = \sigma_\varepsilon^2 (\theta_1 + \theta_1 \cdot \theta_2) \\ \psi_q(2) = \sigma_\varepsilon^2 \cdot \theta_2 \end{cases}$$

$$\begin{cases} \psi_q(k) = 0 & k \geq 2 \end{cases}$$

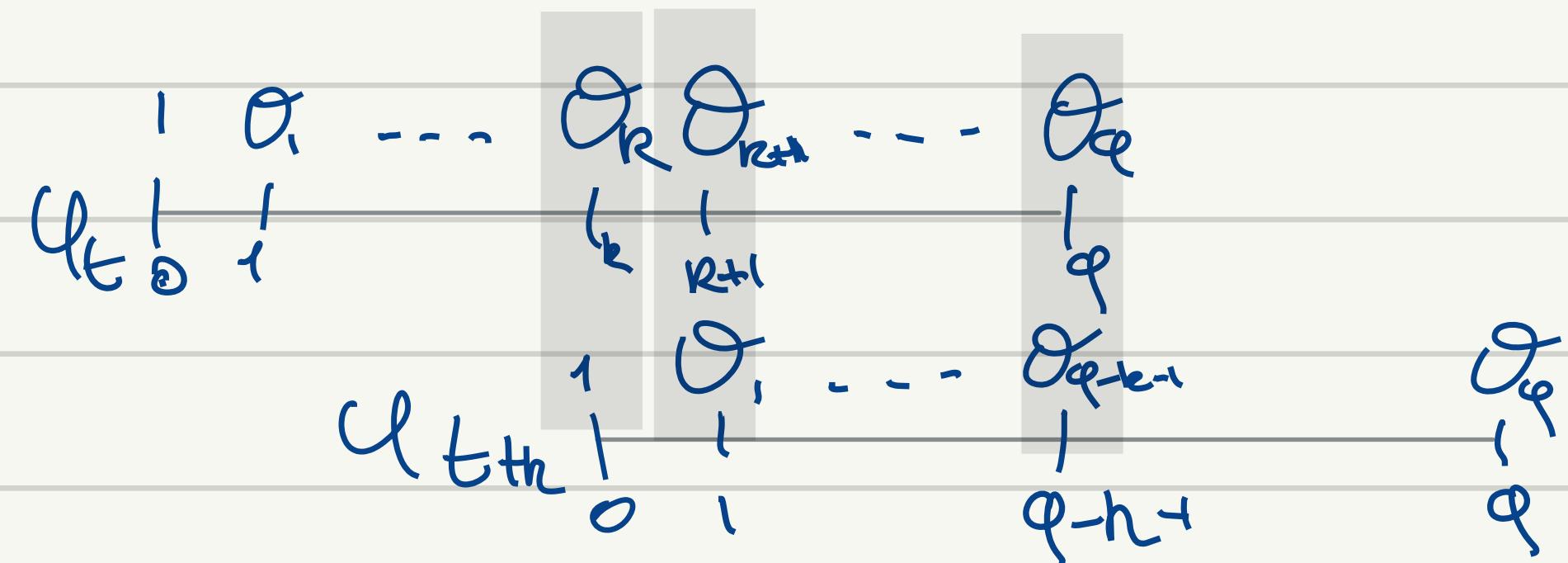
$$\begin{aligned} &\text{Se } \theta_1 = 0 \rightarrow \psi_t = \varepsilon_t \\ &\psi_t = \varepsilon_t + \theta_2 \varepsilon_{t-2} \end{aligned}$$

MA(q)

$$\psi_t \sim MA(q)$$

Stazionario in covarianza

$$\text{con } \psi_q(k) = \begin{cases} \theta_k + \theta_{k+1}\theta_1 + \dots + \theta_q\theta_{q+k-1} & k = 1, \dots, q \\ 0 & k > q \end{cases}$$



$$\psi_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad \text{con } \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

MA(∞)

$$\psi_t \sim MA(\infty)$$

$Var(\psi_t)$ finita e costante

$$\psi_t = \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i} \quad \text{con } \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

con $\psi_0 = 1$

Condizione di assoluto sommabilità
 $\sum_{i=0}^{+\infty} |\psi_i| < +\infty$

$$E[\psi_t] = 0 \quad Var[\psi_t] = \sum_{i=0}^{+\infty} \psi_i^2 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \sum_{i=0}^{+\infty} \psi_i^2$$

$\Rightarrow K$ finito

$\sum_{i=0}^{+\infty} \psi_i^2 < +\infty$

CONDIZIONE DI QUADRATO SOMMABILITÀ

Condiz. più forte garantisce l'ergodicità del t-momento

$$S(k) = E[\psi_t \cdot \psi_{t-k}] = E\left[\left(\sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i}\right) \left(\sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-k-i}\right)\right]$$

→ Quando le seq. num. $t-i$ è diverse de $t-k-i$: \neq

$$= \sigma_\varepsilon^2 \sum_{i=0}^{+\infty} \psi_i \cdot \psi_{i+k}$$

con $\psi_0 = 1$

$$S(k) \xrightarrow{k \rightarrow +\infty} 0$$

TH. di WOLD (38)

Ogni processo stazionario in cov. con $E[\psi_t] = 0$ si può rappresentare come

$$\psi_t = \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i} + z_t \quad \psi_0 = 1, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2) \quad \text{e} \quad \sum_{i=0}^{+\infty} \psi_i^2 < +\infty$$

$z_t \perp \varepsilon_s \forall t, s$ Processo stocastico linearmente deterministico

$$\hat{z}_t = g(z_{t-1}, \dots)$$

$g(\cdot)$ f. lineare nei parametri:

$$E[(z_t - \hat{z}_t)^2] = 0 \quad \forall t > t_0 \text{ con } t_0 \text{ finito}$$

Processo stocastico per
prevedibile

Il th. di Wold vale per tutti i processi MA(q) con $q < +\infty$

I processi stazionari in covariante possono essere approssimati come un MA(∞). Questo vale anche per i modelli AR che possono essere rappresentati come un MA(∞).

Modelli Autoressive

ARC(1)

$Y_t \sim \text{ARCI}$

$$Y_t = \phi_1 \cdot Y_{t-1} + \varepsilon_t$$

$$\text{con } \varepsilon_t \sim \mathcal{WN}(0; \sigma^2_\varepsilon)$$

$$Y_t - \phi_1 Y_{t-1} = \varepsilon_t$$

Processo stocastico
Xké questo è stoc.

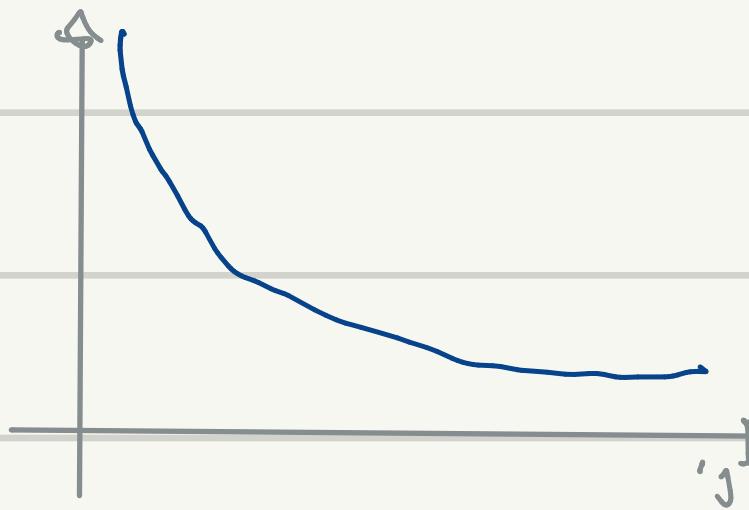
$$\begin{cases} Y_t - \phi_1 Y_{t-1} = 0 & \text{Il ritardo massimo tra gli elementi è di ordine 1 (in questo caso)} \\ Y_0 & \text{Se il ritardo max è } 0 \text{ si dice che è omogenea} \\ \hookrightarrow \text{La soluzione sarà } Y^*(t) \end{cases}$$

$$Y_t = \phi_1^n Y_{t-n} + \sum_{j=0}^{n-1} \phi_1^j \varepsilon_{t-j}$$

$$\text{Se } |\phi_1| < 1 \quad \phi_1^n \xrightarrow{n \rightarrow \infty} 0$$

Processo a media mobile d'ordine infinito

$$X_t = \sum_{j=0}^{+\infty} (\underbrace{\phi_1^j}_{\psi_j}) \varepsilon_{t-j} \sim \text{MA}(\infty)$$



è stazionario in covarienza se $\sum_{j=0}^{+\infty} \psi_j^2 < +\infty$ sse $|\phi_1| < 1$

$$\sum_{j=0}^{+\infty} (\phi_1^j)^2 = \sum_{j=0}^{+\infty} (\phi_1^2)^j \quad \rightarrow \quad \sum_{j=0}^{+\infty} \alpha^j = 1 + \alpha + \alpha^2 + \dots = \frac{1}{1-\alpha} = \left(\frac{1}{1-\phi_1^2} \right)$$

\hookrightarrow Progressione geom con $a = \phi_1^2$

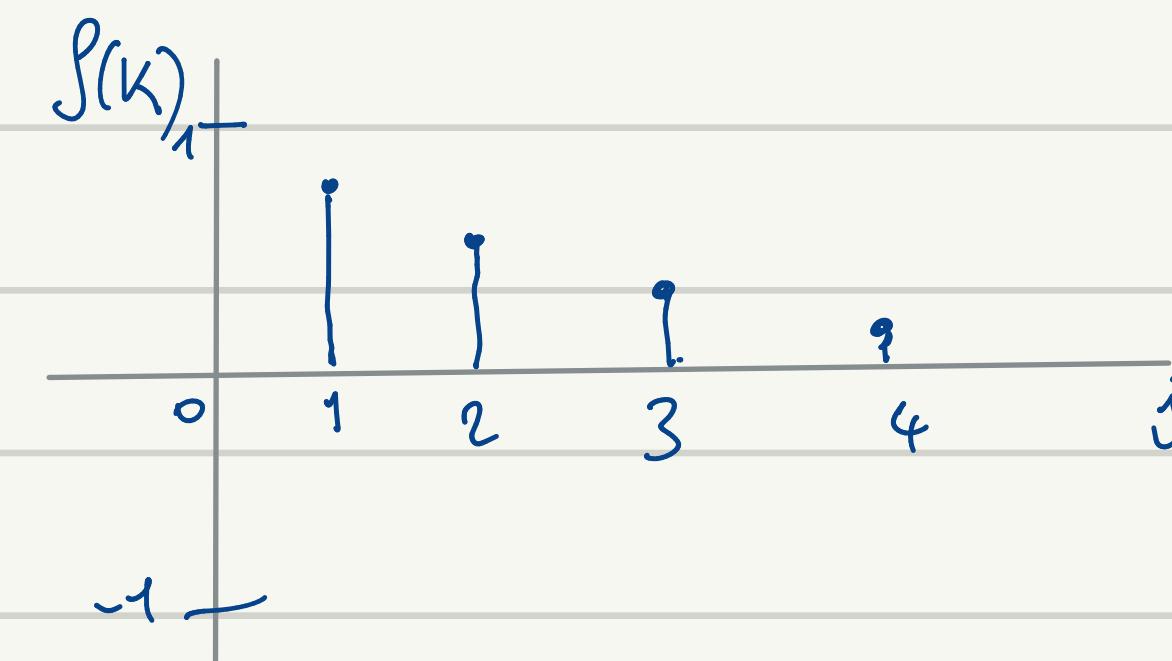
Con

$$E[X_t] = 0$$

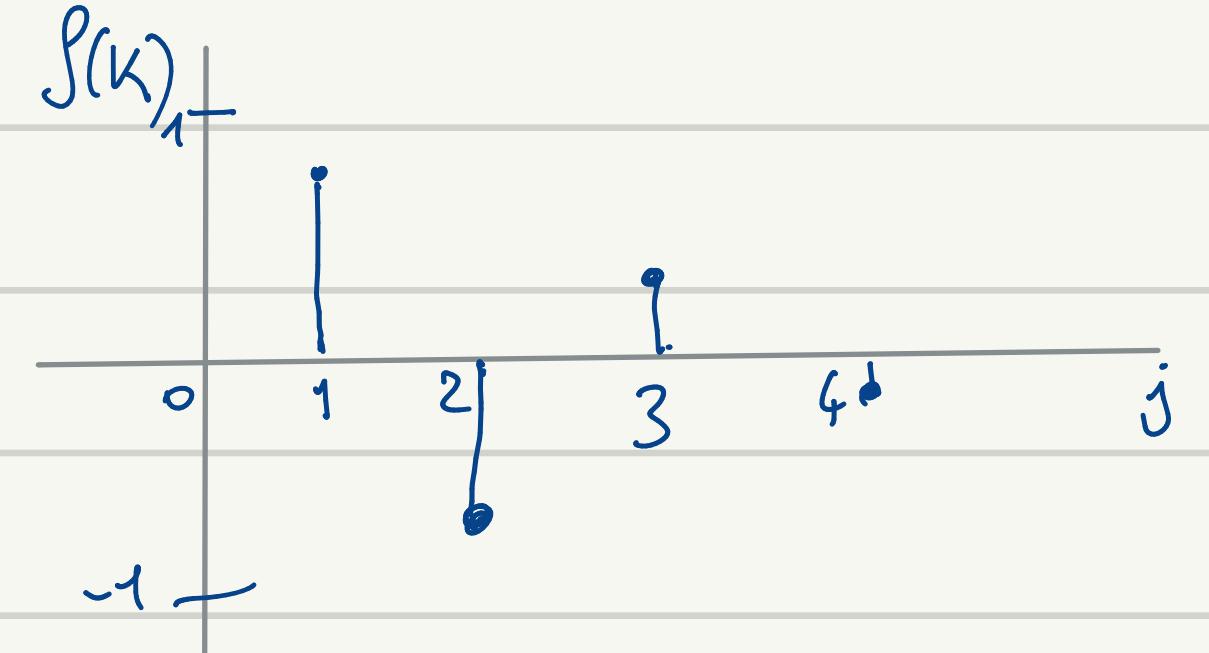
$$\text{Var}[X_t] = \sigma_\varepsilon^2 \cdot \left(\frac{1}{1-\phi_1^2} \right)$$

$$\gamma(k) = \text{Cor}(X_t, X_{t-k}) = \frac{\sigma_\varepsilon^2 \cdot \phi_1^k}{1-\phi_1^2}$$

$$0 < \phi_1 < 1$$



$$-1 < \phi_1 < 0$$



$\rho(k) = \phi_1^k \quad |\rho(k)| \xrightarrow{k \rightarrow +\infty} 0$ in modo esponenziale
Il processo è asintoticamente ^{→ incorrelato} incorrelato

Si ipotizza che $\gamma_0 \perp \nexists \varepsilon_1, \varepsilon_2, \dots, \varepsilon_t, \dots$

$$\gamma_1 = \phi_1 \gamma_0 + \varepsilon_1$$

$$\gamma_2 = \phi_1 \gamma_1 + \varepsilon_2 = \phi_1^2 \gamma_0 + \phi_1 \varepsilon_1 + \varepsilon_2$$

⋮

$$\gamma_t = \underbrace{\phi_1^t \gamma_0}_{\substack{\hookrightarrow \text{Non è mai correlato con gli epsilon futuri}}} + \sum_{j=0}^{t-1} \phi_1^j \varepsilon_{t+j}$$

↪ Non è mai correlato con gli epsilon futuri

Altre condizioni iniziali $\rightarrow \gamma_0$ v.c. $E[\gamma_t] = \phi_1^t E[\gamma_0]$

$$\text{Var}[\gamma_t] = \phi_1^{2t} \text{Var}[\gamma_0] + \sigma_\varepsilon^2 \sum_{j=0}^{t-1} \phi_1^{2j}$$

$$\text{Cov}(\gamma_{t+k}, \gamma_t) = \phi_1^{2t+2k} \text{Var}(\gamma_0) + \phi_1^k \frac{(1-\phi_1^{2t})}{1-\phi_1^2} \sigma_\varepsilon^2 \quad \begin{array}{l} \hookrightarrow \text{Dipende da } t \text{ e } k \rightarrow \text{Non è stazionario} \\ \text{Se l'ipotesi iniziale è generica} \end{array}$$

↪ Aggiungo un'ipotesi

Se $|\phi_1| < 1$, $t \rightarrow +\infty$

$$\lim_{t \rightarrow +\infty} E[\gamma_t] = 0, \quad \lim_{t \rightarrow +\infty} \text{Var}(\gamma_t) = \left(\frac{\sigma_\varepsilon^2}{1-\phi_1^2} \right),$$

$$\text{Cov}(\gamma_{t+k}, \gamma_t) = \frac{\phi_1^k \cdot \sigma_\varepsilon^2}{1-\phi_1^2} \quad \begin{array}{l} \hookrightarrow \text{Non dipende più da } t \\ \text{R} \end{array}$$

$$t \gg 0 \quad Y_t \approx MA(\infty) \quad X_t = \sum_{j=0}^{+\infty} \phi_j^j \varepsilon_{t-j}$$

$$E[Y_t - X_t]^2 \xrightarrow[t \rightarrow +\infty]{} 0 \quad \text{Allora} \quad Y_t \xrightarrow[m.q.]{E \rightarrow +\infty} X_t \quad \text{media quadratica}$$

Se scelgo in modo opportuno

$$\begin{aligned} Y_0 &\text{ v.c. con} \\ E[Y_0] &= 0 \\ \text{Var}[Y_0] &= \frac{\sigma_\varepsilon^2}{1-\phi_1^2} \end{aligned}$$

$$\begin{aligned} \text{Allora} \quad E[Y_0] &= 0 \\ \text{Var}[Y_0] &= \frac{\sigma_\varepsilon^2}{1-\phi_1^2} \\ \text{Cov}(Y_{t+k}, Y_t) &= \frac{\phi_1^k \cdot \sigma_\varepsilon^2}{1-\phi_1^2} \quad \text{f.k.} \quad \forall t=0, 1, \dots \end{aligned}$$

$$L \rightarrow L_{\text{lag}}$$

\rightarrow Applicazione che da una succ. che è il ritardo di y_t

$$\forall t, X_t = L(Y_t) = Y_{t-1}$$

$$(L \cdot Y_t = Y_{t-1}) \rightarrow L^k Y_t = Y_{t-k}$$

\downarrow operatore ritardo

AR(1)

$$Y_t - \phi_1 Y_{t-1} = \varepsilon_t$$

$$L^0 Y_t - \phi_1 L^1 Y_t = \varepsilon_t \quad \rightarrow \quad Y_t = \frac{\varepsilon_t}{L^0 - L^1 \phi_1}$$

Polinomio Auto regressivo

$$Y_t (L^0 - L^1 \phi_1) = \varepsilon_t \quad \rightarrow \quad Y_t \phi(L) = \varepsilon_t$$

\downarrow Polinomio dell'operatore ritardo

Eq. caratteristica del processo AR(1)

$$\phi(z) = \ell - \phi, z$$

$$\left\{ \begin{array}{l} \phi(z) = 0 \\ 1 - \phi_i z = 0 \end{array} \right. \quad \xrightarrow{\text{Reduce inverse di } z_i} \quad \frac{1}{\phi_i} = z_i \quad \Rightarrow \quad \phi(z) = C$$

$$\phi(z) = \frac{(1 - \phi_1 z - \dots - \phi_n z^n)}{\prod_{i=1}^n (z - \lambda_i)}$$

AR(1) asymptotic stat. use $|z| > 1 \rightarrow |\phi_1| < 1$ sse $|\lambda_1| > 1$

$$AR(p) \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad \text{con } \varepsilon_t \sim WN(0; \sigma^2_\varepsilon)$$

$$\cos \phi_p(\text{CL}) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$$

$$\underline{\phi}_p(z) = 0 \quad z \in C$$

↳ Ammette p radici in campo complesso z_1, z_2, \dots, z_p

$(Y_t)_{t \geq 0} \sim AR(p)$ è stat. in cov. se $|z_i| > 1$, $i = 1, \dots, p$

$$\lambda_i = 2^i \quad \text{and} \quad |\lambda_i| < 1$$

$$\phi_1(\ell) = 0$$

1

$z_1 = a + bi$ $z_2 = a - bi$ numeri complessi coniugati

$$z_1 = \alpha - bi \quad |z_1| = |z_2|$$

$$|\lambda| = g^{-1} = \frac{1}{|\gamma|}$$

$|z| > 1$ sse $|x| < 1$

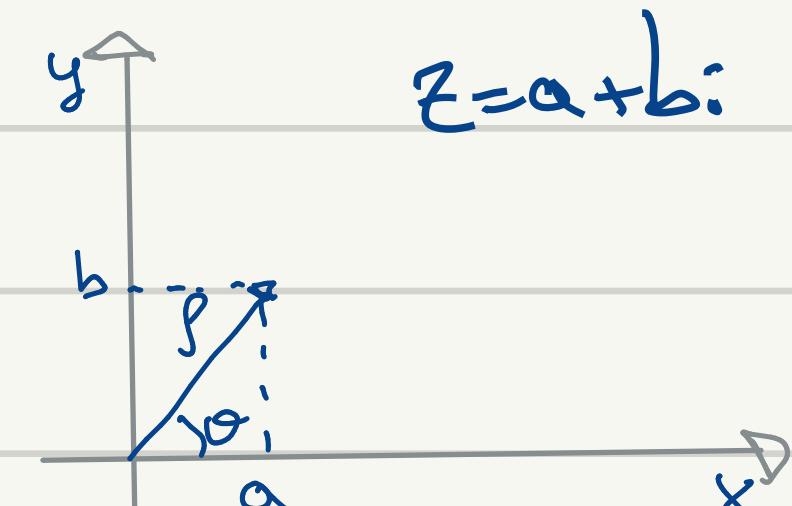
$$z = a + bi$$
$$|z| = \rho = \sqrt{a^2 + b^2} \geq 0$$

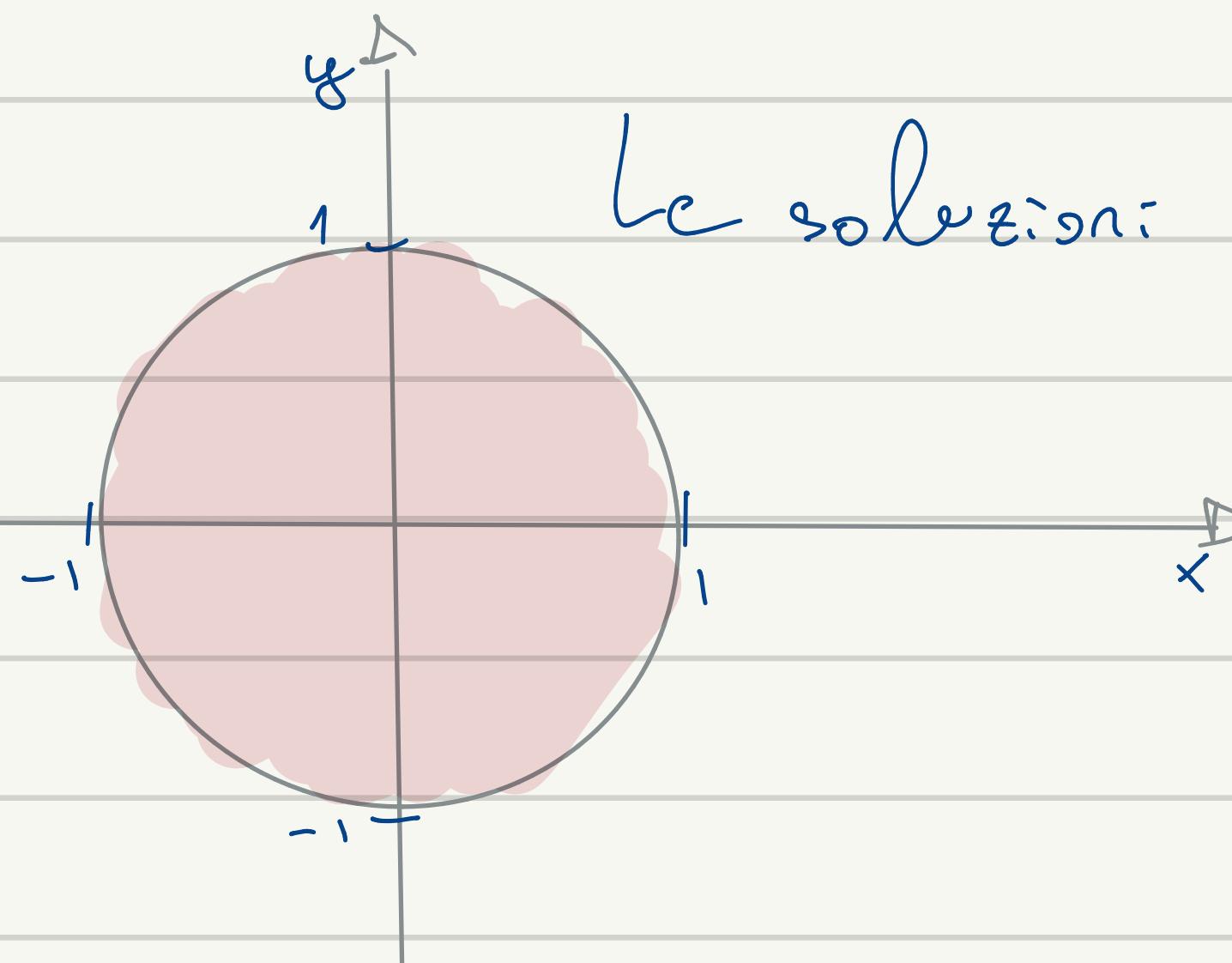
$$z = a + bi = \rho [\cos(\theta) + i \sin(\theta)]$$

$$z^t = \rho^t [\cos(t\theta) + i \cdot \sin(t \cdot \theta)]$$

$$z^{-1} = \rho^{-1} [\cos(-\theta) + i \cdot \sin(-\theta)]$$

$$z = \frac{3}{\cos(-\theta) + i \cdot \sin(-\theta)}$$





Le soluzioni stanno fuori:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t \quad \text{staz. in cov. se } |\phi_1| < 1$$

(solo ϕ_1)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$$\text{AR}(p) \quad \phi_p(L) Y_t = \varepsilon_t \\ (1 - \phi_1 L - \dots - \phi_p L^p) Y_t = \varepsilon_t$$

