

1 REGRESSIONE LINEARE

MODELLO E STIMA

La regressione lineare rappresenta un metodo di stima del valore atteso condizionato di una variabile dipendente, o endogena, Y , dati i valori di altre variabili indipendenti, o esogene, X_1, \dots, X_k . La regressione è un modello statistico dove si assume un legame tra variabile endogena e variabili esogene di tipo lineare.

NOTAZIONE MATRICIALE $\rightarrow \underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, NOTAZIONE VETTORIALE $\rightarrow y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i$

TEOREMA DI GAUSS-MARKOV

In un modello lineare in cui i termini di errore (supposti indipendenti dalle variabili esogene) hanno valore atteso nullo, sono incorrelati e omoschedastici, gli stimatori lineari corretti più efficienti sono gli stimatori ottenuti con il metodo dei minimi quadrati.

$$\text{IPOTESI CLASSICHE GAUSS MARKOV} \left\{ \begin{array}{l} \text{A1) } E(\varepsilon_i) = 0, \quad i = 1, \dots, N \\ \text{A2) } \{\varepsilon_1, \dots, \varepsilon_N\} \text{ e } \{\underline{x}_1, \dots, \underline{x}_N\} \text{ indipendenti (o x deterministica)} \\ \text{A3) } V(\varepsilon_i) = \sigma^2, \quad i = 1, \dots, N, \quad \text{omoschedasticità} \\ \text{A4) } \text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i, j = 1, \dots, N (i \neq j), \quad \text{no autocorrelazione} \end{array} \right.$$

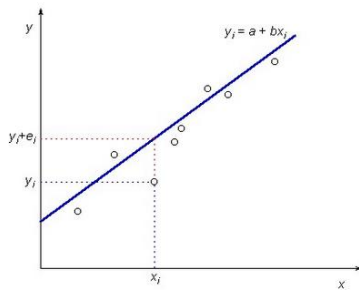
CON IPOTESI GAUSS MARKOV \underline{b} E' BLUE (Best Linear Unbiased Estimatori)

STIMATORE MINIMI QUADRATI (OLS=ORDINARY LEAST SQUARES)

$$\left\{ \begin{array}{l} \underline{b} = \text{STIMATORE MINIMI QUADRATI MATRICE} = (X'X)^{-1}X'y \\ \underline{b} = \text{VETTORE} = \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^N \underline{x}_i y_i \\ \text{REGRESSIONE SEMPLICE} \left\{ \begin{array}{l} b_1 = \bar{y} - b_2 \bar{x} \\ b_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\sigma(x, y)}{\sigma^2(x)} \end{array} \right. \end{array} \right.$$

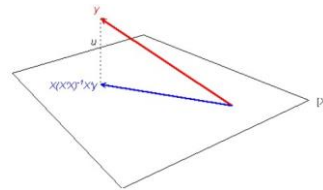
RESIDUI

$$\text{MODELLO STIMATO} \rightarrow \left\{ \begin{array}{l} \hat{y}_i = \underline{x}_i' \underline{b} \\ \underline{\hat{y}} = \underline{X}\underline{b} \end{array} \right., \quad \text{RESIDUI} \rightarrow \left\{ \begin{array}{l} e_i = y_i - \hat{y}_i = y_i - \underline{x}_i' \underline{b} \\ \underline{e} = \underline{y} - \underline{\hat{y}} = \underline{y} - \underline{X}\underline{b} \end{array} \right.$$



Il metodo dei minimi quadrati, nel caso bivariato, deriva una retta che interpola uno scatter di punti minimizzando la somma dei quadrati delle distanze ε_i dei punti stessi dalla retta; il grafico fornisce un'intuizione del procedimento. La scelta di minimizzare i quadrati degli ε_i non è arbitraria. Se si minimizzasse la somma degli ε_i , distanze positive e negative si compenserebbero, rendendo peggiore la qualità dell'interpolazione; se si adottasse una funzione criterio uguale alla somma dei valori assoluti degli ε_i , non essendo la funzione valore assoluto derivabile su tutto l'asse reale, non si potrebbe ricorrere all'elegante metodo di minimizzazione sopra illustrato.

INTERPRETAZIONE GEOMETRICA DELLE STIME OLS



Il vettore di stime OLS $\underline{\hat{y}}$ consente di ottenere i valori previsti ("teorici") per la variabile dipendente $\underline{\hat{y}} = X\underline{\hat{b}} = X(X'X)^{-1}X'y$. Formalmente l'espressione corrisponde a una proiezione ortogonale del vettore delle osservazioni \underline{y} sullo spazio generato dalle colonne della matrice X .

MULTICOLLINEARITA'

Se due o più colonne della matrice dei regressori X sono linearmente dipendenti, non esiste l'inversa $(X'X)^{-1}$ per cui il vettore di stime OLS non può essere determinato. Se da un lato è assai improbabile che questa eventualità si verifichi nelle applicazioni pratiche, è comunque possibile che alcune colonne della matrice X siano prossime alla dipendenza lineare; in tal caso sarà ancora possibile ottenere un vettore di stime OLS, ma sorgerà il problema della multicollinearità.

Si parla di multicollinearità allorché una o più colonne della matrice dei regressori X sono prossime a essere linearmente dipendenti (correlate). L'effetto della multicollinearità è che la matrice $(X'X)^{-1}$ è prossima all'essere singolare. Questo ha due conseguenze di particolare rilievo nelle applicazioni:

- la significatività statistica dei singoli coefficienti risulta modesta (non si riesce a distinguere il contributo informativo delle variabili e le procedure diventano scarsamente significative);
- il fitting della regressione risulta elevato (si osservano elevati valori dell'indice R^2).

$$VIF_i = \text{Variance Inflation factor} = \frac{1}{1 - R_i^2} \rightarrow \begin{cases} R_i^2 \text{ di una regressione sulla colonna } i \text{ e gli altri regressori} \\ \text{Più } R_i^2 \text{ elevato più } VIF_i \text{ elevato} \rightarrow \text{maggiore multicollinearità} \end{cases}$$

Uso di dummy causa della Multicollinearità esatta

ADATTAMENTO AI DATI

L' R^2 è dato dal rapporto tra la devianza della regressione e la devianza totale. L' R^2 esprime con immediatezza quanta parte della variabilità complessiva del fenomeno Y , che si intende spiegare tramite X , si può attribuire al legame lineare stimato mediante la retta di regressione. L' R^2 è un numero compreso tra 0 e 1.

Se esso vale 0 allora la devianza della regressione è zero e quindi la retta stimata è parallela all'asse x (il modello non contribuisce in alcun modo a spiegare la variabile dipendente y_i in aggiunta alla sua media campionaria). Quindi la retta di regressione non ha alcuna capacità interpretativa per Y .

Viceversa se l' R^2 assume valore pari a 1 la devianza della regressione coincide con la devianza totale e, quindi, la devianza residua è nulla, cioè tutti i residui sono nulli e tutti i valori stimati per la Y sono collocati esattamente sulla retta di regressione che passa per tutti i dati osservati (i quali evidentemente sono perfettamente allineati).

$$\left\{ \begin{array}{l} R^2 = \frac{V(\hat{y}_i)}{V(y_i)} = \frac{\frac{1}{N-1} \sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\text{DevianzaRegressione}}{\text{DevianzaTotale}} = [\text{se } \exists \text{ intercetta}] = 1 - \frac{V(e_i)}{V(y_i)} \\ \\ R^2_{\text{senza intercetta}} = R^2_{\text{non centrato}} = \frac{\sum_{i=1}^N \hat{y}_i^2}{\sum_{i=1}^N y_i^2} = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N y_i^2}, \quad y_i^2 \text{ senza media} \\ \\ R^2_{\text{modello non lineare}} = \text{Corr}^2(y_i, \hat{y}_i) = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\left[\sum_{i=1}^N (y_i - \bar{y})^2 \right] \left[\sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \right]}, \quad \left\{ \begin{array}{l} \text{per OLS coincide con altre def} \\ \text{riflette cmq qualità dell'aprox lineare} \end{array} \right. \\ \\ R^2_{\text{corretto}} = 1 - \frac{\frac{1}{N-K} \sum_{i=1}^N e_i^2}{\frac{1}{N-1} \sum_{i=1}^N y_i^2} \rightarrow \text{penalizzazione per inclusione variabili esplicative aggiuntive} \end{array} \right.$$

DUMMY

Esistono Dummy additive $y = \alpha + \beta_1 D + \beta_2 x_2$ e Dummy moltiplicative $y = \alpha + \beta_2 x_2 + \beta_3 x_2 D$. Le Dummy additive modificano l'intercetta.

PROPIETA' DEI STIMATORI

$$\left\{ \begin{array}{l} E[\underline{b}] = \underline{\beta}, \quad V[\underline{b}] = \begin{cases} \sigma^2 (X'X)^{-1} \\ \sigma^2 \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \end{cases} \\ \hat{V}[\underline{b}] = \left[\text{stimo } \sigma^2 \text{ con } \hat{\sigma}^2 = s^2 = \frac{1}{N-K} \sum_{i=1}^N e_i^2 \right] = s^2 (X'X)^{-1} = s^2 \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \\ s^2 \text{ è la stima della varianza dei residui considerando che essi hanno valore atteso nullo} \\ \text{La var stimata di un elemento } b_k \text{ è dato da } s^2 c_{kk}, \text{ dove } c_{kk} \text{ è l'elemento } (k, k) \text{ della matrice } \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \end{array} \right.$$

SCelta TRA PIU' MODELLI - AIC, BIC

AIC e BIC sono criteri alternativi all'R quadro corretto che forniscono un trade-off tra qualità dell'adattamento dei dati, e la semplicità del modello, misurata dal numero dei parametri K .

$$AIC = \text{Akaike Information Criteria} = \log \left(\frac{1}{N} \sum_{i=1}^N e_i^2 \right) + \frac{2K}{N}, \quad BIC = \text{Bayesian ...} = \log \left(\frac{1}{N} \sum_{i=1}^N e_i^2 \right) + \frac{K}{N} \log N$$

Modello con AIC, BIC minore è il migliore. Entrambi prevedono penalizzazione per aggiunta di regressori.