Se $|\phi_1| < 1$

Polinomio inverso

$$(1 - \phi_1 L)^{-1} (1 - \phi_1 L) Y_t = (1 - \phi_1 L)^{-1} \varepsilon_t$$

$$Y_t = (1 - \phi_1 L)^{-1} \varepsilon_t \sim MA(\infty)$$

$$\begin{aligned} &= \sum_{j=1}^{+\infty} \phi_1^j \varepsilon_{t-j} \\ &= [\sum_j \phi_1^j L^j] \varepsilon_t \\ &= \Psi_\infty(L) \varepsilon_t \end{aligned}$$

Def. polinomio inverso nell'espress.
ritardo L del polinomio quel poli-
nomio (di ordine infinito) t.c.
 $\phi(L)^{-1} \cdot \phi(L) = \phi_p(L) \cdot \phi_p(L)^{-1} = 1$

$$\Psi_\infty(L) = \sum_{j=1}^{+\infty} \psi_j L^j$$

con $\psi_0 = 1$
 $\psi_j = \phi_1^j$

T.h.
Dato $\phi_p(L)$ se $\exists \Psi_\infty(L)$ t.c. $\Psi_\infty(L) \cdot \phi_p(L) = 1$ allora $\Psi_\infty(L) = \phi_p(L)^{-1}$

Condizione di invertibilità del polinomio
 $\phi_p(L)$ è invertibile se $|\phi_1| > 1$:

La condizione di invertibilità è la stessa delle condizioni di stazionarietà (debole)
in covarianza del processo

Nel caso AR(1)

$$(1 - \phi_1 L) \varphi_t = \varepsilon_t \quad |\phi_1| < 1 \quad \varphi_t = \varphi_\infty(L) \cdot \varepsilon_t \sim MA(\infty) \text{ staz. in cov.}$$

$$(1 - \phi_1 L)^{-1} = \varphi_\infty(L) = \sum \phi_i^i L^i = 1 + \phi_1 L + \phi_1^2 L^2 + \dots$$

AR(p)

$$\phi_p(L) \varphi_t = \varepsilon_t \quad \text{staz. in cov. sse } |\phi_p(L)|^{-1}$$

$$\phi_p(L)^{-1} \cdot \phi_p(L) \cdot \varphi_t = \phi_p(L)^{-1} \cdot \varepsilon_t$$

staz. in cov.

$$MA(\infty) \quad \varphi_t = \phi_p(L)^{-1} \cdot \varepsilon_t$$

$$\phi_p(L) = \prod_{i=1}^p (1 - \lambda_i L) \quad \phi_p(L)^{-1} = \prod_{i=1}^p (1 - \lambda_i L)^{-1} \quad \text{con } (1 - \lambda_i L)^{-1} = 1 + \lambda_i L + \lambda_i^2 L^2 + \dots$$

Es. $p=2$
 $\exists i \text{ sse } |\lambda_i| < 1$

$$\varphi_\infty(L) = (1 + \lambda_1 L + \lambda_1^2 L^2 + \lambda_1^3 L^3 + \dots)(1 + \lambda_2 L + \lambda_2^2 L^2 + \dots) = (1 + \lambda_2 L + \dots) + (\lambda_1 L + \lambda_1 \lambda_2 L^2 + \lambda_1 \lambda_2^2 L^4 + \dots) + \dots$$

Esistono formule ricorsive che approssimano i risultati:

$$= 1 + \lambda_2 L + (\lambda_1 \lambda_2 + \lambda_2^2) L^2 + \dots$$

Equazioni di YULE-WALKER

$$E[Y_t] = E[\Psi_x(l) \cdot \varepsilon_t] = 0$$

OBIETTIVO: Trovare la funzione di auto covarienza

Ese. AR(2)

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \quad \text{con } \text{cov}(\varepsilon_t, Y_{t-k}) = 0, \quad \forall k > 0$$

$$\text{Var}(Y_t) = E[Y_t \cdot Y_t]$$

$$\text{Cov}(Y_t, Y_{t-k}) = E[Y_t, Y_{t-k}]$$

$$\gamma(0) = \phi_1 E[Y_{t-1} \cdot Y_{t-1}] + \phi_2 E[Y_{t-2} \cdot Y_{t-1}] + E[\varepsilon_t \cdot Y_{t-1}]$$

$$\gamma(1) = \phi_1 E[Y_{t-1} \cdot Y_{t-2}] + \phi_2 E[Y_{t-2} \cdot Y_{t-2}] + E[\varepsilon_t \cdot Y_{t-2}]$$

$$\gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0) + 0$$

:

$$\gamma(k) = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) \quad \forall k > 2$$

$$\rho_1 = \frac{\gamma(1)}{\gamma(0)} = \phi_1 \frac{\gamma(0)}{\gamma(0)} + \phi_2 \frac{\gamma(1)}{\gamma(0)} = \phi_1 + \phi_2 \rho_1 \rightarrow \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

$$\rho_2 = \frac{\gamma(2)}{\gamma(0)} = \phi_1 \frac{\gamma(1)}{\gamma(0)} + \phi_2 \frac{\gamma(0)}{\gamma(0)} = \phi_1 \rho_1 + \phi_2 \rightarrow \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k > 2$$

Asintotica incorrelazione

$$\gamma(k) \xrightarrow{k \rightarrow \infty} 0$$

$$\rho(k) \xrightarrow{k \rightarrow \infty} 0$$

La succ. è convergente e tende a zero se le radici sono < 1

Funzione di autocorrelazione parziale (c.p.)

$$\text{corr}(Y_t, Y_{t-k}) = \beta_k = S(k)$$

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \nu_t$$

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \nu_t$$

$$C.P.(k) = \text{corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k}) = \beta_k$$

$$C.P.(0) = 1$$

$$C.P.(1) = S_1$$

$$C.P.(2) = \hat{\beta}_{22} \longrightarrow Y_t = \beta_{02} + \beta_{12} Y_{t-1} + \beta_{22} Y_{t-2} + \nu_t$$

$$C.P.(3) = \hat{\beta}_{33} \longrightarrow Y_t = \beta_{03} + \beta_{13} Y_{t-1} + \beta_{23} Y_{t-2} + \beta_{33} Y_{t-3} + \nu_t$$

$$C.P.(k) = \hat{\beta}_{kk}$$

Con ACF vero se è MA o meno
con PACF vero di che ordine p è

9/11

MA(∞)

Condizione d'invertibilità per gli MA mentre per AR bisogna di un altro

Un processo $(Y_t)_t$ stazionario in covarianze è invertibile se si può scrivere come un AR(∞)

$$AR(\infty) \quad Y_t = \sum_{i=1}^{+\infty} \bar{\gamma}_i Y_{t-i} + \varepsilon_t$$

con $\sum_{i=1}^{+\infty} |\bar{\gamma}_i| < +\infty$

MAC(1)

$$\varphi_t = \varepsilon_t + \theta \varepsilon_{t-1} \rightarrow \varepsilon_{t-1} = \varphi_{t-1} - \theta \varepsilon_{t-2}$$

$$\varphi_{t-1} = \varepsilon_{t-1} + \theta \varepsilon_{t-2}$$

$$\varphi_t = \varepsilon_t + \theta \varphi_{t-1} - \theta^2 \varepsilon_{t-2} \rightarrow \varepsilon_{t-j} = \varphi_{t-j} - \theta \varepsilon_{t-j-1}$$

\hookrightarrow Errore

AR(1)

$$\varphi_t = \theta \varphi_{t-1} + \nu_t \quad c.p.(1) = \theta$$

$$\varphi_t = \varepsilon_t + \theta \varphi_{t-1} - \theta^2 (\varphi_{t-2} - \theta \varepsilon_{t-3})$$

:

$$\varphi_t = \theta \varphi_{t-1} - \theta^2 \varphi_{t-2} + \theta^3 \varphi_{t-3} - \dots + (-1)^{j+1} \theta^j \varphi_{t-j} - \theta^{j+1} \varepsilon_{t-j-1} + \varepsilon_t$$

\Rightarrow + se dispesi:

$$\hookrightarrow \text{Se } |\theta| < 1 \rightarrow \theta^j \xrightarrow{j \rightarrow \infty} 0$$

Condizione di invertibilità per MAC(1)

DIC: condizioni di inv.

$\theta(L)$

$$\varphi_t = \varepsilon_t + \theta L \varepsilon_t = (1 + \theta L) \varepsilon_t$$

$$\text{Eq. caratteristica } \theta(z) = 0$$

$$1 + \theta z = 0$$

$$z = -\frac{1}{\theta}$$

$$|z| > 1 \rightarrow |-\theta| < 1 \rightarrow |\theta| < 1$$

$$\theta_q(z) = 0, \quad z \in \mathbb{C} \quad \nexists z_1, \dots, z_p \quad \nexists z_i, \quad |z_i| < 1$$

\rightarrow 3! il polinomio inverso di $\theta_q(L)$, $\theta_q(L)^{-1}: \theta_q(L) = 1$

$$\varphi_t = \theta_q(L) \cdot \varepsilon_t$$

$$\hookrightarrow \theta_q(L)^{-1} \varphi_t = \varepsilon_t$$

AR(∞) stat. incor.

$$\Pi_\infty(L) = 1 - \pi_1 L - \pi_2 L^2 - \dots$$

\uparrow

$$\varphi_t = \sum \pi_i \varphi_{t-i} + \varepsilon_t$$

IDENTIFICAZIONE: se abbiamo una serie storica possiamo identificare degli unici parametri θ e σ^2_ε per un processo ARIMA che lo ve a simile. E lo posso fare grazie ACF e PACF.

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad |\theta| < 1$$

$$E[Y_t] = 0$$

$$V[Y_t] = (1+\theta^2) \sigma^2_\varepsilon$$

$$\text{Cov}(Y_t, Y_{t-k}) = \theta \sigma^2_\varepsilon$$

$$\text{Cov}(Y_t, Y_{t-k}) = 0, \quad k > 1$$

$$z_t = \eta_t + \left(\frac{1}{\theta}\right) - \eta_{t-1}$$

$$\eta_t = \theta \cdot \varepsilon_t \sim WN(0; \theta^2 \sigma^2_\varepsilon)$$

$$\begin{aligned} E[z_t] &= 0 \\ V[z_t] &= \theta^2 \sigma^2_\varepsilon + \frac{1}{\theta^2} \theta^2 \sigma^2_\varepsilon = (1+\theta^2) \sigma^2_\varepsilon \\ \text{Cov}(z_t, z_{t-1}) &= \frac{1}{\theta} \theta^2 \sigma^2_\varepsilon = \theta \sigma^2_\varepsilon \end{aligned}$$

PACF (K)

ARMA(p, q)

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \nu_t$$

$$\text{con } \nu_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

$$\text{con } \varepsilon_t \sim WN(0, \sigma^2_\varepsilon)$$

Possiamo dire che è sempre invertibile e stazionario? No, in generale

$$\phi_p(L) Y_t = \theta_q(L) \cdot \varepsilon_t$$

$$\phi_p(L)^{-1} \cdot \phi_p(L) Y_t = \phi_p(L)^{-1} \cdot \theta_q(L) \cdot \varepsilon_t$$

$$Y_t = \left[\frac{\theta_q(L)}{\phi_p(L)} \right] \varepsilon_t$$

Se $\phi_p(z) = 0$ con $|z_i| > 1$ $i := 1, \dots, p$ → Stazionario in cov

$$Y_t = \phi_p(L)^{-1} \cdot \theta_q(L) \cdot \varepsilon_t$$

$(Y_t)_t$ è invertibile sse $\theta_q(z) = 0$ con $|z_j| > 1$, $j = 1, \dots, q$
 $\rightarrow \theta_q^{-1}(L) \cdot \phi_p(L) Y_t = \varepsilon_t \rightarrow \widehat{\Pi}_\infty(L) Y_t = \varepsilon_t$

ARMA(p, q) staz. in cov. e inv.

→ ACF simile AR(p)

→ PACF simile MAC(q)

AR(p) con $E[Y_t] \neq 0 \rightarrow$ uso β_0

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}$$

$$\nu = E[Y_t] = \beta_0 + \phi_1 E[Y_{t-1}] + \dots + \phi_p E[Y_{t-p}] + E[\varepsilon_t]$$

$$E[Y_t] = \nu = \frac{\beta_0}{1 - \phi_1 - \dots - \phi_p} = \beta_0 \cdot \phi_p(1)^{-1}$$

$$\phi_p(L) Y_t = \theta_q(L) \cdot \varepsilon_t \rightarrow E[Y_t] = \nu_t = \beta_0 \phi_p(1)^{-1}$$
$$L = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

I modelli ARIMA hanno il vantaggio che sono PARSIQUIOSI, ovvero possono rappresentare serie complesse con pochi parametri, es. $(\text{ARIMA}(2,4) \sim \text{ARIMA}(14,4))$.

Hanno, però, 2 PROBLEMI

1. Avendo dentro una parte MA hanno il problema dell'invertibilità

2. Anche se sono stazionari e invertibili possono avere un problema di IDENTIFICAZIONE

Ovvero: il problema delle radici comuni

Problema delle radici comuni

Esempio: un ARIMA(1,1)

Avere delle radici comuni significa avere entro una delle radici AR e una delle radici MA uguali

$$(1 - \phi_1 L) Y_t = (1 + \theta_1 L) \varepsilon_t$$

$$\lambda_1 = \phi_1, \quad \tilde{\lambda}_1 = -\theta_1$$

$$\rightarrow \lambda_1 = \tilde{\lambda}_1 \quad \text{Se } \tilde{\lambda}_1 = -\phi_1,$$

$$\begin{aligned} (1 - \phi_1 L) Y_t &= (1 - \tilde{\lambda}_1 L) \varepsilon_t \\ &= (1 - \phi_1 L) \varepsilon_t \end{aligned}$$

$$Y_t = \varepsilon_t \sim WN(0, \sigma^2_\varepsilon)$$

Ottengo così un processo White Noise quando in realtà è un ARIMA(1,1), questo problema non l'avevi avuto con un MA o AR.

Il correlogramma sarà simile al WN

Quindi per l'identificazione dovrò aggiungere il vincolo che non ci siano radici comuni.
Vale anche se le radici sono simili.

LO/11 (Brooks Chap 6)

Vedere slides per capire come scegliere i parametri p,q guardando l'ACF e il PACF

ARIMA(p, d, q)

Negli ARIMA abbiamo il parametro d che, differenziando la serie d volte, la rende stazionaria.

$$D^d y_t = \underbrace{D \cdot D \cdot \dots \cdot D}_{d \text{ volte}} y_t$$

$$D y_t - y_{t-1} = \varepsilon_t$$

$$\begin{cases} y_t \sim I(d) \\ D y_t \sim ARIMA(p, 1, q) \end{cases} \rightarrow y_t = ARIMA(p, 1, q)$$

Ese. R.W. $\sim I(1)$

$\varepsilon_t \sim I(1) \sim ARIMA(0, 1, 0)$

La serie va sempre detrendata

13/11

$$\varepsilon_t \sim WN(0; \sigma_\varepsilon^2)$$

$$\varepsilon_t \sim iid(0; \sigma_\varepsilon^2) \rightarrow OLS$$

$$\varepsilon_t \sim N_{iid}(0; \sigma_\varepsilon^2) \rightarrow ML$$

$$\varepsilon_t \sim Diffringente (D.H.)$$

OMOSCHEDASTICITÀ \times le è indipendente rispetto al passato

$$Var(\varepsilon_t | \text{passato}) = Var(\varepsilon_t) = \sigma_\varepsilon^2$$

Condizioni del processo D.H.

$$1) E[\varepsilon_t] < +\infty \rightarrow E[\varepsilon_t] \text{ è finito}$$

$$2) E[\varepsilon_t | \mathcal{F}_{t-1}] = 0 \rightarrow \text{Non è un processo prevedibile del passato}$$

$$\hookrightarrow \begin{cases} E[\varepsilon_t] = 0 \\ Cov(\varepsilon_t, \varepsilon_{t-k}) = 0 \end{cases} \rightarrow \text{Proprietà della legge del valore atteso iterata} \rightarrow E[\varepsilon_t] = E[E[\varepsilon_t | \mathcal{F}_{t-1}]]$$

$$3) E[\varepsilon_{t+1} | \mathcal{F}_t] = 0 \rightarrow E[\varepsilon_{t+j} | \mathcal{F}_t] = 0 \quad \forall j \geq 1$$

$$\text{DIM: } \times \text{ la 2) } E[\varepsilon_{t+1} | \mathcal{F}_{t+1}] = 0 \text{ allora } E[\varepsilon_{t+1} | \mathcal{F}_t] = E[E[\varepsilon_{t+1} | \mathcal{F}_{t+1}] | \mathcal{F}_t] = E[0 | \mathcal{F}_{t+1}] = 0 \quad \square$$

\rightarrow Si ha che \mathcal{F}_t è set informativo

ε_t è una D.H. rispetto a \mathcal{F}_t se

$$(i) E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$$

$$\begin{cases} a) E[\varepsilon_t] = 0 \\ b) Cov(X, \varepsilon_t) = 0 \quad \forall X \in \mathcal{F}_{t-1} \\ c) E[\varepsilon_{t+j} | \mathcal{F}_t] = 0, \quad \forall j \geq 1 \end{cases}$$

$$Var(\varepsilon_t | \mathcal{F}_{t-1}) = E[\varepsilon_t^2 | \mathcal{F}_{t-1}] = h(t-1) \rightarrow \text{C'è eteroschedasticità}$$

AR(p)

$$0) Y_t = [S + \sum_{i=1}^p \phi_i Y_{t-i}] + \varepsilon_t$$

$$1) E[\varepsilon_t | \underbrace{Y_{t-1}, Y_{t-2}, \dots}_{\mathcal{F}_{t-1}}] = 0 \rightarrow \varepsilon_t \text{ è D.H. rispetto a } \mathcal{F}_t : \begin{cases} i) E[\varepsilon_t] = 0 \\ ii) Cov(Y_{t-j}, \varepsilon_t) = 0 \quad \forall j \geq 1 \end{cases}$$

$$0) \rightarrow E[Y_t | \mathcal{F}_t] = S + \sum_i \phi_i Y_{t-i}$$

$$2) \begin{cases} 2.1) (Y_t)_t \text{ è staz. (richiede che } \phi_p(L) \text{ è invertibile)} \rightarrow Y_t \sim MA(\infty) \text{ cov. staz. in} \\ 2.2) \text{ ASINTOTICA INDIP.} \end{cases}$$

$$E[Y_t] = \frac{S}{1 - \phi_1 - \dots - \phi_p} = \phi_p^{-1}(1)S$$

$$Y_t = \phi_p^{-1}(L) \cdot S + \phi_p(L)^{-1} \cdot \varepsilon_t$$

$$= \phi_p^{-1}(1)^{-1} \cdot S + \phi_p(L)^{-1} \cdot \varepsilon_t$$

$$3) \mathbb{O} < E[\epsilon^4] < +\infty$$

4) Assenza di perfetta collinearità

5)

$$\mathcal{F}_t = \{\gamma_t, \gamma_{t-1}, \dots\} \quad \mathcal{G}_t = \{\epsilon_t, \epsilon_{t-1}, \dots\}$$

Sono equivalenti $\mathcal{F}_t \subseteq \mathcal{G}_t \Rightarrow \mathcal{F}_t = \mathcal{G}_t$

$$\mathbb{O} = E[\epsilon_t | \mathcal{F}_{t-1}] = E[\epsilon_t | \mathcal{G}_{t-1}] = 0 \quad \mathcal{F}_{t-1} = \mathcal{G}_{t-1}$$

Noti gli γ posso ricavarmi gli ϵ con $\epsilon_t = \gamma_t - \delta - \sum \phi_i \gamma_{t-i}$

→ Posso dire che gli errori non sono autocorrelati

Ese. AR(1)

$$E[\epsilon_t | \gamma_{t-1}] = 0 \rightarrow E[\gamma_t | \gamma_{t-1}] = \delta + \phi_1 \gamma_{t-1}$$

$$E[\gamma_t | \gamma_{t-1}, \gamma_{t-2}] = \delta + \phi_1 \gamma_{t-1} + \phi_2 \gamma_{t-2} \quad \text{in caso di AR(1)} \quad \phi_2 \approx 0$$

Gli errori non sono autocorrelati,
sono white-noise

AR(1)

$$\gamma = X\beta + \epsilon$$

$(n \times 1) \quad (n \times (k+1)) \quad (k+1) \times 1 \quad n \times 1$

~~Per comp~~

$$\left\{ \begin{array}{l} \gamma_1, (\gamma_2, \dots, \gamma_T) \\ \gamma_1 = \delta + \phi_1 \gamma_0 + \epsilon_1 \\ \gamma_2 = \delta + \phi_2 \gamma_1 + \epsilon_2 \\ \vdots \\ \gamma_T = \delta + \phi_T \gamma_{T-1} + \epsilon_T \end{array} \right.$$

Non posso usare queste eq.

→ Avrò $T-1$ eq., con un AR(1)
perdo un'eq.

$$\underline{\gamma} = \begin{bmatrix} 1 & \gamma_1 \\ 1 & \gamma_2 \\ \vdots & \vdots \\ 1 & \gamma_{T-1} \end{bmatrix} \begin{bmatrix} \delta \\ \phi_1 \end{bmatrix} + \underline{\epsilon}$$

AR(p) Generalizzazione a un AR(p)

Aumentano i parametri (p) ma si riducono le dimensioni del campione ($n-p$)

$$\hat{\beta} = \begin{bmatrix} \hat{\delta} \\ \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_r \end{bmatrix} = (X'X)^{-1} X' \underline{y}$$

Stimatore distorto $\rightarrow E[\varepsilon | X] \neq 0$ se $E[\varepsilon_t | X] \neq 0$

E.s.) $E[\varepsilon_2 | X] = E[\varepsilon_2 | Y_1, Y_2, \dots, Y_{t-1}] \neq 0$
 $Y_t = \delta + \phi_1 Y_{t-1} + \varepsilon_t \quad \text{Cov}(Y_t, \varepsilon_t) = \phi_1 \sigma_\varepsilon^2 - \sigma_\varepsilon^2 \neq 0$

MAC(∞) $Y_{t+j} = \sum_{i=1}^{+\infty} \phi_i Y_{t+j-i} \rightarrow \text{Cov}(\varepsilon_t, Y_{t+j}) = \phi_j \sigma_\varepsilon^2$

Per la consistenza basta che $\text{Cov}(Y_{t-j}, \varepsilon_t) = 0 \forall j \geq 1$ $\hat{\beta} \xrightarrow{c} \beta$
Quindi vale anche l'asintotica normalità.

TH-di WALD-YANN

Generalizza l'asintotica normalità, senza la necessità di specificare che il processo sia iid.

$$U_t = (X_t \cdot \varepsilon_t)$$

In questo tipo di modelli non c'è autocorrelazione degli errori, quindi potrò usare le formule di WHITE

AR(p) OLS

$$\varepsilon_t | \mathcal{F}_t \sim N(0, \sigma_\varepsilon^2) \rightarrow \varepsilon_t \sim N_{\text{iid}}(0; \sigma_\varepsilon^2)$$

ML $\hat{\beta} = \frac{\partial \ln f(Y_t | \beta)}{\partial \beta} \rightarrow \hat{\beta}$ ML condiz. \equiv OLS mentre $\hat{\sigma}_{\varepsilon, \text{ML}}^2 = \frac{1}{T} \sum \hat{\varepsilon}_t^2$
Condizionale $f(y_1, \dots, y_T | y_1)$
Completa $f(y_1, \dots, y_T)$
 \hookrightarrow Numeri di parametri da stimare
 $\hookrightarrow f(y_1, \dots, y_T) = f(y_1) \cdot f(y_2, \dots, y_T | y_1)$
 \hookrightarrow AR(1) \hookrightarrow Il peso di questa stima è poco influente

$$Y_t \sim N\left(\frac{\delta}{1-\phi_1}, \frac{\sigma_\varepsilon^2}{(1-\phi_1)^2}\right)$$

14/11 Unicase princ Simul

$$\Sigma_t | \mathcal{F}_{t-1}, \dots \sim D.M.$$

$$E[\Sigma_t | \mathcal{F}_{t-1}] = 0$$

$\gamma_t \sim ARMA(p,q)$ stat. in covariante, invertibile e senza radici comuni

$$\gamma_t \sim MA(\infty)$$

stat. in cov

$$\gamma_t \sim AR(\infty)$$

stat. in cov

$\mathcal{F}_{t-1} \subseteq \mathcal{F}_{t-1}$ sono equivalenti

$\{\varepsilon_{t-1}, \dots\}$

$$E[\varepsilon_t | \mathcal{F}_{t-1}] = E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$$

$$\left\{ \text{Cov}(\varepsilon_t, \gamma_{t-j}) = 0 \quad \forall j > 0 \right.$$

$$\left[\text{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0 \quad \forall j \neq 0 \right]$$

ARMA(1,1)

$$\gamma_t = \delta + \phi, \gamma_{t-1} + \theta, \varepsilon_{t-1} + \varepsilon_t = E[\varepsilon_{t-1} | \mathcal{F}_{t-1}] = \varepsilon_{t-1} \rightarrow = 0$$

$$\gamma_{t+1|t} = E[\gamma_t | \mathcal{F}_{t-1}] = \delta + \phi, \gamma_{t-1} + \theta, E[\varepsilon_{t-1} | \mathcal{F}_{t-1}] + E[\varepsilon_t | \mathcal{F}_{t-1}] = \delta + \phi, \gamma_{t-1} + \theta, \varepsilon_{t-1}$$

Precisione del prezzo (avanti)

$$\gamma_{T+1|T} = \delta + \phi, \gamma_T + \theta, \varepsilon_T \rightarrow \text{Generalizzazione}$$

Per stimare i coefficienti posso usare

NLS, GLS
Normal Least Square
Complete
Condizionale

$$\begin{aligned} \varepsilon_t | \mathcal{F}_{t-1} &\sim N(0; \sigma^2_\varepsilon) \\ \varepsilon_t | \mathcal{F}_{t-1} &\sim N(0; \sigma^2_\varepsilon) \end{aligned}$$

es. in ARMA(1,1)
 $\frac{\phi}{1-\phi} \rightarrow \sigma^2 \sum_{i=1}^n \gamma_i^2$

$$\begin{cases} \gamma_t \sim N(E[\gamma_t]; V_{\gamma_t}) \\ \varepsilon_t \sim N(0; \sigma^2_\varepsilon) \\ \varepsilon_t \text{ iid } N(0; \sigma^2_\varepsilon) \end{cases}$$

Molto spesso le soluzioni non sono lineari e bisogna usare tecniche di ricompionimento

$$(y_1, \dots, y_T | y_1, \varepsilon_1) \sim N(\cdot, \Delta = 0 = E[\varepsilon_t])$$

$$\hat{Y}_{T+1|T} = \hat{\delta} + \hat{\phi}_1 Y_T + \hat{\theta}_1 \hat{\varepsilon}_T$$

Con questi modelli possiamo prevedere anche più passi nel tempo.

Otengo 2 tipi di previsioni:

- Statiche → Sempre 1 passo avanti
- Dinamiche →

PREVISIONE STATICÀ

$$y_1, \dots, y_T, Y_{T+1}, Y_{T+2}, \dots$$

$$\hookrightarrow y_{T+1} = \{y_{T+1}, y_{T+2}, \dots\}$$

$$Y_{T+1|T} \quad Y_{T+2|T+1} \quad Y_{T+j|T+j-1}$$

Possso fare solo 1 passo avanti

Può essere usato per le stime dei coefficienti, non comune nell'ambito pratico.

PREVISIONE DINAMICA

$$y_1, \dots, y_T, Y_{T+1}, Y_{T+2}, \dots$$

$$Y_{T+1|T} = E[Y_{T+1} | \mathcal{F}_T]$$

$$Y_{T+2|T} = E[Y_{T+2} | \mathcal{F}_T]$$

$$Y_{T+j|T} = E[Y_{T+j} | \mathcal{F}_T]$$

→ Facciamo il caso AR(1,1)

$$= \delta + \phi_1 Y_T + \theta_1 \varepsilon_T$$

$$= E[\delta + \phi_1 Y_T + \theta_1 \varepsilon_{T+1} + \varepsilon_{T+2} | \mathcal{F}_T] = \delta + \phi_1 E[Y_{T+1} | \mathcal{F}_T] + \theta_1 = (\delta + \phi_1 \delta) + \phi_1^2 Y_T + \phi_1 \theta_1 \varepsilon_T$$

$$= \delta + \phi_1 E[Y_{T+j-1} | \mathcal{F}_T] + \theta_1$$

$$= \delta (1 + \phi_1 + \phi_1^2 + \dots + \phi_1^{j-1}) + \phi_1^j Y_T + \phi_1^{j-1} \theta_1 \varepsilon_T$$

$$= \delta \sum_{i=0}^{\infty} \phi_1^i = \delta \cdot \left(\frac{1}{1 - \phi_1} \right) = E[Y_T]$$

L'errore entra nella previsione future

$$|\phi_1| < 1, |\theta_1| < 1$$

Caso AR(2)

$$Y_{T+1} = \phi_1 Y_T + \phi_2 Y_{T-1} + \varepsilon_{T+1}$$

$$Y_{T+1|T} = \phi_1 Y_T + \phi_2 Y_{T-1} + \theta_1$$

$$Y_{T+2|T} = \phi_1 Y_{T+1|T} + \phi_2 Y_T + \theta_1$$

$$Y_{T+j|T} = \phi_1 Y_{T+j-1|T} + \phi_2 Y_{T+j-2|T} + \theta_1$$

Con stazarietà converge alla media del processo

$\varphi_t \sim I(1)$ AR(1) stat.

$$\Delta Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

$$Y_t = Y_{t-1} + \phi_1 \cdot \Delta Y_{t-1} + \varepsilon_t$$

$$E[Y_{t+1} | \mathcal{F}_t] = Y_t + \hat{\phi}_1 \Delta Y_{t-1}$$

// Slides

Le serie temporali sollevano nuove problematiche tecniche

- Ritardi temporali
- Correlazione nel tempo (*correlazione seriale*, o *autocorrelazione* – già incontrata con i dati panel)
- Calcolo di errori standard quando gli errori sono serialmente correlati

Un buon modo per apprendere riguardo le serie temporali è quello di fare ricerche! Un'ottima fonte di serie temporali macro USA, e alcune internazionali, è il [FRED database](#) della Federal Reserve Bank of St. Louis's

2. Uso di modelli di regressione per la previsione (Paragrafo 14.1)

- Previsione e stima di effetti causali sono obiettivi piuttosto diversi.
- Per la previsione,
 - \bar{R}^2 conta (molto!)
 - La distorsione da variabili omesse non è un problema!
 - Non ci preoccuperemo di interpretare i coefficienti nei modelli di previsione – non serve stimare effetti causali se si vogliono soltanto fare previsioni!
 - La validità esterna è fondamentale: il modello stimato usando dati storici deve valere nel (prossimo) futuro

3. Introduzione alle serie temporali e alla correlazione seriale (Paragrafo 14.2)

Basi per le serie temporali:

- A. Notazione
- B. Ritardi, differenze prime, tassi di crescita
- C. Autocorrelazione (correlazione seriale)
- D. Stazionarietà

A. Notazione

- Y_t = valore di Y nel periodo t .
- Data set: $\{Y_1, \dots, Y_T\}$ sono T osservazioni sulla variabile serie temporale Y
- Consideriamo soltanto osservazioni consecutive, a intervalli uniformi (per esempio mensili, dal 1960 al 1999, senza saltare mesi; dati mancanti e intervalli non uniformi introducono complicazioni tecniche)

B. Ritardi, differenze prime e tassi di crescita

CONCETTO CHIAVE 14.1

Ritardi, differenze prime, logaritmi e tassi di crescita

- Il primo ritardo di una serie temporale Y_t è Y_{t-1} ; il suo j -esimo ritardo è Y_{t-j} .
- La differenza prima di una serie, ΔY_t , è la sua variazione tra il periodo $t - 1$ e il periodo t , cioè $\Delta Y_t = Y_t - Y_{t-1}$.
- La differenza prima del logaritmo di Y_t è $\Delta \ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1})$.
 $\Rightarrow = \ln\left(\frac{Y_t}{Y_{t-1}}\right)$
- La variazione percentuale di una serie temporale Y_t tra i periodi $t - 1$ e t è approssimativamente uguale a $100\Delta \ln(Y_t)$, dove l'approssimazione è più accurata quando la variazione percentuale è piccola.

Le differenze logaritmiche sono utili per analizzare tassi di crescita in quanto

$$\Delta \ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1}) = \ln\left(\frac{Y_t}{Y_{t-1}}\right)$$

\Rightarrow "Tasso di crescita" delle serie storiche

Esempio: tasso trimestrale di crescita del PIL USA annualizzato

PIL = PIL reale negli USA (miliardi di \$ del 1996)

- PIL nel primo trimestre del 2012 (2012:T1) = 15382
- PIL nel secondo trimestre del 2012 (2012:T2) = 15428
- Variazione percentuale nel PIL, dal 2012:T1 al 2012:T2:

$$= 100 \times \left(\frac{15428 - 15382}{15382} \right) = 0,299\%$$

- Variazione percentuale nel PIL, dal 2012:T1 al 2012:T2, *tasso annualizzato* = $4 \times 0,299\% = \mathbf{1,196\%} \approx \mathbf{1,2\%}$ (percentuale per anno)
- Usando l'approssimazione logaritmica alle variazioni percentuali si ottiene $4 \times 100 \times [\log(15428) - \log(15382)] = \mathbf{1,194\%}$

C. Autocorrelazione (correlazione seriale)

La correlazione di una serie con i suoi valori ritardati è detta **autocorrelazione** o **correlazione seriale**.

- La prima **autocovarianza** di Y_t è $\text{cov}(Y_t, Y_{t-1})$
 - La prima **autocorrelazione** di Y_t è $\text{corr}(Y_t, Y_{t-1}) = \frac{\gamma(k)}{\gamma(0)}$
 - Quindi
- $$\text{corr}(Y_t, Y_{t-1}) = \frac{\text{cov}(Y_t, Y_{t-1})}{\sqrt{\text{var}(Y_t) \text{var}(Y_{t-1})}}$$
- Queste sono correlazioni di popolazione, che descrivono la distribuzione congiunta di (Y_t, Y_{t-1})

D. stazionarietà

La stazionarietà indica che la storia è rilevante. Si tratta di un requisito chiave per la validità esterna della regressione di serie temporali.

CONCETTO CHIAVE 14.5

Stazionarietà

Una serie temporale Y_t è **stazionaria** se la sua distribuzione di probabilità non cambia nel corso del tempo, cioè se la distribuzione congiunta di $(Y_{s+1}, Y_{s+2}, \dots, Y_{s+T})$ non dipende da s indipendentemente dal valore di T ; altrimenti, la serie Y_t viene detta **non stazionaria**. Due serie temporali X_t e Y_t , sono dette **congiuntamente stazionarie** se la distribuzione congiunta di $(X_{s+1}, Y_{s+1}, X_{s+2}, Y_{s+2}, \dots, X_{s+T}, Y_{s+T})$ non dipende da s indipendentemente dal valore di T . La stazionarietà impone che il futuro sia come il passato, almeno in senso probabilistico.

Per ora assumiamo che Y_t sia stazionaria (ci torneremo più avanti).

4. Autoregressioni (Paragrafo 14.3)

- Un punto di partenza naturale per un modello di previsione è quello di usare valori passati di Y (cioè Y_{t-1}, Y_{t-2}, \dots) per la previsione di Y_t .
- Un'**autoregressione** è un modello di regressione in cui si esegue la regressione di Y_t rispetto ai suoi valori passati.
- Il numero di ritardi usati come regressori è detto **ordine** dell'autoregressione.
 - In una **autoregressione del primo ordine**, si esegue la regressione di Y_t rispetto a Y_{t-1}
 - In una **autoregressione del p -esimo ordine**, si esegue la regressione di Y_t rispetto a $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$.

Il modello autoregressivo del primo ordine (AR(1))

Il modello di popolazione AR(1) è

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

- β_0 e β_1 non hanno interpretazioni causali
- se $\beta_1 = 0$, Y_{t-1} non è utile per prevedere Y_t
- Il modello AR(1) può essere stimato da una regressione di Y_t rispetto a Y_{t-1} (come la eseguireste, in pratica??)
- La verifica di $\beta_1 = 0$ v. $\beta_1 \neq 0$ fornisce un test dell'ipotesi che Y_{t-1} non sia utile per prevedere Y_t

Previsioni: terminologia e notazione

- I valori predetti sono “dentro il campione” (definizione consueta)
OOB: Out Of the Bag
- Le previsioni sono “fuori campione” – nel futuro
- Notazione:
 - $Y_{T+1|T}$ = previsione di Y_{T+1} basata su Y_T, Y_{T-1}, \dots , usando i coefficienti di popolazione (ignoti)
 - $\hat{Y}_{T+1|T}$ = previsione di Y_{T+1} basata su Y_T, Y_{T-1}, \dots , usando i coefficienti stimati, che sono stimati con i dati al periodo T .
 - Per un AR(1):
 - $Y_{T+1|T} = \beta_0 + \beta_1 Y_T$
 - $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$, dove $\hat{\beta}_0$ e $\hat{\beta}_1$ sono stimati con dati al periodo T .

Errori di previsione

L'errore di previsione futura a un periodo è

$$\text{errore previsione} = Y_{T+1} - \hat{Y}_{T+1|T}$$

La distinzione tra errore di previsione e residuo è la stessa che esiste tra previsione e valore predetto:

- un *residuo* è “dentro il campione”
- un *errore di previsione* è “fuori campione” – il valore di Y_{T+1} non è usato nella stima dei coefficienti di regressione

Il modello AR(p): uso di ritardi multipli per la previsione

Il modello autoregressivo del p -esimo ordine (AR(p)) è

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + u_t$$

- Il modello AR(p) usa p ritardi di Y come regressori
- Il modello AR(1) è un caso particolare
- I coefficienti non hanno un'interpretazione causale
- Per verificare l'ipotesi che Y_{t-2}, \dots, Y_{t-p} non siano utili a prevedere Y_t , oltre a Y_{t-1} , si usa un test F
- Si usano test t o F per determinare p
- Oppure, meglio, si determina p usando un “criterio di informazione” (*ne parleremo più avanti...*)

5. Regressioni temporali con predittori aggiuntivi e modello autoregressivo misto (Paragrafo 14.4)

- Finora abbiamo considerato modelli di previsione che usano solo valori passati di Y
- Ha senso aggiungere altre variabili (X) che potrebbero essere predittori utili di Y , oltre ai valori predittivi dei valori ritardati di Y :

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_p Y_{t-p} + \delta_1 X_{t-1} + \dots + \delta_r X_{t-r} + u_t$$

- Questo è un **modello autoregressivo misto** con p ritardi di Y e r ritardi di X ... **ADL(p,r)**.

Il test dell'ipotesi congiunta che nessuna delle X sia un predittore utile, oltre ai valori passati di Y, si chiama *test di causalità di Granger*

CONCETTO CHIAVE 14.7

Test di causalità di Granger

La statistica per il test di causalità di Granger è la statistica F per la verifica dell'ipotesi nulla che i coefficienti su tutti i valori di una delle variabili dell'equazione (14.20) (per esempio, i coefficienti di $X_{1t-1}, X_{1t-2}, \dots, X_{1t-q_1}$) siano pari a zero. Questa ipotesi nulla implica che i regressori non abbiano ulteriore potere predittivo per Y_t rispetto a quello già posseduto dagli altri regressori; il test di questa ipotesi nulla viene detto test di causalità di Granger.

"Causalità" è un termine sfortunato in questo caso: la causalità di Granger si riferisce semplicemente al contenuto predittivo (marginale).

6. Incertezza e intervalli di previsione

Perché serve una misura dell'incertezza di previsione?

- Per costruire intervalli di previsione
- Per consentire agli utenti della previsione (inclusi voi) di sapere quale grado di precisione attendersi

Si consideri la previsione

$$\hat{Y}_{T+1|T} = \hat{\beta}_0^+ \hat{\beta}_1^- + \hat{\beta}_1^+$$

L'errore di previsione è:

$$Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0^- \beta_0) + (\hat{\beta}_1^- \beta_1) Y_T + (-\hat{\beta}_1^+) X_T]$$

$\hat{\beta}_2^- - \beta_2$

L'errore di previsione quadratico medio (MSFE) è

$$\begin{aligned} E(Y_{T+1} - \hat{Y}_{T+1|T})^2 &= E(u_{T+1})^2 + \\ &+ E[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\beta}_2 - \beta_2)X_T]^2 \end{aligned}$$

- MSFE = $\text{var}(u_{T+1})$ + incertezza dovuta a errore di stima
- Se la dimensione del campione è grande, la parte dovuta all'errore di stima è (molto) più piccola di $\text{var}(u_{T+1})$, nel qual caso

$$\text{MSFE} \approx \text{var}(u_{T+1})$$

- La **radice quadrata dell'errore di previsione quadratico medio (RMSFE)**:

Root Mean Square Forward Error

$$\text{RMSFE} = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}$$

La radice quadrata dell'errore di previsione quadratico medio (RMSFE)

$$\text{RMSFE} = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}$$

- L'RMSFE è una misura della dispersione della distribuzione dell'errore di previsione.
- L'RMSFE è simile alla deviazione standard di u_t , ma si focalizza esplicitamente sull'errore di previsione usando coefficienti stimati, non usando la retta di regressione della popolazione.
- L'RMSFE è una misura dell'ampiezza di un tipico "sbaglio" di previsione

Tre modi per stimare l'RMSFE

1. Usare l'approssimazione $RMSFE \approx \sigma_u$, così da stimare l'RMSFE mediante il SER.
2. Usare un'effettiva cronologia di previsione per $t = t_1, \dots, T$, quindi stimare mediante

$$MSFE = \frac{1}{T - t_1 + 1} \sum_{t=t_1-1}^{T-1} (Y_{t+1} - \hat{Y}_{t+1|t})^2$$

Solitamente non è pratico – richiede la disponibilità di una registrazione storica di previsioni effettive dal modello

3. Usare una cronologia di previsione simulata, cioè che simuli le previsioni che si sarebbero fatte usando il modello in tempo reale... quindi usare il metodo 2, con queste **pseudo previsioni fuori campione**...

Il metodo delle *pseudo previsioni fuori campione*

- Ri-stimare il modello ogni periodo, $t = t_1-1, \dots, T-1$
- Calcolare la “previsione” per la data $t+1$ usando la stima tramite t
- Calcolare la pseudo previsione fuori campione alla data t , usando la stima tramite $t-1$. Questa è $\hat{Y}_{t+1|t}$.
- Calcolare l’errore della pseudo previsione fuori campione, $Y_{t+1} - \hat{Y}_{t+1|t}$
- Inserire questo errore di previsione nella formula MSFE,

$$MSFE = \frac{1}{T-t_1+1} \sum_{t=t_1-1}^{T-1} (Y_{t+1} - \hat{Y}_{t+1|t})^2$$

Perché si usa il termine “pseudo previsioni fuori campione”?

Usare l'RMSFE per costruire intervalli di previsione

Se u_{T+1} ha distribuzione normale, allora un intervallo di previsione al 95% può essere costruito come

$$\hat{Y}_{T|T-1} \pm 1,96 \times RMSFE$$

Note:

- Un intervallo di previsione al 95% non è un intervallo di confidenza (Y_{T+1} non è un coefficiente non casuale, è casuale!)
- Questo intervallo è valido solo se u_{T+1} è normale – ma potrebbe comunque fornire un'approssimazione ragionevole ed è una misura comunemente usata di incertezza della previsione
- Spesso si usano intervalli di previsione “67%” : \pm

RMSFE

7. Scelta della lunghezza dei ritardi usando criteri d'informazione (Paragrafo 14.5)

Come scegliere il numero di intervalli p in un AR(p)?

- La distorsione da variabili omesse è irrilevante per la previsione!
- Si possono usare sequenze di test t o F ; ma i modelli scelti tendono a essere “troppo grandi” (perché?)
xché non hanno elementi di penalizzazione
- Un altro modo – migliore – per determinare la lunghezza dei ritardi è quello di usare un *criterio di informazione*
- I criteri di informazione bilanciano distorsione (troppo pochi ritardi) e varianza (troppi ritardi)
- Due *criteri informativi* sono quello di Bayes (BIC) e quello di Akaike (AIC)...

Il Bayes Information Criterion (BIC)

$$\text{BIC}(p) = \ln\left(\frac{\text{SSR}(p)}{T}\right) + (p+1)\frac{\ln T}{T}$$

- *Primo termine*: sempre decrescente in p (più grande è p , migliore è l'adattamento)
- *Secondo termine*: sempre crescente in p .
 - La varianza della previsione dovuta all'errore di stima aumenta con p – perciò non si vuole un modello di previsione con troppi coefficienti – ma quando sono "troppi"?
 - Questo termine è una "penalità" per l'uso di più parametri – che aumenta la varianza della previsione.
- *Minimizzando il BIC(p)* si bilanciano distorsione e varianza per determinare un valore "migliore" di p per la previsione.

– Il risultato è che

$$\hat{p}^{BIC} \xrightarrow{p} ! \quad (\text{Appendice 14.5})$$

Un altro criterio di informazione: *Akaike Information Criterion (AIC)*

$$AIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$$

$$BIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$$

Il termine di penalità è più piccolo per l'AIC rispetto al BIC ($2 < \ln T$)

- AIC stima più intervalli (p più grande) del BIC
- Questo potrebbe essere utile se si pensa che sia importante avere una maggiore lunghezza degli intervalli.
- Tuttavia, lo stimatore AIC di p non è consistente – può sovrastimare p – la penalità non è abbastanza grande

8. Non stazionarietà I: i trend (Paragrafo 14.6)

Finora abbiamo assunto che i dati fossero stazionari, cioè che la distribuzione di $(Y_{s+1}, \dots, Y_{s+\tau})$ non dipendesse da s .

Se non c'è stazionarietà, le serie si dicono ***non stazionarie***.

Due importanti tipi di non stazionarietà:

- Trend (Paragrafo 14.6)
- Rotture strutturali (instabilità del modello) (Paragrafo 14.7)

B. Trend deterministici e stocastici

Un trend è un movimento o tendenza di lungo termine nei dati.

- Un **trend deterministico** è una funzione non casuale del tempo (per es. $y_t = t$, o $y_t = t^2$).
- Un **trend stocastico** è casuale e varia nel tempo
- Un importante esempio di trend stocastico è una **passeggiata aleatoria**:

$$Y_t = Y_{t-1} + u_t, \text{ dove } u_t \text{ è serialmente incorrelato}$$

Se Y_t segue una passeggiata aleatoria, allora il valore di Y domani è il valore di Y oggi più un disturbo impredicibile.

Trend deterministici e stocastici (continua)

Due caratteristiche chiave di una passeggiata aleatoria:

(i) $Y_{T+h|T} = Y_T$

- La miglior previsione del valore di Y nel futuro è il valore di Y oggi
- In una prima approssimazione, i logaritmi dei prezzi azionari seguono una passeggiata aleatoria (più precisamente, i rendimenti azionari sono impredicibili)

(ii) Supponiamo $Y_0 = 0$. Allora $\text{var}(Y_t) = \sigma^2$.

- Questa varianza dipende da t (aumenta linearmente con t), perciò Y_t non è stazionaria (si ricordi la definizione di stazionarietà).

Trend deterministici e stocastici (continua)

Una passeggiata aleatoria con deriva è

$$Y_t = \beta_0 + Y_{t-1} + u_t, \text{ dove } u_t \text{ è serialmente incorrelato}$$

La "deriva" è β_0 : se $\beta_0 \neq 0$, allora Y_t segue una passeggiata aleatoria attorno a un trend lineare. Potete vederlo considerando la previsione con passo h :

$$Y_{T+h|T} = \beta_0 h + Y_T$$

Il modello della passeggiata aleatoria (con o senza deriva) offre una buona descrizione di trend stocastici in molte serie temporali economiche.

Trend deterministici e stocastici (continua)

Ecco un consiglio pratico:

Se Y_t ha un trend di passeggiata aleatoria, allora ΔY_t è stazionaria e l'analisi di regressione dovrebbe essere svolta usando ΔY_t anziché Y_t .

Fattori che portano a questo consiglio:

- Relazione tra modello della passeggiata aleatoria e AR(1), AR(2), AR(p) ("radice autoregressiva unitaria")
- Il test di Dickey-Fuller per verificare se Y_t ha un trend di passeggiata aleatoria

Trend stocastici e radici autoregressive unitarie

Passeggiata aleatoria (con deriva): $Y_t = \beta_0 + Y_{t-1} + u_t$

AR(1):

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

- La passeggiata aleatoria è un AR(1) con $\beta_1 = 1$.
- Il caso speciale di $\beta_1 = 1$ è detto radice unitaria*.
- Quando $\beta_1 = 1$, il modello AR(1) diventa

$$\Delta Y_t = \beta_0 + u_t$$

*Questa terminologia deriva dal considerare l'equazione

$1 - \beta_1 z = 0$ – la “radice” di questa equazione è $z = 1/\beta_1$, che è uguale a uno (unità) se $\beta_1 = 1$.

Radici unitarie in un AR(2)

AR(2):
$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t$$

Usiamo il trucco di “riscrivere la regressione” dal Paragrafo 7.3:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) Y_{t-1} - \beta_2 Y_{t-1} + \beta_2 Y_{t-2} + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) Y_{t-1} - \beta_2 (Y_{t-1} - Y_{t-2}) + u_t \end{aligned}$$

Sottraiamo Y_{t-1} da entrambi i membri:

$$Y_t - Y_{t-1} = \beta_0 + (\beta_1 + \beta_2 - 1) Y_{t-1} - \beta_2 (Y_{t-1} - Y_{t-2}) + u_t$$

ovvero

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + u_t,$$

dove $\delta = \beta_1 + \beta_2 - 1$ e $\gamma_1 = -\beta_2$.

Radici unitarie in un AR(2) (continua)

Allora il modello AR(2) può essere riscritto come

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + u_t$$

dove $\delta = \beta_1 + \beta_2 - 1$ e $\gamma_1 = -\beta_2$.

$$\frac{-b \pm \sqrt{\Delta}}{2a}$$

Nota: se $1 - \beta_1 z - \beta_2 z^2 = 0$ ha una radice unitaria, allora $\beta_1 + \beta_2 = 1$ (*verificate lo da soli, trovate le radici!*)

Se c'è una radice unitaria, allora $\delta = 0$ e il modello AR(2) diventa

$$\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + u_t$$

Se un modello AR(2) ha una radice unitaria, allora può essere scritto come un AR(1) in differenze prime.

Radici unitarie nel modello AR(p)

$$\text{AR}(p): \quad Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + u_t$$

Questa regressione può essere riscritta come

$$\Delta Y_t = \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

dove

$$\delta = \beta_1 + \beta_2 + \dots + \beta_p - 1$$

$$\gamma_1 = -(\beta_2 + \dots + \beta_p)$$

$$\gamma_2 = -(\beta_3 + \dots + \beta_p)$$

...

$$\gamma_{p-1} = -\beta_p$$

Radici unitarie nel modello AR(p) (continua)

Il modello AR(p) può essere scritto come

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

dove $\delta = \beta_1 + \beta_2 + \dots + \beta_p - 1$.

Nota: se c'è una radice unitaria nel modello AR(p),
allora $\delta = 0$ e il modello AR(p) diventa un modello
AR($p-1$) in differenze prime:

$$\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

C. Quali problemi sono causati dai trend?

1. I coefficienti AR sono fortemente distorti verso zero.
Questo porta a previsioni scadenti.
2. Alcune statistiche t non hanno una distribuzione normale standard, anche in grandi campioni (ci torneremo più avanti).
3. Se Y e X hanno entrambe trend di passeggiata aleatoria, allora possono apparire correlate anche se non lo sono – si possono ottenere “regressioni spurie”. Di seguito un esempio...

D. Come si individuano trend stocastici?

1. Si traccia il grafico dei dati – ci sono movimenti persistenti di lungo termine?
2. Si usa un test di regressione per una passeggiata aleatoria: il test di Dickey-Fuller per una radice unitaria.