IPOTESI DISTRIBUZIONE

$$A5) \underline{\varepsilon} \sim N(0, \sigma^2 I_n) = e_i \sim \text{NID}(0, \sigma^2) \rightarrow \text{sostituisce } A1), A3) \text{ e } A4)$$

$$\text{Se vale } A5) \rightarrow \underline{b} \sim N(\underline{\beta}, \sigma^2 (X'X)^{-1}) \rightarrow b_k \sim N(\beta_k, \sigma^2 c_{kk})$$

Gli stimatori dei minimi quadrati (e le loro proprietà) non derivano da alcuna ipotesi circa la distribuzione delle v.c. ε_i e di conseguenza per n finito nulla si può dire sulla distribuzione esatta degli stimatori OLS; in particolare non si possono derivare test ed intervalli di confidenza esatti per i parametri. Inoltre, non si può discutere di proprietà come la sufficienza e l'efficienza assoluta. Tuttavia con l'ipotesi che i termini di errore sono indipendenti e normali con valori medi nulli e varianze costante il campione osservato (y_1, \dots, y_n) è la realizzazione di un campione casuale $\underline{Y} = (Y_1, \dots, Y_n)$ le cui componenti $Y_i \sim N(\underline{x}_i' \underline{\beta}, \sigma^2)$ costituiscono una n -pla di v.c. indipendenti e somiglianti. Calcolando gli stimatori e le stime ML (Max Likelihood) si vede che essi coincidono con quelle OLS. Gli stimatori ML di $\underline{\beta}$ godono di tutte le proprietà di Gauss-Markov ma sono anche sufficienti ed efficienti in senso assoluto tra tutti gli stimatori non distorti (non solo lineari e non distorti) perché si può dimostrare che essi raggiungono il limite inferiore della disuguaglianza di Cramer-Rao. Inoltre, per qualsiasi n finito, essendo $\underline{\beta}$ delle combinazioni lineari di v.c. Normali gli stimatori ML/OLS \underline{b} sono distribuiti Normalmente con media varianze già esplicitate.

VERIFICA D'IPOTESI

TEST T $\rightarrow \begin{cases} H_0: \beta_k = \beta_k^0 \\ H_1: \beta_k \neq \beta_k^0 \end{cases} \rightarrow t_k = \frac{(b_k - \beta_k^0)}{s\sqrt{c_{kk}}} \sim \mathcal{T}_{N-K} \rightarrow \text{Approx normale } RC(\alpha) \rightarrow \text{rifiuto ipotesi nulla se } |t_k| > 1,96$

TEST RATIO \rightarrow Statistica test da calcolare per verificare ipotesi nulla $\beta_k = 0 \rightarrow \frac{b_k}{se(b_k)}$ con $s\sqrt{c_{kk}} = se(b_k)$

TEST F – CONGIUNTO DI SIGNIFICATIVITA' $\rightarrow J$ dei K ($J < K$) regressori sono $= 0 \rightarrow H_0: \beta_{K-J+1} = \dots = \beta_K = 0$

$S_0 (= \sum e_i^2 \text{ modello vincolato, cioè omettendo } J \text{ regressori}), \quad S_1 (= \text{somma quadrati residui modello completo})$

$f = \frac{(S_0 - S_1)/J}{S_1/(N - K)} = \frac{(R_1^2 - R_0^2)/J}{(1 - R_1^2)/(N - K)} \sim \mathcal{F}_{N-K}^J \rightarrow RC(\alpha): P(f > \mathcal{F}_{N-K,\alpha}^J) = \alpha, \quad \text{con } \begin{cases} J = \text{Numero Regressori Nulli} \\ K = \text{Numero Regressori Totali} \end{cases}$

TEST F per livello significatività del 5% può essere accettato guardando il p-value (se minore di 5% allora i coefficienti sono diversi da zero).

Table 2.2 OLS results wage equation

Dependent variable: wage			
Variable	Estimate	Standard error	t-ratio
constant	-3.3800	0.4650	-7.2692
male	1.3444	0.1077	12.4853
school	0.6388	0.0328	19.4780
exper	0.1248	0.0238	5.2530
$s = 3.0462 \quad R^2 = 0.1326 \quad \bar{R}^2 = 0.1318 \quad F = 167.63$			

La tabella di seguito può essere interpretata nel modo seguente: sotto "Estimate" viene riportato il valore dei coefficienti del modello di regressione (essi possono essere interpretati solo sotto una condizione ceteris paribus, che stabilisce che tutte le altre variabili incluse nel modello devono essere mantenute costanti); lo "Standard Error" definisce la varianza dei coefficienti di regressione; il t-ratio è la statistica test da calcolare per verificare l'ipotesi nulla $\beta_k = 0$ affinché l'ipotesi che il coefficiente sia uguale a zero sia accettata il t-

ratio deve essere compreso tra -1,96 e +1,96; la statistica test F, infine, verifica congiuntamente che tutti e tre i coefficienti di pendenza siano congiuntamente nulli.

TEST DI WALD

$$WALD \rightarrow R\underline{b} = \underline{q} \rightarrow R\underline{b} \sim N(R\underline{\beta}, \sigma^2 R(X'X)^{-1}R') \rightarrow \xi = \frac{(\underline{R\underline{b}} - \underline{q})[R(X'X)^{-1}R']^{-1}(\underline{R\underline{b}} - \underline{q})'}{\sigma^2} \sim \chi^2_J, \begin{cases} Es. R \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \\ Es. \underline{q} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, J = \dim \underline{q} \end{cases}$$

PROPRIETA' PER CORRETTEZZA

Se non vale A5) $\underline{\varepsilon} \sim N(0, \sigma^2 I_n) \rightarrow$ distribuzione \underline{b} *NON NORMALE*, se $E(\underline{\varepsilon}|X) \neq 0 \rightarrow \underline{b}$ *NON CORRETTO*