Il test di Dickey-Fuller in un AR(1)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

ovvero

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$$

$$H_0: \delta = 0 \text{ (cioè } \beta_1 = 1) \text{ v. } H_1: \delta < 0$$

(nota: è monolaterale: $\delta < 0$ significa che Y_t è stazionaria)

Test DF in AR(1) (continua)

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$$
$$H_0: \delta = 0 \text{ (cioè } \beta_1 = 1\text{)} \vee H_1: \delta < 0$$

Test DF: calcolare la statistica t che verifica $\delta = 0$

- Sotto H_0 , questa statistica t **non** ha una distribuzione normale! (la nostra teoria della distribuzione si applica a variabili stazionarie, e Y_t non è stazionaria!)
- Occorre usare la tabella dei valori critici di Dickey-Fuller. Ci sono due casi, che hanno valori critici diversi:
 - (a) $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$ (sola intercetta)
 - (b) $\Delta Y_t = \beta_0 + \mu t + \delta Y_{t-1} + u_t$ (intervetta e trend temporale)

Tabella di valori critici DF

$$(a) \Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$$

(sola intercetta)

$$(b) \Delta Y_t = \beta_0 + \mu t + \delta Y_{t-1} + u_t$$

(intercetta e trend temporale)

Tabella 14.4 Valori critici per grandi campioni della statistica di Dickey-Fuller aumentata.

Regressore deterministico	10%	5%	1%
Intercetta	-2,57	-2,86	-3,43
Intercetta e trend temporale	-3,12	-3,41	-3,96

Respingere se la statistica t DF (la statistica t che verifica $\delta = 0$) è minore del valore critico specificato. Questo è un test monolaterale dell'ipotesi nulla di una radice unitaria (passegiata aleatoria) rispetto all'alternativa che l'autoregressione sia stazionaria.

Il test di Dickey-Fuller in un AR(p)

In un AR(p), il test DF si basa sul modello riscritto,

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t \quad (*)$$

dove $\delta = \beta_1 + \beta_2 + \dots + \beta_p - 1$. Se c'è una radice unitaria (passeggiata aleatoria), $\delta = 0$; se l'AR è stazionario, $\delta < 1$.

Il test DF in un AR(p) (sola intercetta):

1. Si stima (*), ottenendo la statistica t che verifica $\delta = 0$
2. Si respinge l'ipotesi nulla di una radice unitaria se la statistica t è minore del valore critico DF nella Tabella 14.5

Modifica per trend temporali: includere t come regressore in (*)

Quando si dovrebbe includere un trend temporale nel test DF?

La decisione di usare il test DF con sola intercetta o ~~e~~^e quollo con intercetta e trend dipende da quale sia l'alternativa e da come appaiano i dati.

- Nella specifica con sola intercetta, l'alternativa è che Y è stazionaria attorno a una costante – no crescita di lungo termine nella serie
- Nella specifica con intercetta e trend, l'alternativa è che Y è stazionaria attorno a un trend temporale lineare – la serie presenta crescita di lungo termine.

E. Come risolvere/mitigare i problemi generati dai trend

Se Y_t ha una radice unitaria (ha un trend stocastico di passeggiata aleatoria), il modo più facile per evitare i problemi posti da ciò è quello di modellare Y_t in differenze prime.

- Nel caso di AR, questo significa specificare l'AR usando differenze prime di Y_t (ΔY_t)

Riepilogo: come rilevare e affrontare trend stocastici

1. Il modello della passeggiata aleatoria è quello più utilizzato per trend in dati temporali economici
2. Per determinare se Y_t ha un trend stocastico, prima si traccia il grafico di Y_t . Se appare plausibile un trend, si calcola il test DF (decidere quale versione, con sola intercetta o con intercetta + trend)
3. Se il test DF non rifiuta, si conclude che Y_t ha una radice unitaria (trend stocastico di passeggiata aleatoria)
4. Se Y_t ha una radice unitaria, si usa ΔY_t per analisi di regressione e previsione. Se non c'è radice unitaria, si usa Y_t .

9. Non stazionarietà II: rotture (Paragrafo 14.7)

Il secondo tipo di non stazionarietà che consideriamo è che i coefficienti del modello potrebbero non essere costanti sull'intero campione. Chiaramente è un problema per la previsione se il modello che descrive i dati storici è diverso dal modello attuale – per le previsioni si vuole il modello attuale (questo è un problema di validità esterna)

perciò:

- vedremo due modi per rilevare variazioni nei coefficienti: test per rottura e pseudo previsioni fuori campione
- esamineremo un esempio: la previsione del PIL mediante il differenziale dei tassi di interesse a lungo e a breve termine

A. Test per una rottura nei coefficienti di regressione

Caso I: la data della rottura è nota

Si supponga che sia noto che la rottura si è verificata nella data τ . La stabilità dei coefficienti può essere verificata stimando un modello di regressione a integrazione totale. Nel caso dell'ADL(1,1):

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1} + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times X_{t-1}] + u_t$$

Dummy

β_0 *β_1* *δ_1* *γ_0* *γ_1* *γ_2*

dove $D_t(\tau) = 1$ se $t \geq \tau$, e = 0 altrimenti.

Se $\gamma_0 = \gamma_1 = \gamma_2 = 0$, allora i coefficienti sono costanti sull'intero campione.

Se almeno uno dei γ_0 , γ_1 o γ_2 è diverso da zero, la funzione di regressione cambia nella data τ .

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_1 X_{t-1}$$

$$+ \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) \times Y_{t-1}] + \gamma_2 [D_t(\tau) \times X_{t-1}] + u_t$$

dove $D_t(\tau) = 1$ se $t \geq \tau$, e = 0 altrimenti

La **statistica test di Chow** per una rottura nella data τ è la statistica F (robusta all'eteroschedasticità) che verifica:

$$H_0: \gamma_0 = \gamma_1 = \gamma_2 = 0$$

vs. H_1 : almeno uno dei γ_0 , γ_1 , o γ_2 non è zero

- Si noti che è applicabile a un sottoinsieme dei coefficienti, per esempio, soltanto quello di X_{t-1} .
- Sfortunatamente, spesso non si conosce una potenziale data della rottura, cioè non si conosce τ ...

Caso II: la data della rottura è ignota

Perché considerare questo caso?

- Si potrebbe sospettare che vi sia una rottura, senza però sapere quando
- Si potrebbe voler verificare l'ipotesi nulla della stabilità dei coefficienti contro l'alternativa generale che vi sia stata una rottura.
- Anche se si conosce la data della rottura, se tale “conoscenza” è basata sulla precedente osservazione della serie, allora si è in effetti “stimata” la data, e questo invalida i valori critici del test di Chow (*perché?*)

La statistica QLR (Quandt Likelihood Ratio)

(rapporto delle verosimiglianze di Quandt)

Statistica QLR = massima statistica di Chow

- Sia $F(\tau)$ = la statistica di Chow che verifica l'ipotesi che non vi siano rotture nella data τ .
- La statistica test QLR è il **massimo** di tutte le statistiche F di Chow su un intervallo τ , $\tau_0 \leq \tau \leq \tau_1$:

$$QLR = \max[F(\tau_0), F(\tau_0+1), \dots, F(\tau_1-1), F(\tau_1)]$$



- Una scelta convenzionale per τ_0 e τ_1 è costituita dal 70% più interno del campione (scludendo il primo e l'ultimo 15%).
- Si devono usare i consueti valori critici $F_{q,\infty}$?

Il test QLR (continua)

$$QLR = \max[F(\tau_0), F(\tau_0+1), \dots, F(\tau_1-1), F(\tau_1)]$$

- La distribuzione nulla in grandi campioni di $F(\tau)$ per un dato (fisso, non stimato) τ è $F_{q,\infty}$
- Ma se si calcolano due test di Chow e si scegliere il più grande, il valore critico deve essere maggiore del valore critico per un singolo test di Chow.
- Se si calcolano molti test di Chow – per esempio tutte le date nel 70% centrale del campione – il valore critico deve essere ancora più grande!

- **Nota:** in grandi campioni, QLR ha la distribuzione

$$\max_{a \leq s \leq 1-a} \left(\frac{1}{q} \sum_{i=1}^q \frac{B_i(s)^2}{s(1-s)} \right)$$

Non entra nel dettaglio. Basta ricordare in generale

dove $\{B_i\}$, $i = 1, \dots, n$, sono “ponti browniani” indipendenti e tempo-continui su $0 \leq s \leq 1$ (un ponte browniano è un moto browniano deviato dalla sua emdia; un moto brownianno è una passeggiata aleatoria gaussiana [a distribuzione normale] in tempo continuo), e dove $a = 0,15$ (si esclude il primo e l’ultimo 15% del campione)

- I valori critici sono riportati nella Tabella 14.5...

Tabella 14.5 Valori critici della statistica QLR con troncamento del 15%.

Numero di restrizioni (q)	10%	5%	1%
1	7,12	8,68	12,16
2	5,00	5,86	7,78
3	4,09	4,71	6,02
4	3,59	4,09	5,12
5	3,26	3,66	4,53
6	3,02	3,37	4,12
7	2,84	3,15	3,82
8	2,69	2,98	3,57
9	2,58	2,84	3,38
10	2,48	2,71	3,23
11	2,40	2,62	3,09
12	2,33	2,54	2,97
13	2,27	2,46	2,87
14	2,21	2,40	2,78
15	2,16	2,34	2,71
16	2,12	2,29	2,64
17	2,08	2,25	2,58
18	2,05	2,20	2,53
19	2,01	2,17	2,48
20	1,99	2,13	2,43

Nota: questi valori critici si applicano quando $\tau_0 = 0,15T$ e $\tau_1 = 0,85T$ (arrotondato al primo decimale), in modo tale che la statistica F è calcolata per tutte le possibili date di rottura nel 70% centrale del campione. Il numero di restrizioni q è il numero di restrizioni verificate da ogni singola statistica F . I valori critici per altre percentuali di troncamento sono dati in Andrews (2003).

Si noti che questi valori critici sono maggiori dei valori critici $F_{q,\infty}$ – per esempio il valore critico al 5% $F_{1,\infty}$ è 3,84.

B. Valutazione della stabilità del modello usando pseudo previsioni fuori campione

- Il test QLR non lavora bene verso la fine del campione, ma solitamente questa è la parte meno interessante!
- Un modo per verificare se il modello sta lavorando alla fine del campione è quello di vedere se le pseudo previsioni fuori campione (*poos*) sono “in carreggiata” nelle osservazioni più recenti. Si tratta di un approccio diagnostico informale (non di un test formale) che fa da complemento al test formale mediante QLR.

Applicazione al modello PIL–differenziale dei tassi di interesse

- Abbiamo trovato una rottura in 1980:T4 – perciò per questa analisi consideriamo soltanto regressioni con partenza in 1981:T1 – ignorando i dati precedenti dal “vecchio” modello.
- Modello di regressione:

variabile dipendente: $GDPGR_t$

regressori: intercetta, $GDPGR_{t-1}$, $GDPGR_{t-2}$, $TSpread_{t-1}$ e $TSpread_{t-2}$

- Pseudo previsioni fuori campione:

- Calcolare regressione su $t = 1980:T1, \dots, P$
- Calcolare previsione $poos$, $GDPGR_{2012:T4}^{P+1|P}$, ed errore di previsione
- Ripetere per $P = 2003:T1, \dots, 2012:T4$

→ *Prevision Out Of Sample*

Previsioni POOS di GDPGR usando il modello ADL(2,2) con TSpread

$$H_0: E[\epsilon_t] = 0 \quad \text{statistica test}$$

$$\frac{\bar{\epsilon} - 0}{S.E.(\bar{\epsilon})} = \frac{-0,73}{0,39} = -1,87$$

$$Var(\bar{\epsilon}) = \frac{1}{40} \sigma^2_{\epsilon}$$

$$\bar{\epsilon} = \frac{1}{40} \sum \epsilon_t$$

$$Var(\bar{x}) =$$

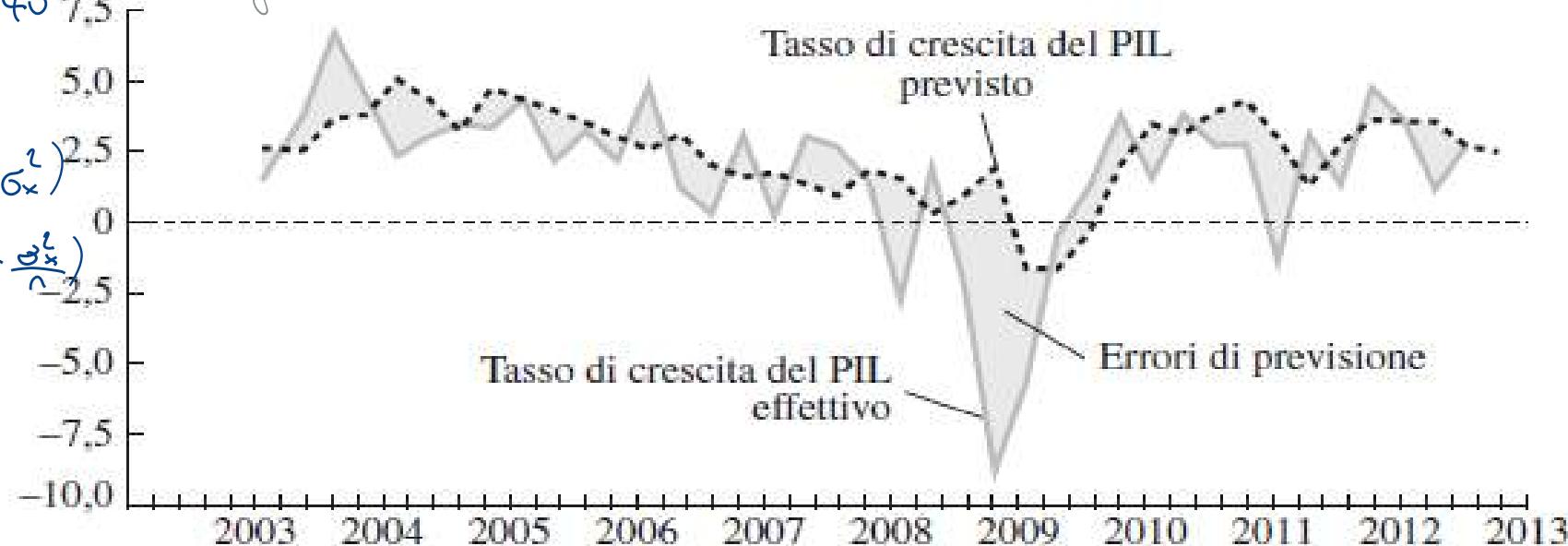
$$\bar{x} \xrightarrow{d} N(\mu; \sigma_x^2)$$

$$\bar{x} \approx N(\mu; \frac{\sigma_x^2}{n})$$

$$R \sqrt{SFE} = 2,54$$

Root Mean Square F... Error

D.R. fatto col 10% ma non al 5%



Il modello prevede il tasso di crescita del PIL dal 2003 al 2012, ma non riesce a prevedere il brusco calo conseguente alla crisi finanziaria del 2008.

10. Riepilogo: modelli di previsione per serie temporali (Paragrafo 14.8)

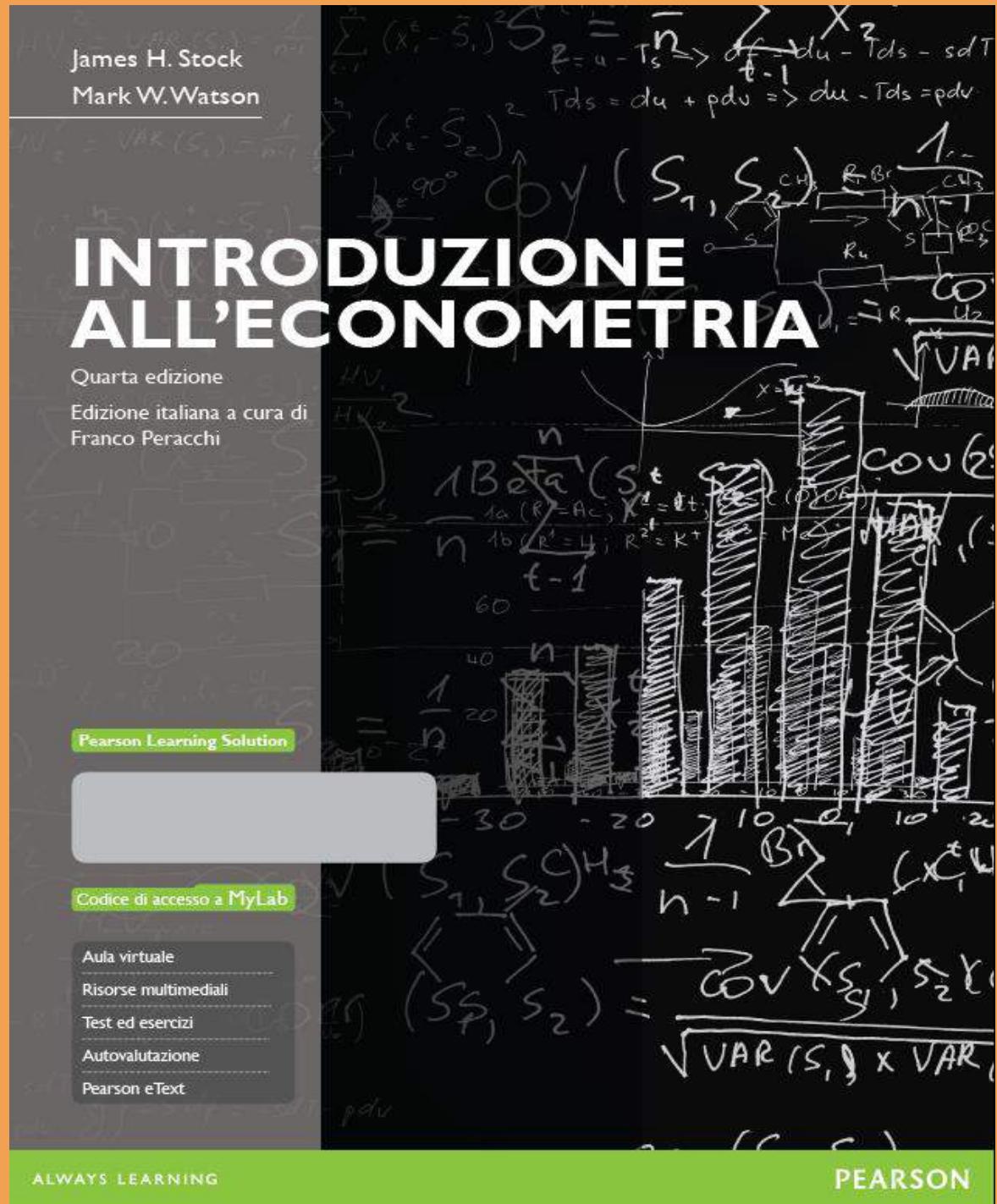
- Per scopi di previsione non è importante avere coefficienti con interpretazione causale!
- Gli strumenti di regressione possono essere usati per costruire modelli di previsione affidabili, anche se non vi è interpretazione causale dei coefficienti:
 - AR(p) – modelli “benchmark”
 - ADL(p,q) – aggiunge q ritardi di X (un altro predittore)
 - Test di causalità di Granger – verificano se una variabile X e i suoi ritardi sono utili per predire Y a partire dai ritardi di Y .

Riepilogo (continua)

- Nuovi concetti e strumenti:
 - stazionarietà
 - intervalli di previsione con RMSFE
 - pseudo previsioni fuori campione
 - BIC per scelta modello
 - Modi per verifiche di non stazionarietà:
 - test di Dickey-Fuller per una radice unitaria (trend stocastico)
 - Test per una violazione nei coefficienti di regressione:
 - test di Chow in una data nota
 - test QLR in una data ignota
 - analisi *poos* per previsioni verso la fine del campione

Capitolo 15

Stima degli effetti causali dinamici



Sommario

1. Gli effetti causali dinamici e i dati sul succo d'arancia
2. Stima degli effetti causali dinamici con regressori esogeni: il modello a ritardi distribuiti
3. Gli errori standard HAC *→ Consistenti rispetto l'eteroschedasticità e Autocorrelati*
4. Applicazione ai prezzi del succo d'arancia
5. Altro sull'esogeneità

Gli effetti causali dinamici e i dati sul succo d'arancia (Paragrafi 15.1 e 15.2)

Un effetto **causale dinamico** è l'effetto su Y di una variazione in X nel tempo.

Per esempio:

- L'effetto prodotto dall'aumento delle tasse sul tabacco sul consumo di sigarette per l'anno in corso, per il prossimo anno, per i prossimi 5 anni.
- L'effetto prodotto sull'inflazione da una modifica sul tasso dei Fed Funds per il mese in corso, per i prossimi 6 mesi, e per il prossimo anno.
- L'effetto prodotto da una gelata verificatasi in Florida sul prezzo del concentrato di succo d'arancia a 1 mese, a 2 mesi a 3 mesi...

Effetti causali dinamici (continua)

In applicazioni a serie temporali, non è possibile condurre l'esperimento casualizzato ideale:

- Esiste un solo mercato del succo d'arancia USA....
- Non è possibile assegnare in maniera casuale gli FDD a repliche diverse del mercato del succo d'arancia USA (cosa vuole dire, in effetti?)
- Non si può misurare il risultato medio (tra i "soggetti") in tempi diversi – esiste solo un unico "soggetto"!
- Quindi non si può stimare l'effetto causale per tempi diversi con lo stimatore delle differenze

Effetti causali dinamici (continua)

Un esperimento alternativo:

- Somministrare in maniera casuale diversi trattamenti *allo stesso soggetto (FDD_t) in tempi diversi*
- Misurare la varianza del risultato ($\%ChgP_t$)
- La “popolazione” dei soggetti è formata dal medesimo soggetto (mercato del succo) *ma in date diverse* – a volte il soggetto è il gruppo di trattamento, a volte il gruppo di controllo!
- Se i “soggetti” (il soggetto in tempi diversi) appartengono alla stessa distribuzione – cioè, se Y_t , X_t sono **stabili** – allora l’effetto causale dinamico può essere dedotto dalla regressione OLS di Y_t sui valori ritardati di X_t .
- Questo stimatore (regressione di Y_t su X_t e sui ritardi di X_t) è chiamato **stimatore a ritardi distribuiti**.

Gli effetti causali dinamici e il modello a ritardi distribuiti

Il modello a ritardi distribuiti è:

$$Y_t = \beta_0 + \beta_1 X_t + \dots + \beta_r X_{t-r} + u_t$$

- $\beta_1 = \text{effetto d'impatto della variazione in } X$ = effetto della variazione in X_t su Y_t , tenendo costante l' X_t precedente
- $\beta_2 = \text{moltiplicatore dinamico periodo 1}$ = effetto della variazione in X_{t-1} su Y_t , tenendo costante $X_t, X_{t-2}, X_{t-3}, \dots$
- $\beta_3 = \text{moltiplicatore dinamico periodo 2}$ (ecc.) = effetto della variazione in X_{t-2} su Y_t , tenendo costante $X_t, X_{t-1}, X_{t-3}, \dots$
- **Moltiplicatori dinamici cumulati**
 - Il moltiplicatore dinamico cumulato del secondo periodo è $\beta_1 + \beta_2 + \beta_3$ = effetto d'impatto + effetto periodo 1 + effetto periodo 2

L'esogeneità nella regressione a serie temporali

Esogeneità (*passato e presente*)

X è **esogena** se $E(u_t|X_t, X_{t-1}, X_{t-2}, \dots) = 0$.

Esogeneità stretta (*passato, presente, e futuro*)

X è **strettamente esogena** se $E(u_t|..., X_{t+1}, X_t, X_{t-1}, ...) = 0$

- L'esogeneità stretta implica l'esogeneità
- Per ora supporremo che X sia esogena – riprenderemo (in breve) il caso dell'esogeneità stretta più tardi.
- Se X è esogena, allora è possibile usare gli OLS per stimare l'effetto causale su Y di una variazione in X

Stima degli effetti causali dinamici con regressori esogeni (Paragrafo 15.3)

Modello a ritardi distribuiti:

$$Y_t = \beta_0 + \beta_1 X_t + \dots + \beta_{r+1} X_{t-r} + u_t$$

Assunzioni del modello a ritardi distribuiti

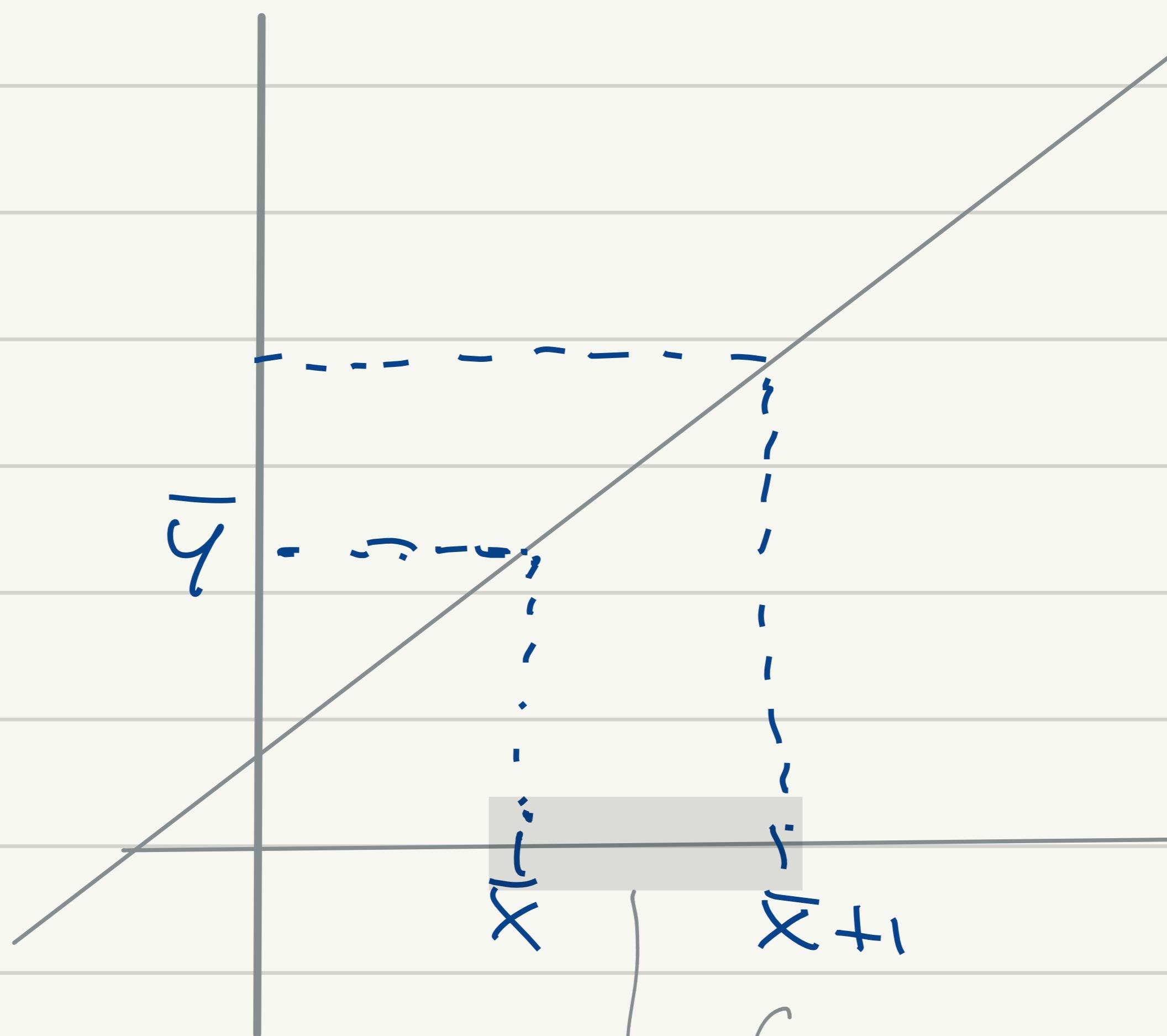
1. $E(u_t | X_t, X_{t-1}, X_{t-2}, \dots) = 0$ (X è esogena)
2. (a) Y e X hanno distribuzioni stabili;
(b) (Y_t, X_t) e (Y_{t-j}, X_{t-j}) diventano indipendenti al crescere di j
3. Y e X presentano otto momenti finiti non nulli
4. Non vi è collinearità perfetta.