CONSISTENZA

La consistenza è una cosiddetta proprietà per grandi campioni e, intuitivamente, afferma che se utilizziamo un numero sempre più grande di osservazioni, la probabilità che lo stimatore si discosti dal vero valore diventa sempre più piccola. Valori di \underline{b} non prossimi a $\underline{\beta}$ diventano sempre più improbabili.

$$\left\{ \begin{array}{l} A6) \frac{1}{N} \sum_{i=1}^N \underline{x}_i \underline{x}_i' \text{ converge matrice non singolare } \Sigma_{xx} + A7) E(\underline{x}_i \varepsilon_i) = 0 \rightarrow \\ \rightarrow \underline{b} \text{ CONSISTENTE} \leftrightarrow \text{plim } \underline{b} = \underline{\beta} \leftrightarrow \lim_{N \rightarrow \infty} P(|b_k - \beta_k| < \epsilon) = 1, \forall \delta \end{array} \right.$$

NORMALITA' ASINTOTICA

$$\text{Sotto ipotesi A1} - \text{A4} + \text{A6} \rightarrow \text{Normalità ASINTOTICA} \rightarrow \underline{b} \sim N\left(\underline{\beta}, \sigma^2 \frac{\Sigma_{xx}^{-1}}{N}\right) \rightarrow \text{stimo } \Sigma_{xx}^{-1} \rightarrow \underline{b} \sim N\left(\underline{\beta}, s^2 \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i'\right)^{-1}\right)$$

PREVISIONE

$$\text{Dato un } \underline{x}_0 \rightarrow \text{Previsione OLS } \hat{y}_0 = \underline{x}_0' \underline{b} \rightarrow V(\hat{y}_0) = \text{VAR PREDITTORE} = V(\underline{x}_0' \underline{b}) = \underline{x}_0' V(\underline{b}) \underline{x}_0 = \sigma^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0$$

$$\text{VARIANZA ERRORE DI PREVISIONE} \rightarrow V(y_0 - \hat{y}_0) = \sigma^2 + \sigma^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0$$

2 INTERPRETAZIONE E CONFRONTO MODELLI DI REGRESSIONE

INTERPRETAZIONE MODELLO LINEARE

Il coefficiente β_k misura la variazione attesa in y_i quando x_{ik} cambia di una unità, ma tutte le altre esplicative in \underline{x}_i' restano costanti.

$$y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \rightarrow \beta_k = \frac{\partial E(y_i | \underline{x}_i)}{\partial x_{ik}} \rightarrow \begin{cases} \text{Misura variazione assoluta (Ceteris paribus: altre var non cambiano)} \\ \text{della variabile } y_i \text{ rispetto ad una variazione di } x_{ik} \end{cases}$$
$$\frac{\partial E(y_i | \underline{x}_i)}{E(y_i | \underline{x}_i)} \bigg/ \frac{\partial x_{ik}}{x_{ik}} = \text{ELASTICITA' MODELLO LINEARE} = \frac{x_{ik}}{\underline{x}_i' \underline{\beta}} \beta_k \rightarrow \begin{cases} \text{Nei modelli lineari} \\ \text{le elasticità non sono costanti} \\ \text{come in quelli logaritmici} \end{cases}$$

INTERPRETAZIONE MODELLO LOGARITMICO

$$\log y_i = (\log \underline{x}_i)' \underline{\gamma} + i \rightarrow \gamma_k = \frac{\partial E(\log y_i | \log \underline{x}_i)}{\partial \log x_{ik}} \approx \left(\log(g(x))' = \frac{g'(x)}{g(x)} \right) \approx \frac{\partial E(y_i | \underline{x}_i)}{E(y_i | \underline{x}_i)} \bigg/ \frac{\partial x_{ik}}{x_{ik}}$$

Nei modelli logaritmici le variazioni attese sono approssimabili come elasticità e sono costanti -> esse, approssimativamente uguali ai coefficienti di regressione, misurano la variazione relativa (percentuale) Ceteris paribus della variabile y_i rispetto ad una variazione di x_{ik} .

DISTORSIONE DA VARIABILI OMESSE

La distorsione da variabile omessa nasce quando viene omessa una variabile dalla regressione, che è una determinante di Y ed è correlata con uno o più dei regressori. L'omissione di variabili rilevanti può rendere le stime OLS inconsistenti. Infatti, la presenza di una variabile omessa potrebbe determinare correlazione tra i residui e la variabile della regressione correlata con la variabile omessa. Una soluzione si ha utilizzando una variabile strumentale.

CONFRONTO MODELLI LINEARI NON NESTED (NESSUNO CASO PARTICOLARE DELL'ALTRO)

$$y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \rightarrow \text{MODELLO A}, \quad y_i = \underline{z}_i' \underline{\gamma} + i \rightarrow \text{MODELLO B}$$

Stimo una regressione uguale al modello A o B aggiungendo le variabili non presenti in un modello, ma presenti nell'altro modello. Guardo alla significatività dei coefficienti aggiunti.

DATA MINING

Stesso insieme di osservazione viene usato sia per la specificazione del modello che per verificare ipotesi.

CONFRONTO MODELLI LINEARI-LOGARITMICI NON NESTED

$$\begin{aligned}
 & y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \rightarrow \text{MODELLO A}, \quad \log y_i = (\log \underline{x}_i)' \underline{\gamma} + i \rightarrow \text{MODELLO B} \\
 & \hat{y}_i = \text{previsioni OLS usando A}, \quad \tilde{y}_i = \text{previsioni OLS usando B} \\
 & \left\{ \begin{aligned} & \text{Verifico linearità} \rightarrow y_i = \underline{x}_i' \underline{\beta} + \delta_{\text{lin}}(\log \hat{y}_i - \log \tilde{y}_i) + \varepsilon_i \rightarrow \text{se } \delta_{\text{lin}} = 0 \rightarrow \text{OK LINEARITA'} \\ & \text{Verifico log - linearità} \rightarrow \log y_i = (\log \underline{x}_i)' \underline{\gamma} + \delta_{\text{log}}(\hat{y}_i - \tilde{y}_i) + i \rightarrow \text{se } \delta_{\text{log}} = 0 \rightarrow \text{OK LOG - LINEARITA'} \end{aligned} \right.
 \end{aligned}$$

TEST FORMA FUNZIONALE

Una strategia per verificare la correttezza della forma funzionale di $E(y_i | x_i) = \underline{x}_i' \underline{\beta}$ consiste nel verificare la significatività di termini aggiuntivi non lineari in \underline{x}_i . Si può usare test t standard, test F, o più in generale test di Wald, ma si deve essere in grado di specificare in dettaglio l'ipotesi alternativa. Altrimenti faccio un test RESET dove tramite una regressione ausiliaria $y_i = \underline{x}_i' \underline{\beta} + a_2 \hat{y}_i^2 + \dots + a_q \hat{y}_i^q + \varepsilon_i$ verifico se le $a_i = 0$. Ovvero verifico che i coefficienti di una regressione ausiliaria costruita sulle previsioni del modello lineare abbia coefficienti non nulli.

TEST DI CHOW O DI BREAK STRUTTURALE

I coefficienti del modello sono diversi per due sottogruppi del campione (ad esempio prima o dopo un importante intervento di politica macroeconomica). Definiamo la variabile indicatrice g_i : $g_1 = 0$ gruppo 1, $g_1 = 1$ gruppo 2.

$$y_i = \underline{x}_i' \underline{\beta} + g_i \underline{x}_i' \underline{\gamma} + \varepsilon_i, \quad \text{dove il vettore } K - \text{dimensionale } g_i \underline{x}_i' \text{ contiene interazioni variabile esplicative con } g_i$$

Il vettore dei coefficienti per il gruppo 1 è $\underline{\beta}$ mentre quello del gruppo 2 è $\underline{\beta} + \underline{\gamma}$

$$H_0: \underline{\gamma} = 0 \rightarrow f = \frac{(S_V - S_{NV})/K}{S_{NV}/(N - 2K)} \sim \mathcal{F}_{N-2K}^K, \quad \begin{cases} K = \text{numero regressori compresa intercetta} = 0 \\ 2K = \text{numero di regressori totali} \rightarrow \text{si noti che } \dim \underline{x}_i' = \dim g_i \underline{x}_i' = K \end{cases}$$

Un test alternativo che non fa uso delle variabili dummy confronta i residui della regressione sui due gruppi con la somma dei residui di due regressioni, una costruita sul gruppo 1 e una sul gruppo 2.

Supponendo che il modello dei dati sia $y_t = a + bx_{1t} + cx_{2t} + \varepsilon_t$. Dividiamo i dati in due gruppi:

$$\begin{cases} y_t = a_1 + b_1 x_{1t} + c_1 x_{2t} + \varepsilon_t \\ y_t = a_2 + b_2 x_{1t} + c_2 x_{2t} + \varepsilon_t \end{cases}$$

L'ipotesi nulla del test di Chow asserisce che $a_1 = a_2, b_1 = b_2, c_1 = c_2$. Sia S_c la somma dei residui del modello completo, S_1 somma dei residui del modello 1, S_2 la somma dei residui del modello 2. Allora la statistica test di Chow è la seguente:

$$f = \frac{(S_c - (S_1 + S_2))/K}{(S_1 + S_2)/(N_1 + N_2 - 2K)} \sim \mathcal{F}_{N_1+N_2-2K}^K$$

3 STIMATORE ALTERNATIVO ETROSCHEDASTICITA' AUTOCORRELAZIONE

La presenza di autocorrelazione (matrice di varianze-covarianze non diagonale e dunque termini di errore diversi sono correlati) ed eteroschedasticità (termini di errore diversi non hanno varianze identiche con la conseguenza che gli elementi sulla diagonale della matrice di covarianza non sono tutti uguali) implica termini di errore del modello non più indipendenti ed identicamente distribuiti. In questi casi lo stimatore OLS può essere comunque corretto e consistente, ma la sua matrice di covarianza è diversa da quella usuale. Può inoltre non essere BLUE.