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-2} + \nu_t$$

$$Y_t = \beta_0 + (\beta_1 + \beta_2) \bar{X}$$

$$\bar{X} = (\bar{x} + 1)$$

Variazione statica media di lungo periodo



→ Cosa succede se c'è una variazione permanente

Il modello a ritardi distribuiti (continua)

- Le assunzioni 1 e 4 sono familiari
 - L'assunzione 3 è familiare, tranne per 8 (non quattro) momenti finiti – ciò ha a che fare con gli stimatori HAC
 - L'assunzione 2 è diversa – prima poneva che (X_i, Y_i) erano i.i.d. – con i dati a serie temporali le cose si fanno più complesse.
2. (a) Y e X hanno distribuzioni stabili; \sim i.d.
- Se sì, i coefficienti non cambiano all'interno del campione (validità interna);
 - e i risultati possono essere estrapolati al di fuori del campione (validità esterna).
 - Questa è la controparte a serie temporali della parte “a distribuzione identica” di i.i.d.

Il modello a ritardi distribuiti, continua

2. (b) (Y_t, X_t) e (Y_{t-j}, X_{t-j}) diventano indipendenti al crescere di $j \sim$ i.i.d. → *Pero non subito*
- Intuitivamente, significa che si hanno esperimenti separati per periodi di tempo molto distanti fra loro.
 - Nei dati sezionali, avevamo supposto che Y e X fossero i.i.d., conseguenza di una semplice campionatura casuale – ciò portava al teorema limite centrale.
 - Una versione del TLC vale per le variabili a serie temporali che diventano indipendenti al crescere della loro separazione temporale – L'assunzione 2(b) è la controparte a serie temporali della parte “a distribuzione indipendente” di i.i.d.

In base ai presupposti del modello a ritardi distribuiti:

- OLS produce stimatori **consistenti** di $\beta_1, \beta_2, \dots, \beta_r$ (dei moltiplicatori dinamici)
- In campioni grandi, la distribuzione campionaria di $(\hat{\beta}_1, \dots)$ è normale
- **MA** la formula per la varianza di questa distribuzione campionaria non è la solita dei dati sezionali (i.i.d.), perché u_t non è i.i.d. – u_t può essere serialmente correlato!
- Ciò significa che i normali errori standard di OLS (in genere le stampe di STATA) sono sbagliati!
- Occorre utilizzare, invece, errori standard che siano robusti sia all'autocorrelazione sia all'eteroschedasticità...

↳ HAC

Errori standard consistenti in presenza di eteroschedasticità e autocorrelazione (HAC) (Paragrafo 15.4)

Il calcolo... per un singolo regressore X_t :

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

Lo stimatore OLS: dall'Appendice 4.3,

$$\hat{\beta}_1 \beta_1 = \frac{\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}) u_t}{\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2}$$
$$\approx \frac{\frac{1}{T} \sum_{t=1}^T v_t u_t}{\sigma_X^2} \quad (\text{In grandi campioni})$$

dove $v_t = (X_t - \bar{X}) u_t$.

Errori standard HAC (continua)

Per cui, in grandi campioni,

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= \text{var}\left(\frac{1}{T} \sum_{t=1}^T v_t\right) / (\sigma_x^2)^2 \\ &= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \text{cov}(v_t, v_s) / (\sigma_x^2)^2 \end{aligned}$$

? $\Sigma_v = \text{matrice var-cov}$
? $\text{Var}(\hat{\beta}_1) = (\sigma_x^2)^{-2} \Sigma_v^{-1} \Sigma_v$

Nei dati i.i.d. sezionali, $\text{cov}(v_t, v_s) = 0$ per $t \neq s$, quindi

$$\text{var}(\hat{\beta}_1) = \frac{1}{T^2} \sum_{t=1}^T \text{var}(v_t) / (\sigma_x^2)^2 = \frac{\sigma_v^2}{T(\sigma_x^2)^2}$$

? Σ_v è diagonale

Questo è il nostro solito risultato per dati sezionali
(Appendice 4.3).

Errori standard HAC (continua)

Ma in dati a serie temporali, $\text{cov}(v_t, v_s) \neq 0$ in genere.

Si ponga $T = 2$:

$$\begin{aligned}\text{var}\left(\frac{1}{T} \sum_{t=1}^T v_t\right) &= \text{var}[1/2(v_1+v_2)] \\ &= 1/4[\text{var}(v_1) + \text{var}(v_2) + 2\text{cov}(v_1, v_2)] \\ &= 1/2 \sigma_v^2 [1/2\rho_1 + \sigma_v^2] \quad (\rho_1 = \text{corr}(v_1, v_2)) \\ &= 1/2 \times f_2, \text{ dove } f_2 = (1+\rho_1)\end{aligned}$$

- In dati i.i.d., $\rho_1 = 0$ quindi $f_2 = 1$, dando la consueta formula
- In dati a serie temporali, se $\rho_1 \neq 0$ allora $\text{var}(\hat{\beta}_1)$ viene data dalla formula consueta.

Espressione per $\text{var}(\hat{\beta}_1)$ *T generico*

$$\text{var}\left(\frac{1}{T} \sum_{t=1}^T v_t\right) = \frac{\sigma_v^2}{T} \times f_T$$

Quindi

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{T} \frac{\sigma_v^2}{(\sigma_x^2)^2} \right] f_T$$

dove

$$f_T = 1 + 2 \sum_{j=1}^{T-1} \left(\frac{T-j}{T} \right) [Eq. (15.13)]$$

- Gli errori standard OLS convenzionali sono sbagliati quando u_t è correlato serialmente (la stampa „r“ di STATA è sbagliata).
- Gli errori standard OLS si discostano in base al fattore f_T
- Deve essere utilizzata una formula di errori standard diversa!!!

Errori standard HAC

- Avendo conosciuto il fattore f_T , sarebbe stato possibile apportare le modifiche necessarie.
 - Nei dati panel, il fattore f_T viene (implicitamente) stimato usando "cluster" – ma questo richiede n grande.
 - Nei dati a serie temporali, serve una formula diversa – f_T deve essere stimato in maniera esplicita

Gli errori standard che usano stimatori di f_T consistenti sono chiamati errori standard consistenti in presenza di eteroschedasticità e autocorrelazione o errori standard HAC (**H**eteroskedasticity- and **A**utocorrelation-
Consistent - **HAC**)

Errori standard HAC (continua)

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{T} \frac{\sigma_v^2}{(\sigma_x^2)^2} \right] \times f_T, \text{ dove } f_T = 1 + 2 \sum_{j=1}^{T-1} \left(\frac{T-j}{T} \right) \rho_j$$

$\sigma_v^2 = \sum_{t=1}^T (\hat{y}_t - \bar{y}) u_t^2$

Lo stimatore di f_T comunemente più utilizzato è:

$$\hat{f}_T = 1 + 2 \sum_{j=1}^{m-1} \left(\frac{m-j}{m} \right) \hat{\rho}_j \quad (\text{Newey-West})$$

- $\hat{\rho}_j$ è uno stimatore di ρ_j
- Questo è lo stimatore “Newey-West”
- m è detto **parametro di troncamento**
- Come scegliere m ?
 - Con il metodo Goldilocks (non troppi, non troppo pochi)
 - O con la regola empirica, $m = 0,75T^{1/3}$

FAQ: è necessario usare errori standard HAC per la stima di un modello AR o ADL?

R: NO.

- Il problema che ha una soluzione negli errori standard HAC si pone solo quando u_t è serialmente correlato: se u_t è serialmente incorrelato, vanno bene gli errori standard OLS
- Nei modelli AR e ADL, gli errori sono serialmente incorrelati se sono stati introdotti sufficienti ritardi di Y
 - Se si inseriscono sufficienti ritardi di Y , allora il termine di errore non può essere previsto usando Y passati, o in maniera equivalente, con u trascorsi – quindi u è serialmente incorrelato

Stima degli effetti causali dinamici con regressori strettamente esogeni (Paragrafo 15.5)

- X è strettamente esogena se $E(u_t | \dots, X_{t+1}, X_t, X_{t-1}, \dots) = 0$
- Se X è strettamente esogena, vi sono modi più efficienti per stimare gli effetti causali dinamici che non una regressione a ritardi distribuiti:
 - Stima dei minimi quadrati generalizzati (GLS)
 - Stima autoregressiva a ritardi distribuiti (ADL)
- Ma la condizione di stretta esogeneità è molto forte, per cui questa condizione nella pratica diventa raramente plausibile – neppure nell'esempio meteo/succo d'arancia (perché?).
- Per cui non tratteremo la stima GLS o ADL degli effetti causali dinamici – per dettagli si rimanda al Paragrafo 15.5.

Una parentesi: calcolo dei moltiplicatori cumulati e dei loro errori standard

I moltiplicatori cumulati possono essere calcolati stimando il modello a ritardi distribuiti, quindi sommando i coefficienti. Tuttavia, si dovrebbero anche calcolare gli errori standard per i moltiplicatori cumulati, e mentre ciò può essere fatto direttamente dal modello a ritardi distribuiti, sono necessarie alcune modifiche.

Siccome i moltiplicatori cumulati sono combinazioni lineari di coefficienti di regressione, è possibile utilizzare i metodi del Paragrafo 7.3 per calcolare i loro errori standard.

Calcolo dei moltiplicatori cumulati (continua)

Un trucco del Paragrafo 7.3 è riscrivere la regressione così che i coefficienti in questa regressione siano quelli che interessano – qui, i moltiplicatori cumulati.

Esempio: riscrivere il modello a ritardi distribuiti con 1 ritardo:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t \\ &= \beta_0 + \beta_1 X_t - \beta_1 X_{t-1} + \beta_1 X_{t-1} + \beta_2 X_{t-1} + u_t \\ &= \beta_0 + \beta_1 (X_t - X_{t-1}) + (\beta_1 + \beta_2) X_{t-1} + u_t \end{aligned}$$

o

$$Y_t = \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} + u_t$$

Calcolo dei moltiplicatori cumulati (continua)

Quindi, si ponga $W_{1t} = \Delta X_t$ e $W_{2t} = X_{t-1}$ e si stimi la regressione,

$$Y_t = \beta_0 + \delta_1 W_{1t} + \delta_2 W_{2t} + u_i$$

Quindi

$\delta_1 = \beta_1$ = effetto d'impatto

$\delta_2 = \beta_1 + \beta_2$ = il primo moltiplicatore cumulato

e gli errori standard (HAC) su δ_1 e δ_2 sono gli errori standard per i due moltiplicatori cumulati.

Calcolo dei moltiplicatori cumulati (continua)

In generale, il modello ADL può essere riscritto come,

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \dots + \delta_{q-1} \Delta X_{t-q+1} + \delta_q X_{t-q} + u_t$$

dove

$$\delta_1 = \beta_1$$

$$\delta_2 = \beta_1 + \beta_2$$

$$\delta_3 = \beta_1 + \beta_2 + \beta_3$$

...

$$\delta_q = \beta_1 + \beta_2 + \dots + \beta_q$$

I moltiplicatori cumulati e i loro errori standard HAC possono essere calcolati direttamente con questa regressione trasformata

L'esogeneità è plausibile? Alcuni esempi (Paragrafo 15.7)

Se X è esogena (e valgono le assunzioni 2-4), allora un modello a ritardi distribuiti offre stimatori consistenti degli effetti causali dinamici.

Come nella regressione multipla con dati sezionali, si deve valutare con occhio critico se X sia esogena in qualsiasi applicazione:

- X è esogena, cioè $E(u_t | X_t, X_{t-1}, \dots) = 0$?
- X è strettamente esogena, cioè $E(u_t | \dots, X_{t+1}, X_t, X_{t-1}, \dots) = 0$?

L'esogeneità (continua)

- È necessario valutare l'esogeneità e l'esogeneità stretta caso per caso
- Spesso l'esogeneità non è plausibile nei dati relativi serie temporali per la presenza della causalità simultanea
- L'esogeneità stretta è raramente plausibile nei dati relativi a serie temporali a causa del feedback.

Stima degli effetti causali dinamici: Riepilogo (Paragrafo 15.8)

- Gli effetti causali dinamici sono misurabili in teoria mediante un esperimento casualizzato controllato con rilevamenti ripetuti nel tempo.
- Quando X è esogena, è possibile stimare gli effetti causali dinamici con una regressione a ritardi distribuiti
- Se u è serialmente correlato, gli errori standard OLS convenzionali sono sbagliati; si devono usare errori standard HAC
- Per decidere se X è esogena, si deve riflettere bene sulle particolarità del problema!

Stazionarietà congiunta - vettoriale

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} \text{ è stazionario (congiunt.) in covarianza}$$

Assunzioni:

I) $E[\underline{z}_t] = \mu$ e finito $\forall t$

II) $\text{Var}[\underline{z}_t] = \Sigma_z$ def. pos. e simmetrica e finita $\forall t$

III) $\text{Cov}(\underline{z}_t, \underline{z}_{t-k}) = \Gamma(z_t, z_{t-k}) = \Gamma(k)$

Gomme meiuse.

$\hookrightarrow F_2$. scalare di k .

F_2 . d:

covarianza
incrociata

$$\begin{aligned}\gamma_{11}(k) &= \text{cov}(y_t, y_{t-k}) \\ \gamma_{22}(k) &= \text{cov}(x_t, x_{t-k}) \\ \gamma_{21}(k) &= \text{cov}(x_t, y_{t-k}) \\ \gamma_{12}(k) &= \text{cov}(y_t, x_{t-k})\end{aligned}$$

$$\Sigma(k) = \begin{pmatrix} y_t & x_t \\ y_{t-k} & x_{t-k} \end{pmatrix}$$

$$\Sigma_v = \Gamma(0) = \begin{pmatrix} y_t & x_t \\ y_{t-k} & x_{t-k} \end{pmatrix}$$

$$\begin{matrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_x^2 \end{matrix}$$

$\sigma > 0$
Def pos e
Simmetrica

$$\sigma_{11} = \text{cov}(y_t, y_t) = \text{Var}(y_t)$$

$$\sigma_{22} = \text{cov}(x_t, x_t)$$

$$\sigma_{12} = \sigma_{21} = \text{cov}(x_t, y_t)$$

23/11

$y_t \sim \text{VAR}(p)$ Modello Auto regressivo Vettoriale

$$(m \times 1) \quad y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{bmatrix}$$

Es. AR(1) $y_t = \beta_0 + \phi y_{t-1} + \varepsilon_t$

ai g:20

VAR(1)

$$\underline{Y}_t = \underline{\delta} + \Phi \underline{Y}_{t-1} + \underline{\varepsilon}_t$$

Processo vettoriale degli errori

$m=2$

$$\begin{cases} Y_{1,t} = \delta_1 + \phi_{11} Y_{1,t-1} + \phi_{12} Y_{2,t-1} + \varepsilon_{1,t} \\ Y_{2,t} = \delta_2 + \phi_{21} Y_{1,t-1} + \phi_{22} Y_{2,t-1} + \varepsilon_{2,t} \end{cases} \quad \begin{array}{l} Y_{1,t} \sim ADL(1,1) \\ Y_{2,t} \sim AD2(1,1) \end{array}$$

Possiamo stimare i coefficienti separatamente con gli OLS. Lo stimatore è assolt. normale

Nel caso non poniamo vincoli lo stimatore è meno efficiente

Assunzioni:

1) $E[\underline{\varepsilon}_t | \mathcal{F}_{t-1}, \underline{Y}_{t-2}, \dots] = 0 \rightarrow \underline{\varepsilon}_t$ processo vettoriale a differenze martingale (VDM) rispetto a \mathcal{F}_t

$$\hookrightarrow a) E[\underline{\varepsilon}_t] = 0$$

$$b) \text{Cov}(\underline{\varepsilon}_t, \underline{Y}_{t-j}) = 0$$

c) Non autocorrelazione degli errori (vedi dopo)

2) { 2.a) Stazionarietà del processo vettoriale $(\underline{Y}_t)_t \sim \text{VAR}(p)$

2.b) Processo $(\underline{Y}_t)_t$ assintoticamente indipendente

symmetrica identica

$$I_m \underline{Y}_t - \Phi_1 \underline{Y}_{t-1} - \dots - \Phi_p \underline{Y}_{t-p} = \underline{\delta} + \underline{\varepsilon}_t$$

→ Polinomio matriciale $\Phi(L)$ di grado p

$$(I_m - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p) \underline{Y}_t = \underline{\delta} + \underline{\varepsilon}_t$$

$$\Phi(L)^{-1} \cdot \Phi(L) = I_m$$

$$\Phi(L) \cdot \underline{Y}_t = \underline{\delta} + \underline{\varepsilon}_t$$

Se $\exists! \Phi^{-1}(L) \rightarrow \underline{Y}_t \sim VMA(\infty)$ staz. in covarianza

↪ Nel caso VAR(1) $\underline{Y}_t = (I_m - \Phi)^{-1} \underline{\delta} + \sum_{j=0}^{+\infty} \Phi^j \underline{\varepsilon}_{t-j}$

$$\Phi^0 = I_m$$

→ Non c'è più un'equazione per una matrice, perché Φ è fix.

$\Phi(z) \rightarrow \det(\Phi(z)) = 0 \quad z \in \mathbb{C}$

$\text{VAR}(1) \sim \underline{Y}_t \quad \det(I_2 - \Phi \cdot z) = z \quad \text{eq. in } z \text{ di grado } m-p=2$

$$|z_i| > 1 \quad i=1, \dots, m-p \quad \lambda_i = z_i^{-1} \text{ sono } |\lambda_i| < 1 \quad \forall i$$

Sc \downarrow riferito a 1) a b e c
 \downarrow c) \mathcal{Y}_t è stazionario $\rightarrow \mathcal{F}_{t-1} = \{\varepsilon_{t-1}, \dots\}$ equiv. a $\mathcal{F}_{t-1} = \{\mathcal{Y}_{t-1}, \dots\}$

$$\text{c)} \quad \text{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0 \quad j \neq 0$$

$$\text{Cov}(\varepsilon_{1,t}, \varepsilon_{1,t-j}) = 0 \quad \vdots$$

$$\text{Cov}(\varepsilon_{2,t}, \varepsilon_{2,t-j}) = 0 \quad "$$

$$\text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t-j}) = 0 \quad j \neq 0$$

In generale $\text{Cov}(\varepsilon_{1,t}, \varepsilon_{2,t}) \neq 0$

Possono essere correlati nello stesso istante temporale ma non nel passato

$$\mathcal{Y}_{t|t-1} = E[\mathcal{Y}_t | \mathcal{F}_{t-1}] = \delta + \sum_{j=1}^{m-1} \phi_j \mathcal{Y}_{t-j} \rightarrow \text{Fz. di regressione vettoriale (o multivariata) lineare e dinamica}$$

$$\text{Previsione dinamica: } \mathcal{Y}_{T+j|T} = E[\mathcal{Y}_{T+j} | \mathcal{F}_T]$$

27/11 Criteri informativi per la scelta del ritardo

Da chapter 7 (slides 48)

I modelli VAR sono overparametrizzati.

es. $\begin{cases} q=3 \\ k=3 \end{cases} \Rightarrow 30 \text{ parametri}$

A d esempio g equazioni e k ritardi per ogni equazione, quindi $g+kq^2$ parametri

Cross-Equation Restrictions

Faccio un test che mi indica il più ottimale fra un set di valori

Per questo test devo fare le stime di sistema, ottenendo la distribuzione asintotica di tutti gli stimatori.

$$LR = T [\log |\hat{\Sigma}_r| - \log |\hat{\Sigma}_0|] \sim \chi_{r-u} \xrightarrow{\text{Numero di restrizioni}}$$

dove $|\hat{\Sigma}_r|$ è la matrice di varianza covarianza del modello ristretto
 $|\hat{\Sigma}_0|$ --
 T è la sample size

Metodo dei criteri informativi:

Abbiamo bisogno di scegliere un criterio informativo multivariato come il AIC, SBIC, HQIC.

E' importante che i modelli confrontati abbiano le stesse dimensione compionaria

Questo metodo è più usato confronto al precedente

// Spiegazione di GRETL

Il Test Portmanteau è lo stesso del test di Ljung-Box ma è di sistema, sul vettore degli errori.

↳ I residui sono indipendentemente distribuiti
Autocorrelazione degli errori

Se si va su Analisi > Radici del VAR si può controllare se il modello è stazionario, ovvero se le soluzioni stanno all'interno del cerchio ($|z| < 1$)

27/11

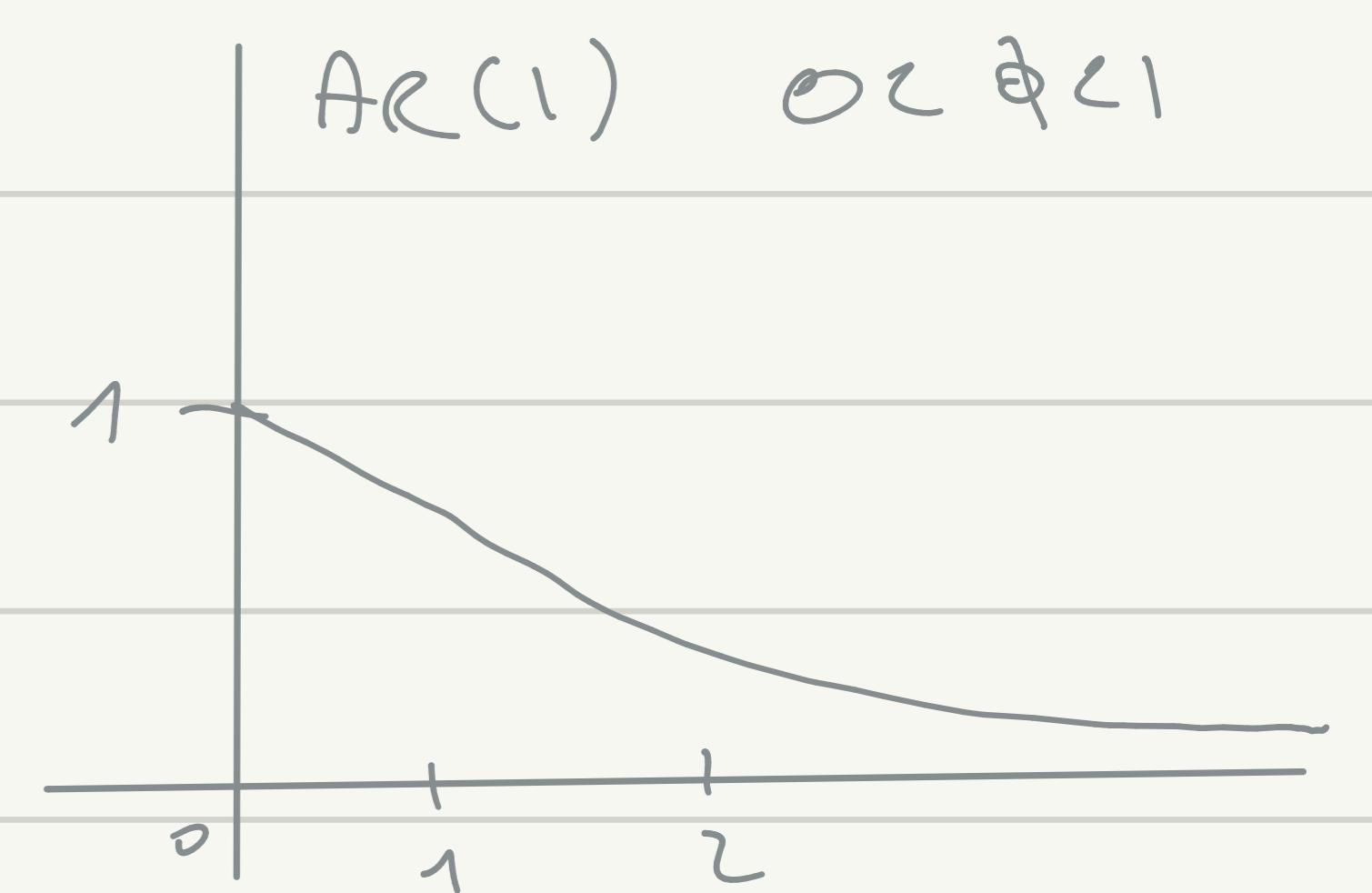
Impulse Response Function (i.r.f.) → Heijette

$Y_t \sim AR(p)$ staz. in cov. $\rightarrow Y_t \sim MA(\infty)$

$$Y_t = \mu + \sum_{i=0}^{+\infty} \Psi_i \varepsilon_{t-i} \quad \text{con } \Psi_0 = 1$$

$$\frac{\partial Y_t}{\partial \varepsilon_{t-i}} = \hat{\Psi}_i \quad i=0, 1, \dots$$

$$\hat{\sigma}_{\varepsilon}^2 \cdot \hat{\Psi}_i$$



$Y_t \sim VAR(p)$ staz. in cov.

$Y_t \sim VMA(\infty)$

$$\begin{matrix} i=1, \dots, m \\ j=1, \dots, m \\ k=0, 1, 2, \dots \end{matrix}$$

$$\frac{\partial Y_t}{\partial \varepsilon_{j,t-k}}$$

$$Y_t = \mu + \sum_{i=0}^{+\infty} \Psi_i \varepsilon_{t-i}$$

$$\Psi_0 = I_m$$

$$\Psi_t = \nu + \sum_{i=0}^{+\infty} \Psi_i \underline{\xi}_{t-i} = \nu + \Psi_0 \cdot \underline{\xi}_t + \Psi_1 \underline{\xi}_{t-1} + \dots$$

Caso $m=2$

$$\Psi_0 = I_2$$

$$\begin{aligned}\frac{\partial y_{1t}}{\sum_{1t}} &= 1 \\ \frac{\sum_{1t}}{\sum_{2t}} &= 0\end{aligned}$$

$$\begin{aligned}\frac{\partial y_{2t}}{\sum_{1t}} &= 0 \\ \frac{\partial y_{1t}}{\sum_{2t}} &= 1\end{aligned}$$

$$\begin{bmatrix} 1 \cdot \underline{\xi}_{1t} + 0 \cdot \underline{\xi}_{2t} \\ 0 \cdot \underline{\xi}_{1t} + 1 \cdot \underline{\xi}_{2t} \end{bmatrix} = \Psi_0 \cdot \underline{\xi}_t$$

$$\begin{aligned}\frac{\partial y_{1t}}{\partial \underline{\xi}_{1t-k}} &= \Psi_{11}(k) \\ \frac{\partial y_{1t}}{\partial \underline{\xi}_{2t-k}} &= \Psi_{12}(k)\end{aligned}$$

$$\begin{aligned}\frac{\partial y_{2t}}{\partial \underline{\xi}_{1t-k}} &= \Psi_{21}(k) \\ \frac{\partial y_{2t}}{\partial \underline{\xi}_{2t-k}} &= \Psi_{22}(k)\end{aligned}$$

Effetti: casuali dinamici d: sistema

Relazione tra le variabili

Non è diagonale $\rightarrow \sum_{\underline{\xi}} \text{Corr}(\underline{\xi}_{1t}, \underline{\xi}_{2t}) \neq 0$

$$\nu = 0 \quad \underline{\Psi}_t = \Psi_\infty(L) \cdot \underline{\xi}_t \quad \text{Var}(\underline{\xi}_t) = \Sigma_{\xi} \text{ non diag. def. positive}$$

\exists una matrice P invertibile t.c. $P \cdot \Sigma_{\xi} P' = I_m$

$$\text{Se applico una trasformazione: } \underline{\eta}_t = P \underline{\xi}_t \rightarrow \text{Var}(\underline{\eta}_t) = P \Sigma_{\xi} P' = I_m$$

Errore (o shock) ortonormali

$$P \underline{\Psi}_t = P \Psi_\infty(L) \cdot \underline{\xi}_t \rightarrow \underline{\Psi}_t = [\Psi_\infty(L) \cdot P^{-1}] \cdot [P \cdot \Sigma_{\xi}]$$

$$\underline{\Psi}_t = \sum_{i=0}^{+\infty} (\Psi_i \cdot P^{-1}) \cdot \underline{\eta}_t \rightarrow \text{Var}(\underline{\eta}_t) = I_m$$

$$\begin{aligned}\underline{\Psi}_t &= \sum_{i=0}^{+\infty} \tilde{\Psi}_i \cdot \underline{\eta}_t \\ \Psi_i \cdot P^{-1} &= \tilde{\Psi}_i\end{aligned}$$

$$\tilde{\Psi} = I_m P^{-1} \neq I_m \quad \frac{\partial \Psi_{it}}{\partial \eta_{j,t-k}} = \tilde{\Psi}_{ij}(k) \quad -1 \text{ s.d. } (\underline{\xi}_{it})$$

C'è un problema di identificazione, ovvero se sono gli unici coefficienti ortonormali
La matrice I_m non è unica

$$T^{-1} T \Sigma_{\xi} T' = I_m$$

Decomposizione di Choleski:

$\exists ! T$ invertibile e triangolare

T^{-1} è triangolare

$$T = \begin{bmatrix} t_{11} & 0 & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & t_{mm} \end{bmatrix}$$

$$\text{con } t_{ii} > 0 \quad \forall i \quad \text{t.c. } T \Sigma_{\xi} T' = I_m$$

$$\Sigma_{\xi} = T T' = (T^{-1}) (T^{-1})'$$

$$\underline{Y}_t \sim \text{Var}(1)$$

$$\underline{Y}_t = \underline{\delta} + \phi \underline{Y}_{t-1} + \underline{\varepsilon}_t$$

Se moltiplicavo per P non avrei l'identificazione

$$T \cdot \underline{Y}_t = T \cdot \underline{\delta} + T \cdot \phi \cdot \underline{Y}_{t-1} + T \cdot \underline{\varepsilon}_t$$

Matrice delle interindipendenze delle variabili esogene

Modello multi-equationale dinamico

Esempio caso $m=2$

$$P \cdot \underline{Y}_t = \begin{bmatrix} Y_{1t} & -a Y_{1t} \\ -b Y_{1t} & Y_{2t} \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & -a \\ -b & 1 \end{bmatrix}$$

Matrice piena e invertibile

$$\begin{cases} Y_{1t} = a Y_{2t} + \dots + \eta_t \\ Y_{2t} = b Y_{1t} + \dots + \eta_t \end{cases}$$

Catene ricorsive

$$\begin{aligned} t_{11} Y_{1t} &= 0 + \dots \rightarrow Y_{1t} = 0 Y_{2t} + \dots \\ t_{21} Y_{2t} &= -t_{21} Y_{1t} + \dots \rightarrow Y_{2t} = -\frac{t_{21}}{t_{22}} Y_{1t} + \dots \end{aligned}$$

$$T = \begin{bmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{bmatrix} \Rightarrow T \cdot \underline{Y}_t = \begin{bmatrix} t_{11} Y_{1t} + 0 \cdot Y_{2t} \\ t_{21} Y_{1t} + t_{22} Y_{2t} \end{bmatrix}$$

Catene casuale ricorsive indotte dalla decomposizione Cholesky

Esempio caso $m=3$

$$\begin{cases} Y_{1t} = 0 \cdot Y_{2t} + 0 \cdot Y_{3t} + \dots \\ Y_{2t} = a \cdot Y_{1t} + 0 \cdot Y_{3t} + \dots \\ Y_{3t} = b_1 \cdot Y_{1t} + b_2 Y_{2t} + \dots \end{cases}$$

→ GDP spread

Modelli per dati: PANEL

y_{it}

$i=1, \dots, n$

$t=1, \dots, T$

$i=1, \dots, n$

$t=1, \dots, T_i$

Panel bilanciato: non ci sono dati mancanti

Panel sbilanciato: ogni unità può avere intervalli di tempo diversi

Nei dati: Cross-Section

Soluzione

Possso avere

$\text{Cov}(X_i, v_i) \neq 0$

$$Y_i = \beta_0 + \beta_1 X_i + \nu_i \rightarrow z_i: \text{variabile omessa correlata con } X_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 z_i + v_i \rightarrow \text{Cov}(X_i, v_i) = 0$$

$\text{Cov}(X_{it}, v_{it}) \neq 0 \quad \forall i, t$

Nei dati: Panel

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_{it} \rightarrow z_i: \text{variabile omessa correlata con } X_{it}$$

Supponiamo $z_{it} = z_i, \forall t$

Caso 2 osservazioni Y_{it}, X_{it} non osservo $z_{it} = z_i \forall t \rightarrow$ Ipotizzo che non varia nel tempo

$$\begin{cases} Y_{i2} = \beta_0 + \beta_1 X_{i2} + \nu_{i2} \rightarrow \beta_2 \cdot z_i + v_{i2} \\ Y_{i1} = \beta_0 + \beta_1 X_{i1} + \nu_{i1} \rightarrow \beta_2 \cdot z_i + v_{i1} \end{cases} \neq 0$$

$$\Delta Y_i = 0 + \beta_1 \Delta X_i + \Delta \mu_i$$

$$\text{Cov}(\Delta X_i, \Delta \mu_i) = 0 \quad \Delta v_i = v_{i2} - v_{i1}$$

Controllo le variabili z_i senza doverle osservare

28/11 GRETL Modelli VAR

Usa lo stimatore POOL OLS che significa andare direttamente al campione completo e scrivere il modello, anziché la cross-section ma con un modello panel con 2 indici

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \nu_{it}$$

Devo minimizzare gli errori al quadrato

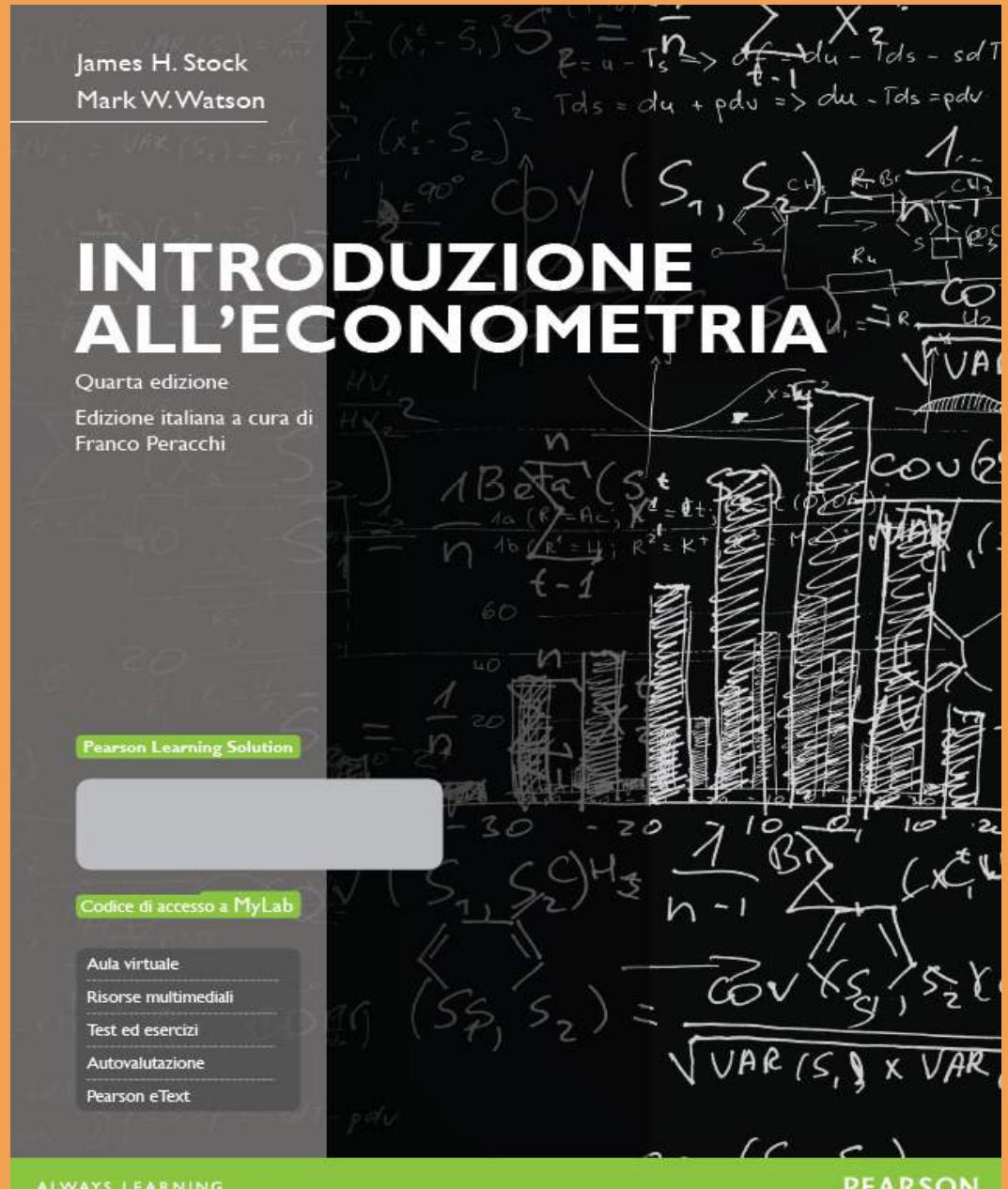
$$\min_{\{\beta_0, \beta_1\}} \sum_i \sum_t \nu_{it}^2$$

$i=1, \dots, n$

$t=1, \dots, T$

Capitolo 10

Regressione con dati panel



Sommario

1. Dati panel: cosa e perché
2. Dati panel con due periodi temporali
3. Regressione con effetti fissi
4. Regressione con effetti temporali
5. Errori standard per regressione con effetti fissi
6. Applicazione a guida in stato di ebbrezza e sicurezza stradale

Dati panel: cosa e perché (Paragrafo 10.1)

Un **panel** contiene osservazioni su più unità (individui, stati, imprese) in cui ogni entità è osservata in due o più istanti temporali diversi.

Esempi:

- Dati su 420 distretti scolastici della California nel 1999 e ancora nel 2000, per 840 osservazioni in totale.
- Dati su 50 stati USA, ognuno è osservato per 3 anni, per un totale di 150 osservazioni.
- Dati su 1000 individui, in quattro mesi diversi, per 4000 osservazioni in totale.

Notazione per dati panel

Un doppio pedice distingue unità (stati) e periodi temporali (anni)

i = unità (stato), n = numero di entità,
perciò $i = 1, \dots, n$

t = periodo temporale (anno), T = numero di periodi temporali
perciò $t = 1, \dots, T$

Dati: supponiamo di avere 1 regressore. I dati sono:

$$(X_{it}, Y_{it}), i = 1, \dots, n, t = 1, \dots, T$$

Notazione per dati panel (continua)

Dati panel con k regressori:

$$(X_{1it}, X_{2it}, \dots, X_{kit}, Y_{it}), i = 1, \dots, n, t = 1, \dots, T$$

n = numero di unità (stati)

T = numero di periodi temporali (anni)

Un po' di gergo...

- I dati panel sono chiamati anche ***dati longitudinali***
- ***panel bilanciato***: non ci sono osservazioni che mancano, cioè tutte le variabili sono osservate per tutte le unità (stati) e tutti i periodi temporali (anni)

Perché sono utili i dati panel?

Con i dati panel possiamo controllare per fattori che:

- Variano tra le unità ma non nel tempo
- Potrebbero causare distorsione da variabili omesse se fossero omessi
- Sono inosservati o non misurati, e perciò non possono essere inclusi in una regressione multipla

Ecco l'idea chiave:

Se una variabile omessa non varia nel tempo, allora qualsiasi variazione in Y nel tempo non può essere causata dalla variabile omessa.

Questi fattori omessi potrebbero causare distorsione da variabili omesse.

Esempio 1: densità del traffico. Supponiamo:

- I. Elevata densità del traffico significa più morti sulle strade
- II. Gli stati con minore densità di traffico (all'ovest) hanno imposte sugli alcolici minori
- Allora le due condizioni per la distorsione da variabili omesse sono soddisfatte. Nello specifico, "imposte elevate" potrebbero riflettere "alta densità di traffico" (perciò il coefficiente OLS sarebbe distorto positivamente
 - imposte elevate, più morti)
- I dati panel ci consentono di eliminare la distorsione da variabili omesse quando le variabili omesse sono costanti nel tempo in un dato stato.

Dati panel con due periodi temporali (Paragrafo 10.2)

Consideriamo il modello dei dati panel,

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + u_{it}$$

Z_i è un fattore che non cambia nel tempo (densità), almeno durante gli anni per cui abbiamo dati.

- Supponiamo che Z_i non sia osservato, perciò la sua omissione potrebbe comportare distorsione da variabili omesse.
- L'effetto di Z_i può essere eliminato usando $T = 2$ anni.

↳ ΔY_t

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 U_i + \gamma_i \rightarrow \text{Effetti individuali}$$

Se faccio le differenze dovrebbero sparire tutte le variabili che variano nel tempo e non negl. stati

Regressione con effetti fissi

(Paragrafo 10.3)

$$Y_{it} = \beta_0 + \beta_1 t + \beta_2 X_{it} + u_{it}$$

β_0 → variabile latente

t → intercetto del modello che varia tra uno stato e l'altro

E se si hanno più di 2 periodi temporali ($T > 2$)?

$$v_{it} = \gamma_i + u_{it}$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}, i = 1, \dots, n, T = 1, \dots, T \quad \text{Cor}(v_{it}, X_{it}) \neq 0$$

$$\downarrow E[u_{it}|X_{i1}, X_{i2}, \dots] = 0$$

Possiamo riscriverlo in due modi utili: \hookrightarrow Prime condizioni del modello

1. modello di regressione "con $n-1$ regressori binari"
2. modello di regressione "con effetti fissi"

Prima lo riscriviamo nella forma "con effetti fissi".

Supponiamo di avere $n = 3$ stati: California, Texas e Massachusetts.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

Regressione per la California ($i = CA$):

$$\begin{aligned} Y_{CA,t} &= \beta_0 + \beta_1 X_{CA,t} + \beta_2 Z_{CA} + u_{CA,t} \\ &= (\beta_0 + \beta_2 Z_{CA}) + \beta_1 X_{CA,t} + u_{CA,t} \end{aligned}$$

O

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

- $\alpha_{CA} = \beta_0 + \beta_2 Z_{CA}$ non cambia nel tempo
- α_{CA} è l'intercetta per CA, e β_1 è la pendenza
- L'intercetta è specifica per CA, ma la pendenza è la stessa in tutti gli stati: rette parallele.

Per TX:

$$\begin{aligned}Y_{TX,t} &= \beta_0 + \beta_1 X_{TX,t} + \beta_2 Z_{TX} + u_{TX,t} \\&= (\beta_0 + \beta_2 Z_{TX}) + \beta_1 X_{TX,t} + u_{TX,t}\end{aligned}$$

o

$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}, \text{ where } \alpha_{TX} = \beta_0 + \beta_2 Z_{TX}$$

Mettendo insieme le rette dei tre stati:

$$Y_{CA,t} = \alpha_{CA} + \beta_1 X_{CA,t} + u_{CA,t}$$

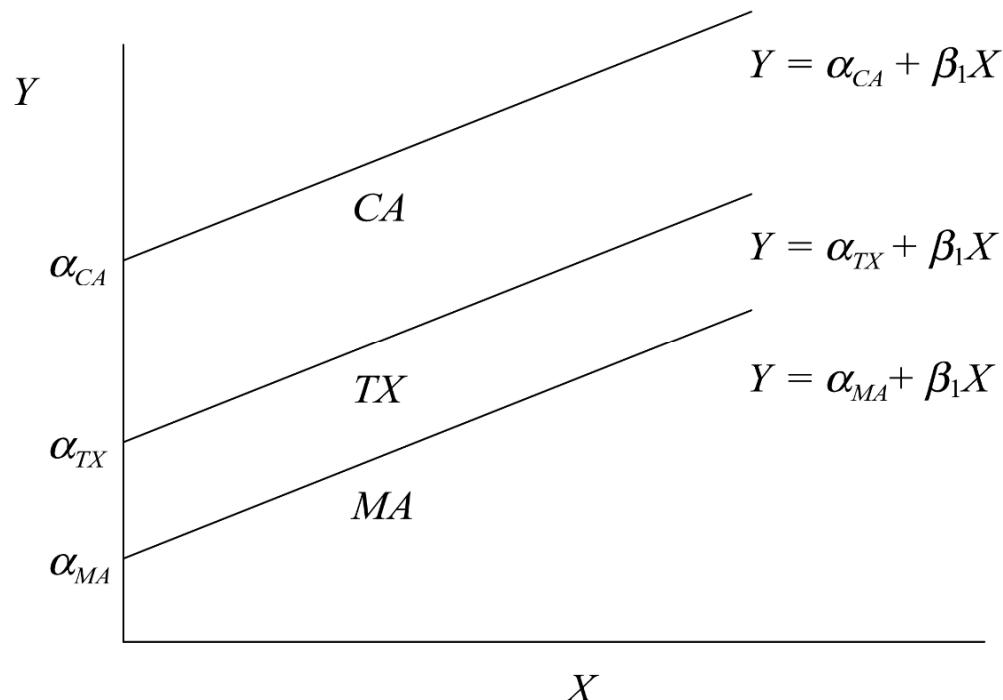
$$Y_{TX,t} = \alpha_{TX} + \beta_1 X_{TX,t} + u_{TX,t}$$

$$Y_{MA,t} = \alpha_{MA} + \beta_1 X_{MA,t} + u_{MA,t}$$

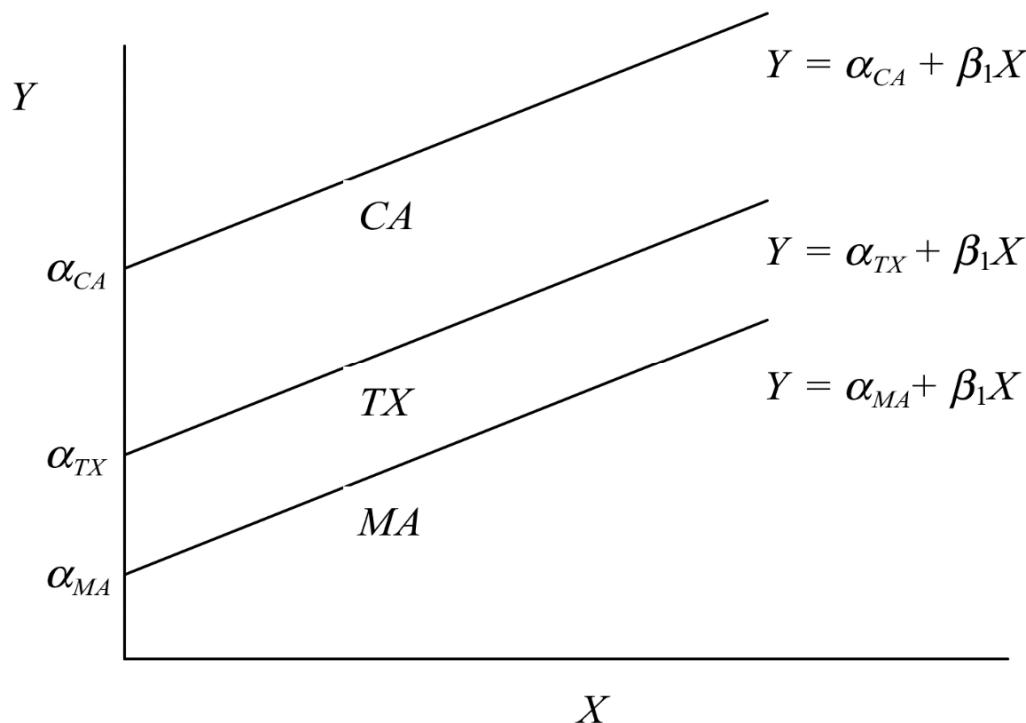
o

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}, i = CA, TX, MA, T = 1, \dots, T$$

Le rette di regressione per ciascuno stato



Si ricordi che gli spostamenti nell'intercetta possono essere rappresentati mediante regressori binari...



Nella forma con regressori binari:

$$Y_{it} = \beta_0 + \gamma_{CA} DCA_i + \gamma_{TX} DTX_i + \beta_1 X_{it} + u_{it}$$

- $DCA_i = 1$ se lo stato è CA , $= 0$ altrimenti
- $DTX_t = 1$ se lo stato è TX , $= 0$ altrimenti
- si lascia fuori DMA_i (*perché?*)

Riepilogo: due modi per scrivere il modello con effetti fissi

1. Forma con “ $n-1$ regressori binari” più costante

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \dots + \gamma_n Dn_i + u_{it}$$

dove $D2_i = \begin{cases} 1 & \text{per } i=2 \text{ (stato n. 2)} \\ 0 & \text{altrimenti} \end{cases}$, ecc.

2. Forma con “effetti fissi” :

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

- α_i è chiamato “effetto fisso dello stato” o “effetto dello stato” – è l’effetto costante (fisso) di trovarsi nello stato i

Regressione con effetti fissi: stima

Tre metodi di stima:

1. Regressione OLS con “ $n-1$ regressori binari”
 2. Regressione OLS con “unità in deviazioni dalle medie”
 3. Specificazione “prima e dopo”, senza un’intercetta (funziona solo per $T = 2$)
-
- Questi tre metodi producono identiche stime dei coefficienti di regressione e identici errori standard.
 - Abbiamo già utilizzato la specificazione “prima e dopo” (1988 meno 1982) – che però funziona solo per $T = 2$ anni
 - I metodi 1 e 2 funzionano per un generico T
 - Il metodo 1 è praticabile solo quando n non è troppo grande

1. Regressione OLS con “ n -1 regressori binari”

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + \dots + \gamma_n Dn_i + u_{it} \quad (1)$$

dove $D2_i = \begin{cases} 1 & \text{per } i=2 \text{ (stato n. 2)} \\ 0 & \text{altrimenti} \end{cases}$ etc.

- Prima si creano le variabili binarie $D2_i, \dots, Dn_i$
- Poi si stima (1) mediante OLS
- L'inferenza (verifiche di ipotesi, intervalli di confidenza) è come di consueto (con errori standard robusti all'eteroschedasticità)
- Non è pratico quando n è molto grande (per esempio se $n = 1000$ lavoratori)

2. Regressione OLS con “unità in deviazioni dalle medie”

Modello di regressione con effetti fissi:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

Le medie delle unità soddisfano:

$$\frac{1}{T} \sum_{t=1}^T Y_{it} = \alpha_i + \beta_1 \frac{1}{T} \sum_{t=1}^T X_{it} + \frac{1}{T} \sum_{t=1}^T u_{it}$$

Deviazioni dalle medie:

$$Y_{it} - \frac{1}{T} \sum_{t=1}^T Y_{it} = \beta_1 \left(X_{it} - \frac{1}{T} \sum_{t=1}^T X_{it} \right) + \left(u_{it} - \frac{1}{T} \sum_{t=1}^T u_{it} \right)$$

Regressione OLS con “unità in deviazioni dalle medie” (continua)

$$Y_{it} - \frac{1}{T} \sum_{t=1}^T Y_{it} = \beta_1 \left(X_{it} - \frac{1}{T} \sum_{t=1}^T X_{it} \right) + \left(u_{it} - \frac{1}{T} \sum_{t=1}^T u_{it} \right)$$

o

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}$$

dove $\tilde{Y}_{it} = Y_{it} - \frac{1}{T} \sum_{t=1}^T Y_{it}$ e $\tilde{X}_{it} = X_{it} - \frac{1}{T} \sum_{t=1}^T X_{it}$

- \tilde{X}_{it} e \tilde{Y}_{it} sono dati “in deviazioni dalle medie”
- Per $i=1$ e $t = 1982$, \tilde{Y}_{it} è la differenza tra il tasso di mortalità in Alabama nel 1982 e il suo valor medio in Alabama calcolato in media su tutti e sette gli anni.