CONSEGUENZE STIMATORE OLS

$$\left\{ \begin{array}{l} V(\underline{\varepsilon} | X) = \sigma^2 \Psi, \Psi \text{ matrix def positiva} \rightarrow \text{unica ipotesi correttezza } \underline{b} \text{ OLS } \text{ è } E(\varepsilon|X) = 0 \text{ stimatore ancora corretto} \\ \text{varianza cambia} \rightarrow V(\underline{b} | X) = \sigma^2 (X'X)^{-1} X' \Psi X (X'X)^{-1} \\ \text{test F, t non più validi, non vale più teorema di gauss markov} \end{array} \right.$$

STIMATORE ALTERNATIVO

Poiché conosciamo l'espressione del miglior stimatore lineare e corretto sotto le ipotesi di Gauss-Markov, possiamo trasformare il modello in modo da ricondurci sotto le condizioni del teorema, in modo tale che il modello trasformato abbia termini d'errore homoschedastici e non autocorrelati.

$$\left\{ \begin{array}{l} \Psi^{-1} = P'P, \quad \text{supponiamo di saper calcolare } P, \text{ poichè } \Psi \text{ è definita positiva esiste sempre } P \\ \Psi = (P'P)^{-1} = P^{-1}(P')^{-1} \rightarrow P\Psi P' = I \\ \underline{y} = X\underline{\beta} + \underline{\varepsilon} \rightarrow \text{trasformo in } P\underline{y} = PX\underline{\beta} + P\underline{\varepsilon} \rightarrow \underline{y}^* = X^*\underline{\beta} + \underline{\varepsilon}^* \text{ che soddisfa Gauss Markov} \\ \underline{\hat{\beta}} = \text{MINIMI QUADRATI GENERALIZZATI (GLS)} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}\underline{y} \\ V(\underline{\hat{\beta}}) = \sigma^2(X'\Psi^{-1}X)^{-1} \\ \text{Se sostituisco a } \Psi \text{ una stima ottengo} \rightarrow \text{MINIMI QUADRATI GENERALIZZATI CALCOLABILI (EGLS, FGLS)} \end{array} \right.$$

ETEROSCHEDASTICITA'

OLS rimangono corretti, ma non più efficienti e i loro standard error vengono stimati con una formula errata (test t, F non sono più validi). O si usa uno stimatore GLS e si recupera l'efficienza oppure si stimano gli standard error dei parametri apportando qualche correzione alla matrice di covarianze.

$$V(\underline{\varepsilon} | X) = \sigma^2 \text{diag}(h_i^2) = \sigma^2 \Psi \rightarrow P = \text{diag}(h_i^{-1}) \rightarrow P\underline{y} = PX\underline{\beta} + P\underline{\varepsilon} \rightarrow \frac{y_i}{h_i} = \left(\frac{x_i}{h_i}\right)' \underline{\beta} + \frac{\varepsilon_i}{h_i} \rightarrow$$

$$\underline{\hat{\beta}} = \text{STIMATORE GLS DEI MINIMI QUADRATI PONDERATI} = \left(\sum_{i=1}^N h_i^{-2} \underline{x}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^N h_i^{-2} \underline{x}_i y_i$$

Si chiama ponderato perché rappresenta uno stimatore dei minimi quadrati in cui ogni osservazione è ponderata con l'inverso della radice della varianza dell'errore corrispondente.

$$V[\underline{\hat{\beta}}] = \sigma^2 \left(\sum_{i=1}^N h_i^{-2} \underline{x}_i \underline{x}_i' \right)^{-1}$$

$$\text{Se sostituisco a } h_i^2 \text{ una stima } \hat{h}_i^2 \rightarrow \underline{\hat{\beta}}^* = \text{STIMATORE EGLS} = \left(\sum_{i=1}^N \hat{h}_i^{-2} \underline{x}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^N \hat{h}_i^{-2} \underline{x}_i y_i$$

$$\text{Se non ho } \hat{h}_i^2 \rightarrow \text{formula white} \rightarrow \text{utilizzo } \underline{b} \text{ con } \hat{V}[\underline{b}] = \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^N e_i^2 \underline{x}_i \underline{x}_i' \left(\sum_{i=1}^N \underline{x}_i \underline{x}_i' \right)^{-1}$$

ETEROSCHEDASTICITA' MOLTIPLICATIVA

Varianza degli errori dipende da un insieme di variabili esogene raccolte in un vettore J dimensionale z_i .

$V(\varepsilon_i | X) = \sigma_i^2 = \sigma^2 \exp\{a_1 z_{i1} + \dots + a_j z_{ij}\} = \sigma^2 \exp\{z_i' \underline{a}\}$, \underline{z}_i = vettore di variabili osservate funzione di \underline{x}_i
se gli $h_i^2 = \exp\{z_i' \underline{a}\}$ sono ignoti perché ignoto $\underline{a} \rightarrow$ stimo \underline{a} regredendo $\log e_i^2 = \log \sigma + z_i' \underline{a} + v_t$, passando ai logaritmi da $V(\varepsilon_i | X) = \sigma^2 \exp\{z_i' \underline{a}\}$.

MODELLO E TEST PER DUE VARIANZE

Situazione di due varianze ignote e cioè la varianza di ε_i è pari a σ_A^2 se osservazione appartiene al gruppo A, σ_B^2 se osservazione appartiene al gruppo B.