Regressione OLS con “unità in deviazioni dalle medie” (continua)

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{\mu}_{it} \quad (2)$$

dove $\tilde{Y}_{it} = Y_{it} - \frac{1}{T} \sum_{t=1}^T Y_{it}$, ecc.

- Prima si costruiscono le unità in deviazioni dalle medie \tilde{Y}_{it} e \tilde{X}_{it}
- Poi si stima (2) con la regressione di \tilde{Y}_{it} su \tilde{X}_{it} usando OLS
- È simile all'approccio “prima e dopo”, ma con Y_{it} in deviazione dalla media al posto di Y_{i1} .
- Gli errori standard vanno calcolati in un modo che tenga conto della natura “panel” dei dati (ne parleremo più avanti)
- Si può fare con un unico comando in STATA

Esempio (continua). Per $n = 48$, $T = 7$:

$$\widehat{FatalityRate} = -0,66BeerTax + \text{State fixed effects}$$
$$(0,29)$$

- Va riportata l'intercetta?
- Quanti regressori binari includereste per una stima con il metodo del "regressore binario"?
- Si confrontino pendenza ed errore standard con la stima per il 1988 v. 1982 della specificazione "prima e dopo" ($T = 2$, $n = 48$) (*si noti che è inclusa un'intercetta - ci torniamo poi*):

$$\widehat{FR}_{1988} - \widehat{FR}_{1982} = -0,072 - 1,04(BeerTax_{1988} - BeerTax_{1982})$$
$$(0,065) \quad (0,36)$$

Regressione con effetti temporali (Paragrafo 10.4)

Una variabile omessa potrebbe variare nel tempo ma non tra gli stati:

- auto più sicure (air bag, ecc.); modifiche nelle leggi nazionali
- producono intercette che variano nel tempo
- Sia S_t l'effetto combinato di variabili che cambiano nel tempo ma non tra gli stati ("auto più sicure").
- Il modello di regressione risultante è:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + \beta_3 S_t + u_{it}$$

Soli effetti fissi temporali

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_3 S_t + u_{it}$$

Questo modello può essere ricomposto con un'intercetta che varia da un anno al successivo:

$$\begin{aligned} Y_{i,1982} &= \beta_0 + \beta_1 X_{i,1982} + \beta_3 S_{1982} + u_{i,1982} \\ &= (\beta_0 + \beta_3 S_{1982}) + \beta_1 X_{i,1982} + u_{i,1982} \\ &= \lambda_{1982} + \beta_1 X_{i,1982} + u_{i,1982}, \end{aligned}$$

dove $\lambda_{1982} = \beta_0 + \beta_3 S_{1982}$ Similmente,

$$Y_{i,1983} = \lambda_{1983} + \beta_1 X_{i,1983} + u_{i,1983},$$

dove $\lambda_{1983} = \beta_0 + \beta_3 S_{1983}$, ecc.

Due formulazioni di regressione con effetti temporali

1. Formulazione con “ $T-1$ regressori binari”:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \dots \delta_T BT_t + u_{it}$$

dove $B2_t = \begin{cases} 1 & \text{quando } t=2 \text{ (anno n. 2)} \\ 0 & \text{altrimenti} \end{cases}$, ecc.

2. Formulazione con “effetti fissi” temporali:

$$Y_{it} = \beta_1 X_{it} + \lambda_t + u_{it}$$

Effetti temporali: metodi di stima

1. Regressione OLS con “ $T-1$ regressori binari”

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_{it} + \dots \delta_T BT_{it} + u_{it}$$

- Si creano variabili binarie $B2, \dots, BT$
- $B2 = 1$ se $t = \text{anno n. } 2$, $= 0$ altrimenti
- Si esegue la regressione di Y su $X, B2, \dots, BT$ con OLS
- Dov'è $B1$?

2. Regressione OLS “in deviazione dalle medie dell'anno”

- Si devia Y_{it}, X_{it} dalle medie dell'anno (non dello stato)
- Si stima con OLS usando dati “in deviazione dalle medie dell'anno”

Stima con effetti fissi stato ed effetti fissi temporali

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

- Quando $T = 2$, calcolare la differenza prima e includere una costante è equivalente a (fornisce esattamente la stessa regressione di) includere effetti individuali e temporali.
- Quando $T > 2$, esistono vari modi equivalenti di incorporare effetti individuali e temporali:
 - deviazione dalle medie e $T - 1$ indicatori temporali (viene fatto nel seguente esempio con STATA)
 - deviazione dalle medie temporali e $n - 1$ indicatori individuali
 - $T - 1$ indicatori temporali e $n - 1$ indicatori individuali
 - deviazione dalle medie individuali e temporali

Le assunzioni e gli errori standard della regressione con effetti fissi (Paragrafo 10.5 e Appendice 10.2)

Sotto le assunzioni dei minimi quadrati nella versione per dati panel, lo stimatore OLS con effetti fissi di β_1 ha distribuzione normale.

Tuttavia, è necessario introdurre una nuova formula dell'errore standard, quella per dati raggruppati, o "clustered". Questa nuova formula è necessaria perché le osservazioni per la stessa unità non sono indipendenti (è la stessa unità!), anche se le osservazioni di unità diverse sono indipendenti se tali unità sono ottenute mediante campionamento casuale semplice.

Qui consideriamo il caso di effetti fissi individuali. Gli effetti temporali possono semplicemente essere inclusi quali regressori binari aggiuntivi.

Assunzioni dei minimi quadrati per dati panel

Si consideri una singola X :

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

1. $E(u_{it}|X_{i1}, \dots, X_{iT}, \alpha_i) = 0$.
2. $(X_{i1}, \dots, X_{iT}, u_{i1}, \dots, u_{iT}), i = 1, \dots, n$, sono i.i.d. dalla distribuzione congiunta.
3. (X_{it}, u_{it}) hanno momenti quarti finiti.
4. Non vi è collinearità perfetta (molteplicità di X)

Le assunzioni 3 e 4 sono identiche al caso dei minimi quadrati, le assunzioni 1 e 2 sono diverse.

Assunzione 1: $E(u_{it}|X_{i1}, \dots, X_{iT}, \alpha_i) = 0$

- u_{it} ha media zero, dato l'effetto fisso e l'intera storia delle X per l'unità corrispondente
- Questa è un'estensione della precedente assunzione 1 della regressione multipla
- Ciò significa che non vi sono effetti passati omessi (qualsiasi effetto passato di X deve essere incluso esplicitamente)
- Inoltre, non c'è feedback da u su X futuri:
 - il fatto che uno stato abbia un tasso di mortalità particolarmente alto quest'anno non influisce sull'aumento delle imposte sulla birra.
 - talvolta questa assunzione di "assenza di feedback" è plausibile, talvolta no. Ci torneremo quando affronteremo le serie temporali.

Assunzione 2: $(X_{i1}, \dots, X_{iT}, u_{i1}, \dots, u_{iT})$, $i = 1, \dots, n$, sono i.i.d. dalla distribuzione congiunta.

- È un'estensione dell'assunzione 2 per la regressione multipla con dati sezionali
- È soddisfatta se le unità sono prese a caso dalla popolazione mediante campionamento casuale semplice.
- **Non** richiede che le osservazioni siano i.i.d. *nel tempo* per la stessa unità – sarebbe irrealistico. Il fatto che uno stato abbia un'imposta sulla birra elevata quest'anno è un buon predittore del (è correlato con) fatto che avrà un'imposta sulla birra elevata l'anno seguente. Similmente, il termine d'errore per un'unità in un anno è plausibilmente correlato con il suo valore l'anno dopo, cioè $\text{corr}(u_{it}, u_{it+1})$ è plausibilmente diverso da zero.

Autocorrelazione (correlazione seriale)

Supponiamo che una variabile Z sia osservata in diverse date t , perciò le osservazioni sono su Z_t , $t = 1, \dots, T$. (consideriamo che vi sia una sola unità). Allora Z_t è detta **autocorrelata** o **serialmente correlata** se $\text{corr}(Z_t, Z_{t+j}) \neq 0$ per date $j \neq 0$.

- “Autocorrelazione” significa correlazione con se stesso.
- $\text{cov}(Z_t, Z_{t+j})$ è detta **j -esima autocovarianza** di Z_t . $\delta(j)$
- Nell’esempio della guida in stato di ebbrezza, u_{it} include la variabile omessa delle condizioni meteo dell’anno per lo stato i . Se gli inverni nevosi si presentano in gruppi (uno segue l’altro), allora u_{it} sarà autocorrelata (*perché?*)
- In molte applicazioni con dati panel, u_{it} è plausibilmente autocorrelata.

Indipendenza e autocorrelazione in dati panel, un'immagine:

	$i = 1$	$i = 2$	$i = 3$	\vdots	$i = n$
$t = 1$	u_{11}	u_{21}	u_{31}	\vdots	u_{n1}
M	M	M	M	L	M
$t = T$	u_{1T}	u_{2T}	u_{3T}	L	u_{nT}

← Campionamento i.i.d. tra le unità →

- Se le unità sono ottenute per campionamento casuale semplice, allora (u_{i1}, \dots, u_{iT}) è indipendente da (u_{j1}, \dots, u_{jT}) per unità diverse con $i \neq j$.
- Ma se i fattori omessi compresi in u_{it} sono serialmente correlati, allora u_{it} è serialmente correlato.

Sotto le assunzioni dei minimi quadrati per dati panel:

- Lo stimatore OLS con effetto fisso $\hat{\beta}_1$ è non distorto, consistente e ha distribuzione asintotica normale
- Tuttavia, i consueti errori standard OLS (sia di omoschedasticità pura sia robusti all'eteroschedasticità) saranno in generale sbagliati perché assumono che u_{it} non sia serialmente correlata.
 - In pratica, gli errori standard OLS spesso sottostimano l'incertezza del campionamento reale: se u_{it} è correlato nel tempo, non si hanno molte informazioni (molta variazione casuale) come si avrebbero se u_{it} fosse incorrelata.
 - Il problema si risolve usando errori standard “clustered”.

Errori standard per dati raggruppati

- Gli errori standard per dati raggruppati stimano la varianza di $\hat{\beta}_1$ quando le variabili sono i.i.d. tra le unità ma sono potenzialmente autocorrelate in una unità.
- È più facile comprenderli se si considera prima il problema più semplice di stimare la media di Y usando dati panel...

Errori standard clustered per la media stimata con dati panel

$$Y_{it} = \mu + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T$$

Lo stimatore della media μ è $\bar{Y} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T Y_{it}$.

È utile scrivere \bar{Y} come media tra le unità del valore medio per ciascuna unità:

$$\bar{Y} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T Y_{it} = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T Y_{it} \right) = \frac{1}{n} \sum_{i=1}^n \bar{Y}_i ,$$

dove $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$ è la media campionaria per l'unità i .

Poiché le osservazioni sono i.i.d. tra le entità, $(\bar{Y}_1, \dots, \bar{Y}_n)$ sono i.i.d. Quindi, se n è grande, vale il TLC e

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n \bar{Y}_i \xrightarrow{d} N(0, \sigma_{\bar{Y}_i}^2 / n), \text{ dove } \sigma_{\bar{Y}_i}^2 = \text{var}(\bar{Y}_i).$$

- L'errore standard di \bar{Y} è la radice quadrata di uno stimatore di $\sigma_{\bar{Y}_i}^2 / n$.
- Lo stimatore naturale di $\sigma_{\bar{Y}_i}^2$ è la varianza campionaria di \bar{Y}_i , $s_{\bar{Y}_i}^2$. Questo ci fornisce la formula dell'errore standard per dati raggruppati per \bar{Y} calcolata usando dati panel:

$$\text{Errore standard Clustered di } \bar{Y} = \sqrt{\frac{s_{\bar{Y}_i}^2}{n}}, \text{ dove } s_{\bar{Y}_i}^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{Y}_i - \bar{Y})^2$$

Che cos'hanno di speciale gli errori standard per dati raggruppati?

- Non molto, in verità – la procedura di derivazione vista in precedenza è la stessa usata nel Capitolo 3 per derivare l'errore standard della media campionaria, con la differenza che qui i "dati" sono le medie di unità i.i.d. ($\bar{Y}_1, \dots, \bar{Y}_n$) anziché una singola osservazione i.i.d. per ciascuna unità.
- C'è una caratteristica importante: nella derivazione dell'errore standard per dati raggruppati non abbiamo mai assunto che le osservazioni siano i.i.d. *in* una unità. Quindi abbiamo implicitamente consentito la correlazione seriale in una unità.
- E la correlazione seriale, dov'è finita? Determina $\sigma_{\bar{Y}_i}^2$, la varianza di $\bar{Y}_i \dots$

La correlazione seriale in Y_{it} entra in $\sigma_{\bar{Y}_i}^2$:

$$\begin{aligned}\sigma_{\bar{Y}_i}^2 &= \text{var}(\bar{Y}_i) \\ &= \text{var}\left(\frac{1}{T} \sum_{t=1}^T Y_{it}\right) = \frac{1}{T^2} \text{ var}(Y_{i1} + Y_{i2} + \dots + Y_{iT}) \\ &= \frac{1}{T^2} \{ \text{var}(Y_{i1}) + \text{var}(Y_{i2}) + \dots + \text{var}(Y_{iT}) \\ &\quad + 2\text{cov}(Y_{i1}, Y_{i2}) + 2\text{cov}(Y_{i1}, Y_{i3}) + \dots + 2\text{cov}(Y_{iT-1}, Y_{iT}) \}\end{aligned}$$

- Se Y_{it} è serialmente incorrelata, tutte le autocovarianze = 0 e abbiamo la consueta derivazione del Capitolo 3.
- Se queste autocovarianze non sono zero, la formula consueta (che le pone a 0) sarà errata.
- Se queste autocovarianze sono positive, la formula consueta sottostimerà la varianza di \bar{Y}_i .

La “magia” degli errori standard per dati raggruppati è che, operando al livello delle unità e delle loro medie \bar{Y}_i , non occorre preoccuparsi di stimare le autocovarianze sottostanti, che sono stimate automaticamente dalla formula dell’errore standard. Ecco i calcoli:

Errore standard clustered di \bar{Y} = $\sqrt{s_{\bar{Y}}^2 / n}$, dove

$$\begin{aligned}s_{\bar{Y}}^2 &= \frac{1}{n-1} \sum_{i=1}^n (\bar{Y}_i - \bar{Y})^2 \\&= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T Y_{it} - \bar{Y} \right)^2 \\&= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T (Y_{it} - \bar{Y}) \right)^2\end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T (Y_{it} - \bar{Y}) \right) \left(\frac{1}{T} \sum_{s=1}^T (Y_{is} - \bar{Y}) \right) \\
 &= \frac{1}{n-1} \sum_{i=1}^n \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T (Y_{is} - \bar{Y})(Y_{it} - \bar{Y}) \\
 &= \frac{1}{T^2} \sum_{i=1}^T \sum_{s=1}^T \left[\frac{1}{n-1} \sum_{j=1}^n (Y_{js} - \bar{Y})(Y_{ji} - \bar{Y}) \right]
 \end{aligned}$$

- Il termine finale tra quadre, $\frac{1}{n-1} \sum_{i=1}^n (Y_{is} - \bar{Y})(Y_{it} - \bar{Y})$, stima l'autocovarianza tra Y_{is} e Y_{it} . Quindi la formula dell'errore standard clustered implicitamente stima tutte le autocovarianze, usandole per stimare $\sigma_{\bar{Y}_i}^2$!
- Per contrasto, la formula "consueta" pone a zero queste autocovarianze omettendo tutti i termini misti – il che è valido solo se queste autocovarianze sono tutte zero.

Errori standard clustered per lo stimatore con effetti fissi nella regressione con dati panel

- Il concetto di errori standard clustered per dati panel è del tutto analogo al precedente caso della media per dati panel – solo più complesso per notazione e formule. Si veda l'Appendice 10.2.
- Gli errori standard clustered per dati panel sono l'estensione logica di quelli robusti all'eteroschedasticità per dati sezionali. Nella regressione con dati sezionali, gli errori standard robusti all'eteroschedasticità sono validi indipendentemente dal fatto che vi sia eteroschedasticità. Nella regressione con dati panel, gli errori standard clustered sono validi indipendentemente dal fatto che vi sia eteroschedasticità e/o correlazione seriale.
- *Tra l'altro...* Il termine “clustered” deriva dal fatto che si consente correlazione in un “cluster” (o “gruppo”) di osservazioni (in una entità) ma non tra cluster.

Perché i dati panel potrebbero aiutare?

- Potenziale distorsione da variabili omesse per variabili che variano tra stati ma sono costanti nel tempo:
 - cultura del bere e del guidare
 - qualità delle strade
 - età delle automobili sulle strade
 - usa effetti fissi di stato
- Potenziale distorsione da variabili omesse per variabili che variano nel tempo ma sono costanti tra stati:
 - miglioramenti nella sicurezza delle auto nel tempo
 - mutamento atteggiamenti verso la guida in stato di ebbrezza a livello nazionale
 - usa effetti temporali

Riepilogo: regressione con dati panel (Paragrafo 10.7)

Vantaggi e limitazioni della regressione con effetti fissi

Vantaggi

- Si può controllare per variabili non osservate che:
 - variano tra stati ma non nel tempo e/o
 - variano nel tempo ma non tra stati
- Più osservazioni forniscono più informazioni
- La stima coinvolge estensioni relativamente semplici della regressione multipla

- La regressione con effetti fissi si può eseguire in tre modi:
 1. Metodo “prima e dopo” quando $T = 2$
 2. “ $n-1$ regressori binari” quando n è piccolo
 3. Regressione “in deviazione dalle medie”
- Metodi simili si applicano alla regressione con effetti temporali e a quella con effetti fissi e temporali
- Inferenza statistica: come nella regressione multipla.

Limitazioni/problems aperti

- Necessaria la variazione in X nel tempo nelle entità
 - es. manovre covid
→ > 2 settimane
- Gli effetti di ritardo temporale possono essere importanti – anche se non ne abbiamo tenuto conto nel modello dell'imposta sulla birra
- È necessario usare errori standard clustered per evitare la possibilità che u_{it} sia autocorrelato

11/12 9:28

$\lambda_i = \text{effetto fisso} \cdot \text{cov}(\lambda_i, X_{it}) \neq 0$

$$\hat{\beta}_{i, \text{LSDV}} = \sum_i \sum_t \tilde{Y}_{it} \cdot X_{it} / \sum_i \sum_t \tilde{X}_{it} \cdot X_{it} = \frac{\sum_i \sum_t \tilde{Y}_{it} \tilde{X}_{it}}{\sum_i \sum_t \tilde{X}_{it}^2} = \hat{\beta}_{i, \text{wg}} = \hat{\beta}_{i, \text{FE}} \rightarrow \text{Fixed Effect}$$

$$\hat{\lambda}_{i, \text{LSDV}} = \bar{Y}_i = \hat{\beta}_{i, \text{LSDV}} \cdot \bar{X}_i = \bar{Y}_i - \hat{\beta}_{i, \text{wg}} \cdot \bar{X}_i = \hat{\lambda}_{i, \text{FE}}$$

$$\text{OLS } \tilde{Y}_{it} = \beta_i \cdot \tilde{X}_{it} + \tilde{\nu}_{it}$$

$$Y_{it} = \sum_j \lambda_j \cdot D_j + \beta_i X_{it} + \nu_{it}$$

$$\underline{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{bmatrix}$$

$$\underline{Y}_n = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{nT} \end{bmatrix}$$

$$\underline{\nu} = \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix}$$

$$\underline{\delta} = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_n \end{bmatrix}$$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$X = \begin{bmatrix} [1] & [0] & \cdots & [0] \\ [0] & [1] & \cdots & [0] \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \cdots & [1] \end{bmatrix}$$

$$\underline{Y} = X \cdot \underline{\delta} + \underline{\nu}$$

$$\hat{\underline{\delta}}_{\text{LSDV}} = (X'X)^{-1} X' \underline{Y}$$

$$\hat{\beta}_{i, \text{FE}}$$

$$\text{con } T=2$$

$$\hat{\beta}_{i, \Delta} = \hat{\beta}_{i, \text{wg}} = \hat{\beta}_{i, \text{LSDV}}$$

$$E[\nu_{it} | X_{i1}, X_{i2}, \lambda_i] = 0 \quad \stackrel{(1, T \rightarrow +\infty)}{\circ}$$

$\hat{\beta}_{i, \text{FE}}$ è consistente e asintoticamente normale

Se $\nu_{it} \sim i.i.d.$ allora $\hat{\beta}_{i, \text{FE}}$ è corretto e inoltre $\hat{\lambda}_{i, \text{FE}}$ è corretto

($\forall i, t$)

Se n fix. e $T \rightarrow +\infty$ $\hat{\beta}_{i,FE}$ sono consistenti.

$$\lambda_i = \sum_{j=1}^k \tilde{\beta}_j \cdot z_{ji}$$

$$(\lambda_i + \lambda_t + \mu_{it})$$

$$\lambda_t = \sum_{h=1}^m S_h \cdot z_{ht} \rightarrow \text{Effetto temporale fisso: } \text{cov}(\lambda_t, x_{it}) \neq 0$$

$$Y_{it} = \beta_1 X_{it} + (\lambda_i + \lambda_t + \mu_{it}) \\ = \lambda_i + \lambda_t + \beta_1 X_{it} + \nu_{it}$$

$$\lambda_t = \sum_{t=1}^T \lambda_t (\beta_t)_t \quad (\beta_t)_t < 0 \quad \text{altrimenti:}$$

$$I) E[\nu_{it} | \lambda_i, \lambda_t, X_{i1}, \dots, X_{iT}] = 0$$

Molto spesso tutto le variabili temporali come dummy perché sono poche generalmente (< 10)

Quindi:

$$Y_{it} = \beta_0 + \sum_{s=1}^T \lambda_s (\beta_s)_t + \beta_1 \quad \| 10:10 \quad 1/12 \quad \text{righe}$$

$$(Y_{it} - \bar{Y}_t) = \beta_1 (X_{it} - \bar{X}_t) + \dots$$

$$Y_{it} = \beta_0 + \sum_{j=2}^n \gamma_j (D_j)_i + \sum_{s=2}^T \tilde{\lambda}_s (\beta_s)_t + \beta_1 X_{it} + \nu_{it}$$

Di solito usare strategie miste:

$$\tilde{Y}_{it} = (Y_{it} - \bar{Y}_i) = \sum_{s=2}^T \tilde{\lambda}_s \cdot (\tilde{\beta}_s)_t + \beta_1 \tilde{X}_{it} + \tilde{\nu}_{it}$$

$\left(\begin{array}{l} n \rightarrow +\infty \\ T \text{ fix.} \end{array} \right) \hat{\lambda}_{s,FE}$ sono consistenti. $\hat{\beta}_{i,FE}$ è consistente e asintoticamente normale.

λ_i = effetti casuali (random effects)

ipotesi: $\text{Cov}(\lambda_i, X_{it}) = 0 \rightarrow \beta_0 + \gamma_i = \lambda_i$

1) $E[\nu_{it} | \lambda_i, X_{it}, \dots, X_{iT}] = 0 \rightarrow \text{Cov}(\nu_{it}, X_{it}) = 0$ l'errore non è correlato con il regressore. $V_{it} = (\lambda_i + \nu_{it})$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \nu_{it}$$

POLS è consistente per β_0 e β_1 ed è anche corretto per β_1 e asintoticamente normale

$\hat{\beta}_{1,\text{POLS}} \xrightarrow{P} \beta_1$ anche se le X_{it} è debolmente esogena

1) $E[V_{it} | X_{it}, X_{it-1}, \dots, X_{i1}] = 0$

$$\text{Cov}(V_{it}, X_{it-s}) = 0, s > 0$$

OLS applicato ad ogni cross-section separata è consistente

Stimatore between-group (BG)

$$\bar{Y}_i = \beta_0 + \beta_1 (\bar{X}_i) + \bar{\nu}_i \quad i=1, \dots, n$$

$\text{Cov}(\bar{X}_i, \bar{\nu}_i) = 0$ se c'è stretta esogenità $\rightarrow \hat{\beta}_{1,BG} \xrightarrow{P} \beta_1$ è consistente ma asintoticamente inefficiente

Stimatore within-group (WG)

$\hat{\beta}_{1,WG} \xrightarrow{P} \beta_1$ stimatore WG è consistente

X_{it} è strettamente esogeno

I migliori stimatori sono POLS e WG però per WG X_{it} dev'essere strettamente esogeno

λ_i random

$$Y_{it} = (\beta_0 + \gamma_i) + \beta_1 X_{it} + \nu_{it}$$

$$E[\gamma_i] = 0 \quad \text{Var}[\gamma_i] = \sigma_\gamma^2 K_i \quad \text{e} \quad \gamma_i \sim \text{i.i.d.}(0, \sigma_\gamma^2) \quad \text{e} \quad \nu_{it} \sim \text{i.i.d.}(0, \sigma_n^2) \quad \text{e} \quad (i, t)$$

4/12

Continuazione modello Panel a effetti Random

Strettamente esogeno γ_i

$$Y_{it} = \beta_0 + \beta_1 X_{it} + V_{it}$$

$$\left. \begin{array}{l} \text{con } V_{it} = \bar{\gamma}_i + \mu_{it} \\ \bar{\gamma}_i \sim \text{i.i.d. } (0, \sigma_{\gamma}^2) \\ \mu_{it} \sim \text{i.i.d. } (0, \sigma_{\mu}^2) \end{array} \right\} \forall i, t$$

$$\bar{\gamma}_i \perp \mu_{is}, \forall i, s \quad \mu_{it} \sim \text{i.i.d. } (0, \sigma_{\mu}^2)$$

Ipotesi restrittive ma ci permette di usare lo \rightarrow

stimate GLS anziché il POLS per avere stime più affidabili

 \hookrightarrow Metodo dei LS pesati \rightarrow BLUE

$$E[V_{it}|X_{i1}, \dots, X_{iT}] = 0$$

$$E[\mu_{it}|X_{i1}, \dots, X_{iT}] = 0$$

$$E[\bar{\gamma}_i|X_{i1}, \dots, X_{iT}] = 0$$

$$\text{Cov}(\bar{\gamma}_i, X_{it}) = 0 \rightarrow E[\bar{\gamma}_i] = 0$$

 \hookrightarrow Necessaria per la consistenza

$$E[V_{it}] = 0$$

$$\text{Var}[V_{it}] = \sigma_{\gamma}^2 + \sigma_{\mu}^2 \quad \forall i, t$$

$$\text{Se } i \neq j, \forall t, s \quad \text{Cov}(V_{it}, V_{js}) = 0$$

$$\text{Se } i = j, \forall t \neq s \quad \text{Cov}(V_{it}, V_{js}) = \text{Cov}(\bar{\gamma}_i, \bar{\gamma}_j) = \text{Var}(\bar{\gamma}_i) = \sigma_{\gamma}^2 \geq 0$$

in questo caso $\bar{\gamma}_i = \bar{\gamma}_j$

$$\left. \begin{array}{l} V_{js} = \bar{\gamma}_j + \mu_{js} \\ V_{it} = \bar{\gamma}_i + \mu_{it} \end{array} \right.$$

Gli errori dello stesso individuo
nel tempo saranno correlati:
La matrice

$$\underline{Y} = X \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \underline{V}$$

$\underline{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix}$

$$\text{Var}(\underline{V}_i) = \begin{bmatrix} (\sigma_{\mu}^2 + \sigma_{\gamma}^2), & \sigma_{\gamma}^2, & \dots, & \sigma_{\gamma}^2 \\ \sigma_{\gamma}^2, & \ddots, & & \vdots \\ \vdots, & \ddots, & \ddots, & (\sigma_{\mu}^2 + \sigma_{\gamma}^2) \end{bmatrix} = \Omega = \sigma_{\mu}^2 \cdot I_T + \sigma_{\gamma}^2 \underbrace{11' - 11'}_{T \times 1 \quad T \times 1}$$

$$\tilde{\Omega} = \text{Var}(\underline{V}) = \begin{bmatrix} \Omega & [0] & \dots & [0] \\ [0] & \Omega & \dots & ; \\ \vdots & \vdots & \ddots & ; \\ [0] & \dots & \dots & \Omega \end{bmatrix} = I_n \otimes \Omega$$

Prendo tutti i vettori

\hookrightarrow Se non ci sono effetti individuali il matrice è una matrice scalare e posso usare il POLS perché è BLUE, altrimenti è una matrice a blocchi e devo usare l'estimatore Random Effects che si basa sui GLS ed è POLS

$$\underline{Y} = X \underline{\beta} + \underline{V}$$

$$\text{Var}(\underline{V}) = \tilde{\Omega} \text{ def. positive note}$$

$$\exists P \text{ invertibile t.c. } P \tilde{\Omega} P^T = \sigma_N^2 \cdot I_{NT} \rightarrow \tilde{\Omega} = \sigma_N^2 \cdot P^{-1}(P^T)^T$$

→ La P non è unica

$$\tilde{\Omega}^{-1} = \frac{1}{\sigma_N^2} \cdot P^T P = P^T P = \sigma_N^2 \cdot \tilde{\Omega}^{-1}$$

$$P\underline{Y} = P X \underline{\beta} + P \cdot \underline{V}$$

$$\underline{Y}^* = X^* \underline{\beta} + \underline{V}^*, \text{ con } \text{Var}(\underline{V}^*) = \sigma_N^2 \cdot I_{NT}$$

OLS

$$\hat{\underline{\beta}}_{OLS} = \hat{\underline{\beta}}_{RE} = (X^{*T} X^*)^{-1} X^{*T} \underline{Y}^* = (X^T (P^T P) X)^{-1} X^T (P^T P) \underline{Y} = (X^T \tilde{\Omega}^{-1} X)^{-1} X^T \tilde{\Omega}^{-1} \underline{Y}$$

$\tilde{\Omega}$ è diagonale a blocchi → P diagonale a blocchi

$$P = \begin{bmatrix} P^0 & \dots & P^T \end{bmatrix}$$

$$P^T \tilde{\Omega} P^T = \sigma_N^2 \cdot I_T$$

Quasi deviazione?

$$(Y_{it} - \bar{Y}_i)$$

$$\sum_{t=1}^T P^T Y_i = P^T X_i \underline{\beta} + P^T V_i$$

$$Y_{it} = \beta_0 + \beta_1 x_{it} + v_{it} \quad | \quad (Y_{it} - \bar{Y}_i) = \beta_0 (1-\theta) + \beta_1 (x_{it} - \theta \bar{x}_i) + (v_{it} - \theta \bar{v}_i)$$

OLS

Devo ancora scoprire di θ v_{it} n.i.d. $(0, \sigma_v^2)$

$$\hat{\theta} = 1 - \hat{\Psi}^{1/2}$$

$$\tilde{\Omega} \rightarrow \hat{\Omega}$$

$$\hat{\Psi} = \frac{\hat{\sigma}_N^2}{\hat{\sigma}_N^2 + T \hat{\sigma}_\theta^2}$$

GLS

$$\min_{\{\underline{\beta}\}} \underline{V}^T \underline{V} = \min_{\{\underline{\beta}\}} \underline{V}^T P^T P \underline{V}$$

$$\underline{V}^* = P \underline{V}$$

$$= \min_{\{\underline{\beta}\}} \underline{V}^T \tilde{\Omega}^{-1} \underline{V}$$

$$(P^T P) \propto \tilde{\Omega}^{-1}$$

$$\tilde{v}_{it} = v_{it} - \bar{v}_{it} \rightarrow \tilde{Y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{v}_{it}$$

$$\frac{\sigma_N^2}{\sigma_N^2 + T \hat{\sigma}_\theta^2}$$

$$\hat{\sigma}_N^2 \xrightarrow{n \rightarrow \infty} \sigma_N^2$$

T fix

$$\rightarrow Y_{it} = \alpha_i + \beta_1 x_{it} + v_{it} \xrightarrow{\sim \text{iid}(0, \sigma_v^2)} \rightarrow Y_{it} - \hat{\alpha}_i - \hat{\beta}_1 x_{it} = \hat{v}_{it} - \tilde{v}_{it}$$

WG-FE Vado a usare lo stimatore Within group perché è consistente

Continue →

$$\hat{\sigma}_\nu^2 = \frac{\sum_i \sum_t \hat{\nu}_{it}^2}{n \cdot T - (n+1)}$$

Stimatore BG

$$\forall i=1, \dots, n \quad \bar{Y}_i = \beta_0 + \beta_1 \bar{X}_i + e_i$$

$\frac{1}{T} \sum_{t=1}^T v_{it} = e_i \rightarrow$ Media nel tempo comprende γ e ν
 \rightarrow distribuiti uguali e non correlati tra loro

$$\sigma_e^2 = \text{Var}(e_i) = \hat{\sigma}_y^2 + \frac{1}{T} \hat{\sigma}_\nu^2$$

$$T \cdot \hat{\sigma}_e^2 = (\hat{\sigma}_y^2 + T \cdot \hat{\sigma}_\nu^2)$$

BG è consistente perché \times è strettamente esogena

RE = Stimatore di Boestra - Nohel

Quando non abbiamo effetti individuali

$$\hat{\sigma}_y^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_\nu^2 + T \cdot \hat{\sigma}_\nu^2}$$

Lo stimatore Random Effect è uguale al WG quando $\hat{\theta} = 0$ ($\psi = 1$) ovvero

RE e WG coincidono quando $T \rightarrow +\infty$

1) Supponiamo che $\gamma_i \sim \text{RE}$ ma che ν_{it} sia eterosched./autocorrelato

$\nu_{it} \sim \text{iid}(\sigma_\nu^2)$ non vale più

2) $(X_{i1}, \dots, X_{iT}, \nu_{i1}, \dots, \nu_{iT}) \sim \text{iid } f_i$

$\text{Var}(V) \neq \Sigma \rightarrow \hat{\theta}$

$$\text{POLS } (U_{it} - \hat{\theta} \bar{Y}_i) = \tilde{\beta}_0 + \beta_1 (X_{it} - \hat{\theta} \bar{X}_i) + W_{it}$$

$$\begin{array}{c} \xrightarrow{\text{non correlati}} \\ \downarrow \\ V_{it} - \hat{\theta} \bar{Y}_i \end{array}$$

non sono correlati \rightarrow RE consistente e efficiente

(2) $\gamma_i \sim RE$, però X_{it} è debolmente esogene. $Cov(\nu_{it}, X_{i,t-s}) = 0 \quad s \geq 0$

$$\begin{cases} Y_t = \beta_0 + \beta_1 X_t + \nu_t \\ X_t = \gamma_0 + \gamma_1 Y_{t-1} + v_t \end{cases}$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + (\gamma_1 + \nu_{it})$$

Debolmente esogene
↳ RE

POLS è consistente
S.E. HAC

$\left\{ \begin{array}{l} \text{Se } X_{it} \text{ è endogene rispetto a } \nu_{it} \\ \gamma_i \sim RE \end{array} \right.$ $Cov(X_{i,t-s}, \nu_{it}) = 0 \quad \forall s > 0$

\hookrightarrow WG, RE, POLS sono inconsistenti \rightarrow Devo usare il metodo delle variabili strumentali (I.V.)

$$Z_{it} = X_{i,t-1} \quad \forall s > 0$$

Stime $t = 2, \dots, T$

\hookrightarrow Perdo un'osservazione

5/12

POLS

S.E. HAC

$$(X_{i1}, \dots, X_{iT}, \nu_{i1}, \dots, \nu_{iT}) \sim i.i.d. f_i$$

$$Y_{it} = \mu + \nu_{it}$$

$$\hat{\mu}_{\text{POLS}} = \frac{1}{n \cdot T} \sum_i \sum_t Y_{it} = \mu + \frac{1}{n \cdot T} \sum_i \sum_t \nu_{it}$$

$$\text{Var}(Y_{it}) = \text{Var}(\nu_{it})$$

$$\text{Cov}(Y_{it}, Y_{is}) = \text{Cov}(\nu_{it}, \nu_{is}) \neq 0$$

Assunzioni del modello

$$\begin{cases} E[\nu_{it}] = 0 \rightarrow E[\bar{\nu}_{it}] = 0 \rightarrow E[\nu_{i:}] = 0 \\ \sigma_{\nu_{i:}}^2 = \sigma_{\eta_i}^2 = \text{Var}(\nu_{i:}) = T \cdot \text{Var}(\bar{\nu}_i) \\ \eta_i \perp \eta_j \quad \eta_i \sim i.i.d. (0, \sigma_{\eta_i}^2) \end{cases}$$

$$\sqrt{n} \cdot \sqrt{T} (\hat{\mu} - \mu) = \frac{1}{\sqrt{n} \cdot T} \sum_i \sum_t \nu_{it} \sqrt{n} \cdot \sqrt{T}$$

$$= \sqrt{n} \cdot \frac{1}{n} \sum_i \sqrt{T} \frac{1}{T} \sum_t \nu_{it}$$

$\eta_i \rightarrow$ posso applicare il TLC

$$= \sqrt{n} \cdot \frac{1}{n} \sum_i (\sqrt{T} \cdot \bar{\nu}_i)$$

$$= \sqrt{n} \cdot \frac{1}{n} \sum_i \eta_i \xrightarrow[n \rightarrow \infty]{\text{CLT}} N(0, \sigma_{\eta_i}^2)$$

$$= \frac{1}{n} \sum_i \eta_i \xrightarrow[n \rightarrow \infty]{} N(0, \frac{\sigma_{\eta_i}^2}{n})$$

Dobbiamo calcolare questo stimatore

$$\eta_i = \sqrt{T} \cdot \bar{\nu}_i \quad \eta_i \perp \eta_j \quad i=1, \dots, n \quad n \gg 0 \quad \eta_i \sim \text{iid}$$

$$\hat{\sigma}_{\eta}^2 = \left(\frac{1}{n}\right) \sum_i \eta_i^2$$

$$\hat{\sigma}_{\hat{\eta}}^2 = \left(\frac{1}{n}\right) \sum_i \hat{\eta}_i^2$$

$$\text{con } \hat{\eta}_{it} = \sqrt{T} \cdot \frac{1}{T} \sum_t \hat{\mu}_{it} \neq 0 \\ \text{e } \sum_t \sum_i \hat{\mu}_{it} = 0$$

I residui vengono stima di un parametro, quindi:

$$S_{\hat{\eta}}^2 = \left(\frac{1}{n-1}\right) \sum_i \hat{\eta}_i^2 \rightarrow \text{Varianza campionaria}$$

$$\begin{aligned} \sigma_{\eta}^2 &= T \cdot \text{Var}(\bar{\nu}_i) = T \cdot \text{Var}\left(\frac{1}{T} \sum_t \nu_{it}\right) \\ &= T \cdot \frac{1}{T^2} \text{Var}(\nu_{i1} + \dots + \nu_{iT}) \\ &= \frac{1}{T} \sum_t \sum_s \text{Cov}(\nu_{it}, \nu_{is}) \end{aligned}$$

(HAC)

$$\begin{aligned} S_{\hat{\eta}}^2 &= \left(\frac{1}{n-1}\right) \sum_i \hat{\eta}_i^2 = \frac{1}{n-1} \sum_i T \left(\frac{1}{T} \sum_t \hat{\mu}_{it}\right)^2 \\ &= \frac{1}{n-1} \sum_i \frac{T}{T^2} \left(\sum_t \hat{\mu}_{it}\right)^2 = \frac{1}{n-1} \cdot \sum_i \left[\frac{1}{T} \sum_t \sum_s \hat{\mu}_{it} \cdot \hat{\mu}_{is} \right] \\ &= \left(\frac{1}{T}\right) \sum_t \sum_s \underbrace{\left(\frac{1}{n-1}\right) \sum_i \hat{\mu}_{it} \hat{\mu}_{is}}_{\text{Cov}(\nu_{it}, \nu_{is})} \end{aligned} \rightarrow \text{Tengo conto di tutte le covarianze della matrice}$$

Lo S.E. ($\hat{\beta}$) = $\frac{S_{\hat{\eta}}}{\sqrt{n}}$ è robusto perché sta già implicitamente stimando le varianze e cov.

Per l'elenco non servono i conti, basta sapere che è necessario lavorare solo sugli η_i .

GRETl 10:10 5/12

$$\hat{\eta} = \hat{\beta}_{WG} - \hat{\beta}_{RE} \quad \text{Var}(\hat{\eta}) \rightarrow \text{matrice asint.}$$

$$\rightarrow \text{Test: } \hat{\eta}' (\text{Var}(\hat{\eta}))^{-1} \hat{\eta} \sim \chi^2(k)$$

$2_i \sim FE$

$$Y_{it} = \lambda_t + \alpha_i + \beta_i X_{it} + \nu_{it}$$

X_{it} sia debolmente esogeno $\text{Cov}(X_{i,t-s}; \nu_{it}) = 0 \quad \forall s \geq 0$

Non posso usare direttamente il WG
Posso usare le differenze prime anche
se è inefficiente perché perdo la prima
osservazione

$$t=2, \dots, T \quad \Delta Y_{it} = (\lambda_t - \lambda_{t-1}) + \alpha_i + \beta_i \Delta X_{it} + \Delta \nu_{it} \quad \leftarrow \downarrow$$

$$\Delta Y_{it} = \beta_0 + \sum_{j=3}^T \gamma_j \cdot (\beta_j)_{it} + \beta_i \Delta X_{it} + \Delta \nu_{it}$$

Per adesso ipotesi che non ci siano

$$\Delta Y_{it} = \beta_i \Delta X_{it} + \Delta \nu_{it}$$

I.V.

Possiamo non considerarli corrett.? No, perché non abbiamo assunto che
sia strett. esogeno. $\text{Cov}(\Delta X_{it}, \nu_{it}) \neq 0$

$$Z_{it} = X_{i,t-1}$$

Perché $\text{Cov}(X_{i,t-1}; \Delta \nu_{it}) = 0$

$$\text{Cov}(X_{i,t-1}; \Delta X_{it}) \neq 0?$$

$$\text{Cov}(\Delta X_{i,t-1}, \Delta \nu_{it}) = 0$$

$$= \text{Cov}(\Delta X_{it}, \nu_{it}) - \text{Cov}(\Delta Z_{it}, \nu_{it})$$

$$\Delta X_{i,3}$$

$$\Delta X_{i,2}$$

Si preferisce a . Caso dell'essotta identificazione

11/12

Generalized Moment Method

Stimatore GMM di Hansen

Metodo delle variabili strumentali per l'identificazione

Metodo GIV di Sargan

Metodo delle variabili strumentali generalizzato

Stesso approccio del 2SLS (2 stage least square)

Siamo nell'ambito della sovraidentificazione

Es. modello su dati cross-section

$$Y_i = \beta_0 + \beta_1 X_i + \nu_i$$

$$\text{Cov}(X_i, \nu_i) \neq 0$$

Condizioni:

$$1. E[X_i \cdot \nu_i] = E[\nu_i] = 0 \rightarrow \text{C'è esatta identificazione}$$

$$2. E[Z_{1i} \cdot \nu_i] = 0 \rightarrow \text{Orthogonalità}$$

Con il metodo dei momenti usiamo il principio di Pearson che dice che possiamo sostituire i momenti delle variabili strumentali con i momenti campionari:

$$\text{RM} \begin{cases} \frac{1}{n} \sum_i \nu_i = 0 \\ \frac{1}{n} \sum_i Z_{1i} \nu_i = 0 \end{cases} \rightarrow 1 \text{ sol. unica}$$

$$\hat{\beta}_{0,\text{RM}} = \hat{\beta}_{0,\text{IV}} = \hat{\beta}_{0,\text{2SLS}}$$

$$\hat{\beta}_{1,\text{RM}} = \hat{\beta}_{1,\text{IV}} = \hat{\beta}_{1,\text{2SLS}}$$

2 strumenti > 1 regressore endogeno \rightarrow C'è sovraidentificazionee supponiamo per comodità $\beta_0 = 0$ Z_{1i} e Z_{2i} strumenti

$$\begin{cases} E[Z_{1i} \cdot \nu_i] = 0 \\ E[Z_{2i} \cdot \nu_i] = 0 \end{cases} \rightarrow \begin{cases} \frac{1}{n} \sum_i Z_{1i} \cdot \nu_i = 0 \\ \frac{1}{n} \sum_i Z_{2i} \cdot \nu_i = 0 \end{cases}$$

non esiste sol. unica

GLV

$$\begin{cases} m_1(\hat{\beta}_1) = 0 \\ m_2(\hat{\beta}_2) = 0 \end{cases} \text{ non è possibile con GLV}$$

In termini vettoriali devo arrivare il più possibile $\underline{m}(\beta) \rightarrow 0$

$$\underline{m}(\beta_i) = \begin{bmatrix} m_1(\beta_1) \\ m_2(\beta_2) \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\min_{\{\beta_i\}} \underline{m}(\beta_i)' \cdot I_2 - \underline{m}(\beta_i) = \min_{\{\beta_i\}} m_1(\beta_1)^2 + m_2(\beta_2)^2$$

Peso 1 a entrambi i momenti quadratici

↳ Logica OLS

$$\rightarrow A = I_2$$

WLS - Weighted Least Square

$$\min_{\{\beta_i\}} \alpha_{11}(m_1(\beta_1))^2 + \alpha_{22}(m_2(\beta_2))^2 \rightarrow A = \text{diag}(\alpha_{11}, \alpha_{22})$$

GLS - Generalized Least Square

$$\min_{\{\beta_i\}} \underline{m}(\beta_i)' A \cdot \underline{m}(\beta_i) = \min_{\{\beta_i\}} \alpha_{11}(m_1(\beta_1))^2 + \alpha_{22}(m_2(\beta_2))^2 + 2\alpha_{12} \cdot m_1(\beta_1) \cdot m_2(\beta_2)$$

$S(\beta_i)$ } } }
A def. positiva simmetrica

Sargan $\nu_i \sim \text{iid}(0; \sigma_u^2) \rightarrow$ Considero omoschedasticità

$$\min_{\{\beta_i\}} S(\beta_i) \rightarrow \frac{\partial S(\beta_i)}{\partial \beta_i} = 0$$

$$S(\beta_i) \geq 0 \quad \hat{\beta}_{i, \text{GLV}} \quad A = I \quad (m_1(\beta_i) - 0)^2 + (m_2(\beta_i) - 0)^2$$

• Se c'è sovraidentificazione $S(\hat{\beta}_{i, \text{GLV}}) > 0$

• Se c'è esatta identificazione $\hat{\beta}_{i, \text{GLV}} = \hat{\beta}_{i, \text{IV}} = \hat{\beta}_{i, \text{ML}}$ e $S(\hat{\beta}_{i, \text{GLV}}) = 0$

• A è ininfluenza $\rightarrow \forall A \rightarrow \hat{\beta}_{i, \text{GLV}} = \hat{\beta}_{i, \text{ML}}$ owo l'uso delle matrice dei pesi non cambia il risultato

Con sovridentificazione

1) $\hat{\beta}_{1,GIV}$ è consistente e asintotica normalmente $\forall A$

Però la stima $\hat{\beta}_{1,GIV}$ dipende da A

2) $\exists A$ ottimale che dà luogo ad uno stimatore GIV anche asintoticamente efficiente

$\rightarrow A^{\text{ottimale}} \subset H^{-1}$ con H matrice di varianze covarianze asintotiche dei momenti campionari
 $H = \text{AVAR}(\underline{m}(\beta_1))$

$$\begin{aligned}\hat{H} &\xrightarrow{\text{def}} H & \min_{\{B_1\}} \underline{m}(\beta_1)' \hat{A} \underline{m}(\beta_1) \\ \hat{A} &\propto \hat{H}^{-1} & \hat{A} \xrightarrow{\text{def}} A\end{aligned}$$

$$\text{Var}(\nu_i | Z_{ii}, Z_{2i}) = \sigma_\nu^2 \quad E[\underline{\nu} \cdot \underline{\nu}'] = \text{Var}(\underline{\nu})$$

$$\frac{1}{n} \sum_i \nu_{1i} = 0 \quad \nu_{1i} = Z_{ii} \cdot \nu_i$$

$$\frac{1}{n} \sum_i \nu_{2i} = 0 \quad \nu_{2i} = Z_{2i} \cdot \nu_i$$

$$H = \sigma_\nu^2 \cdot \Sigma_z \quad H^{-1} = \frac{1}{\sigma_\nu^2} \cdot \Sigma_z \quad A = \Sigma_z \quad \hat{A} = \left(\frac{1}{n} Z' Z \right)^{-1} \quad \text{Stimatore di Sargan}$$

GIV $\hat{\beta} = Z'Z$ è lo stimatore asint. efficiente \rightarrow 2SLS è lo stimatore asintot. efficiente sotto l'ipotesi di omoschedasticità

Stimatore GMM ottimale

1) sceglie una matrice dei pesi A ad esempio $A = I$

Trovò lo stimatore GMM₁ del primo peso che è consist. e asint. normale

2) $\hat{\beta}_{1,GMM} \rightarrow \hat{H} \quad \hat{H} \xrightarrow{\text{def}} H \quad \xrightarrow{\text{def}} \hat{A}$

$$\text{GMM}_2 \quad \min_{\{B_1\}} \underline{m}(\beta_1)' \hat{A} \underline{m}(\beta_1)$$

Test J di Sargan

Applicabile solo a un stimatore assintotici efficienti

Ipotizziamo heteroschedasticità

$$n \cdot S(\hat{\beta}_1) \underset{H_0}{\approx} \chi^2(2-1) = \chi^2_1$$

\hookrightarrow Regressore endogeno
 \hookrightarrow Strumenti

Tendenzialmente gli s.e. sono più piccoli di quello che sono realmente

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \mu_{it}$$

$$X_{it} \text{ deb. esogeno} \quad \text{Cov}(X_{i,t-s}, \mu_{it}) = 0 \quad \forall s > 0$$

$$\Delta Y_{it} = \beta_1 \Delta X_{it} + \Delta \mu_{it} \quad t = 2, 3, 4 \quad \text{per } x_{it} \text{ ipotesi}$$

$$Z_{1t} = X_{i,t-1} \quad E[X_{i,t-1} \cdot \Delta \mu_{it}] = 0$$

$$Z_{2t} = X_{i,t-2} \quad E[X_{i,t-2} \cdot \Delta \mu_{it}] = 0$$

⋮

$$Z_{jt} = X_{i,t-j} \quad E[X_{i,t-j} \cdot \Delta \mu_{it}] = 0 \quad \forall j \geq 1$$

Strumenti esogeni e forti

Se disponibili "sovridentif."

$$t=2 \quad \Delta Y_{i2} = \beta_1 \Delta X_{i2} + \Delta \mu_{i2}$$

$$Z_{1i2} = X_{i,1}$$

$$m_1(\beta_1) = \frac{1}{n} \sum_i X_{i,1} \cdot \Delta \mu_{i2}$$

$$t=3 \quad \Delta Y_{i3} = \beta_1 \cdot \Delta X_{i3} + \Delta \mu_{i3}$$

$$\begin{pmatrix} X_{i,2} \\ X_{i,3} \end{pmatrix}$$

$$m_2(\beta_1) = \frac{1}{n} \sum_i X_{i,2} \cdot \Delta \mu_{i3}$$

$$t=4 \quad \Delta Y_{i4} = \beta_1 \cdot \Delta X_{i4} + \Delta \mu_{i4}$$

$$\begin{pmatrix} X_{i,3} \\ X_{i,2} \\ X_{i,1} \end{pmatrix}$$

$$m_3(\beta_1) = \frac{1}{n} \sum_i X_{i,3} \cdot \Delta \mu_{i4}$$

$$m_4(\beta_1) = \frac{1}{n} \sum_i X_{i,2} \cdot \Delta \mu_{i4}$$

$$m_5(\beta_1) = \frac{1}{n} \sum_i X_{i,1} \cdot \Delta \mu_{i4}$$

$$m_6(\beta_1) = \frac{1}{n} \sum_i X_{i,1} \cdot \Delta \mu_{i4}$$

GMM 1

$$\min_{\{\beta_1\}} m(\beta_1)' A m(\beta_1) \quad \text{con } A = I_6$$

GMM 2

Nel caso di endogeneità

$$\Delta Y_{it} = \beta_1 \Delta X_{it} + \Delta u_{it}$$

X_{it} endogeno rispetto a u_{it} strumento

→

$$\begin{pmatrix} X_{i,t-2} \\ X_{i,t-3} \\ \vdots \end{pmatrix}$$

non sono correlati con Δu_{it}

↔

Possiamo fare le stime ma dobbiamo rinunciare
a un tempo
per $t=2$ non abbiano strumenti

library (PLR) per manipolare dati panel in R
Arellano and Bond

FINE