Dal risultato $\begin{cases} X \sim N(\mu, \sigma^2) \\ Y \sim N(v, t^2) \end{cases} \rightarrow T_{n,m} = \frac{s_{x,n}^2}{s_{y,m}^2} \sim \frac{\sigma^2}{t^2} F_{n-1, m-1}$ sotto l'ipotesi nulla $\sigma^2 = t^2$ e vale il risultato seguente:

$$\begin{cases} H_0: \sigma_A^2 = \sigma_B^2 \\ H_1: \sigma_A^2 \neq \sigma_B^2 \end{cases} \rightarrow Rc(\alpha): \begin{cases} s_A^2/s_B^2 \geq F_{(\alpha/2, N_A-K, N_B-K)} \\ s_A^2/s_B^2 \leq F_{(1-\alpha/2, N_A-K, N_B-K)} \end{cases} \rightarrow \underline{\hat{\beta}}^* = \left(\sum_{i \in A} s_A^{-2} \underline{x}_i \underline{x}_i' + \sum_{i \in B} s_B^{-2} \underline{x}_i \underline{x}_i' \right)^{-1} \left(\sum_{i \in A} s_A^{-2} \underline{x}_i y_i + \sum_{i \in B} s_B^{-2} \underline{x}_i y_i \right)$$

RESIDUI AL QUADRATO E VARIANZA DEI RESIDUI

I residui al quadrato del modello originale molto spesso vengono utilizzati come una proxy della varianza del termine di errore, infatti il termine di errore è assunto a media zero e la varianza di una variabile casuale con media zero è l'aspettativa del suo quadrato.

TEST DI ETEROSCHEDASTICITA' MOLTIPLICATIVA

$$\sigma_i^2 = \sigma^2 \exp\{z_i' \underline{a}\}, \quad H_0: \underline{a} = 0 \text{ vs } H_1: \underline{a} \neq 0$$

Test standard su \underline{a} nella regressione $\log e_i^2 = \log \sigma^2 + z_i' \underline{a} + v_t \sim F_{N-J-1}^J$. Il test vale solo approssimativamente perché v_t non rispetta le ipotesi di Gauss Markov.

TEST DI ETEROSCHEDASTICITA' DI BREUSCH-PEGAN

$$\sigma_i^2 = \sigma^2 h\{z_i' \underline{a}\}, \quad H_0: \underline{a} = 0 \text{ vs } H_1: \underline{a} \neq 0$$

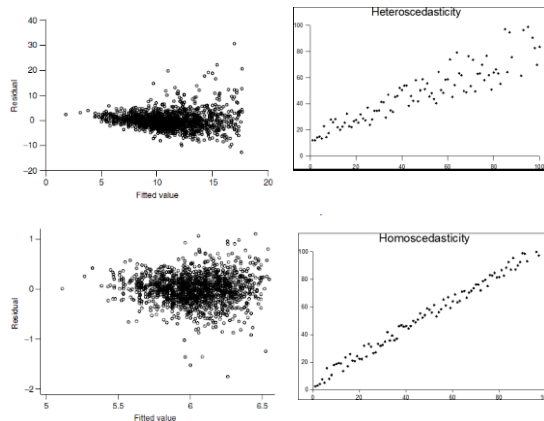
Regredisco e_i^2 su \underline{z}_i (J dimensionale) e una costante, sotto l'ipotesi nulla vale: $\xi = NR^2 \rightarrow \text{asintoticamente} \rightarrow \chi_{(J)}^2$

TEST DI ETEROSCHEDASTICITA' DI WHITE

Regredisco e_i su una costante più momento primo, secondo, momenti incrociati dei regressori originali e $\xi = NR^2 \rightarrow \chi_{(p)}^2$. Dove $p = \text{numero regressori} + \text{intercetta}$. Ad esempio:

$$e_i^2 = \text{cost} + aE(x) + bE(z) + cE(x^2) + dE(z^2) + fE(xz)$$

Se l'errore è omoschedastico tutti i coefficienti dovrebbero risultare prossimi a zero.



Il grafico a sinistra mostra eteroschedasticità: confrontando i residui con i valori stimati, i residui non si disperdono entro una banda. Anche il secondo grafico mostra eteroschedasticità, perché le osservazioni si disperdono attorno alla retta di interpolazione non con la stessa varianza.

Il grafico a sinistra mostra omoschedasticità: confrontando i residui con i valori stimati, i residui si disperdono entro una banda (ipotesi di omoschedasticità). Anche il secondo grafico mostra omoschedasticità perché le osservazioni si disperdono attorno alla retta di interpolazione con la stessa varianza.

AUTOCORRELAZIONE

Uno o più termini d'errore consecutivi sono correlati. Minimi quadrati restano corretti, ma divengono inefficienti e il loro standard error si stimano usando una formula sbagliata. Autocorrelazione si verifica solo usando campioni in serie storica (invece che i da 1 a N , t da 1 a T). Presenza di autocorrelazione negli errori di un modello in serie storica e molto spesso è sintomatologia di una specificazione dinamica insufficiente del modello (predico consumo di gelato in base al prezzo del gelato e reddito senza contare la variabile tempo).

AUTOCORRELAZIONE PRIMO ORDINE

$$\left\{ \begin{array}{l} y_t = \underline{x}_t' \underline{\beta} + \varepsilon_t, \quad \text{con } \varepsilon_t = \rho \varepsilon_{t-1} + v_t, \quad \text{con } v_t \text{ media nulla e varianza } \sigma_v^2, \text{ privo di correlazione seriale} \\ \varepsilon_t = \rho \varepsilon_{t-1} + v_t \text{ interpretabile con AR(1),} \quad \text{si assume stazionarietà } (|\rho| < 1) \text{ e che } E[\varepsilon_1] = 0 \rightarrow \\ \text{Per avere ipotesi Gauss - Markov valide trasformo errori in errori non autocorrelati } \rightarrow v_t = \varepsilon_t - \rho \varepsilon_{t-1} \\ \rightarrow y_t - \rho y_{t-1} = (\underline{x}_t' - \rho \underline{x}_{t-1}') \underline{\beta} + v_t, \quad t = 2, 3, \dots, T \rightarrow \text{stima GLS approssimato per mancanza 1 osservazione} \\ \text{se } \rho \text{ è ignoto lo stimo con OLS su regressione dei residui } e_t = \rho e_{t-1} + v_t \rightarrow \hat{\rho} = \left(\sum_{t=2}^T e_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^T e_t e_{t-1} \right) \end{array} \right.$$

TEST DI DURBIN WATSON

Considero i residui OLS e_t e il suo ritardo e_{t-1} . Per applicarlo si ipotizza che è possibile trattare le \underline{x}_t come deterministiche e che in esse sia incluso un termine di intercetta. La statistica di Durbin-Watson è tale che:

$$dw \approx 2 - 2\hat{\rho}$$

Un valore di dw prossimo a 2 indica che ρ è prossimo a zero e quindi di assenza di autocorrelazione, un valore di dw 'molto più piccolo di 2' indica autocorrelazione positiva $\rho > 0$; se dw è 'molto più grande di 2' si può ipotizzare autocorrelazione negativa. Esistono tavole costruite da Durbin-Watson con regioni di incertezza.

TEST DI AUTOCORRELAZIONE DEL PRIMO ORDINE-LAGRANGE BREUSCH E GODFREY

Considero la regressione del residuo OLS e_t sul suo ritardo e_{t-1} . Questa regressione ausiliaria produce una stima $\hat{\rho}$. Analizzo R^2 della regressione ausiliaria (inclusendo il termine di intercetta). Un R^2 prossimo a zero implica che i residui ritardati non spiegano quelli correnti. Inoltre si dimostra che $R^2 * (T - 1) \sim \chi_1^2$.

AUTOCORRELAZIONE DI ORDINE SUPERIORE

Per modelli con errori autoregressivi di ordine superiore, ad esempio $y_t = \underline{x}_t' \underline{\beta} + \varepsilon_t$, con $\varepsilon_t = \rho \varepsilon_{t-1} + \rho \varepsilon_{t-2} + v_t$, il processo è uguale a quello del primo ordine. Test di Breusch-Godfrey continua a valere ampliando la regressione a più regressori ritardati e calcolando sempre $l'R^2$.

AUTOCORRELAZIONE A MEDIA MOBILE

Una specificazione autoregressiva come $y_t = \underline{x}_t' \underline{\beta} + \varepsilon_t$, con $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$ implica che tutti i termini di errore siano correlati fra loro, anche se la correlazione fra elementi lontani nel tempo è molto piccola, al punto di essere quasi trascurabile. Se invece la teoria economica suggerisce che solo alcuni termini d'errore sono correlati si può ricorrere a un processo per gli errori a media mobile.

Un esempio di errori a media mobile è $y_t^* = \underline{x}_t^{*'} \underline{\beta} + \varepsilon_t$, con $\varepsilon_t = v_t + v_{t-1}$.

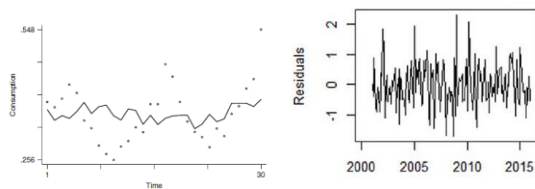
Questi modelli di regressione sono sensibilmente più difficili da stimare a causa della complessità della trasformazione che genera errori con le proprietà di Gauss-Markov.

STANDARD ERROR DI NEWEY WEST

$y_t = \underline{x}_t' \underline{\beta} + \varepsilon_t$, con $E(\underline{x}_t \varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_{t-s}) = 0$ per $s = H, H+1, \dots$, (correlazione nulla oltre un certo ritardo) utilizzo OLS (che è consistente) e stimo la matrice di covarianza con:

$$\hat{V}(\underline{b}) = \left(\sum_{t=1}^T \underline{x}_t \underline{x}_t' \right)^{-1} T S^* \left(\sum_{t=1}^T \underline{x}_t \underline{x}_t' \right)^{-1}, \text{ con } S^* = \frac{1}{T} \sum_{t=1}^T e_t^2 \underline{x}_t \underline{x}_t' + \frac{1}{T} \sum_{j=1}^{H-1} w_j \sum_{s=j+1}^T e_s e_{s-j} (\underline{x}_s \underline{x}_{s-j}' + \underline{x}_{s-j} \underline{x}_s')$$

Questi standar error sono consistenti in caso di eteroschedasticità e autocorrelazione (HAC, Heteroskedasticity and Autocorrelation Consistent).



Nella figura a sinistra viene riportato un esempio di una regressione con residui autocorrelati, come si può vedere le osservazioni oscillano attorno ai valori previsti dal modello. Nella secondo figura si osserva un grafico con residui autocorrelati, come si può vedere essi presentano un andamento ciclico.

4 ENDOGENITA', VARIABILI STRUMENTALI E GMM

Tutti i risultati presentati precedentemente, anche quando non esplicitato, sono validi soltanto quando vale l'ipotesi A2) ovvero che la variabile esplicativa \underline{x}_i è indipendente dal termine di errore ε_i . Se questa ipotesi non vale per non intaccare l'ipotesi di consistenza dello stimatore OLS (anche in caso di presenza di eteroschedasticità e autocorrelazione grazie alla correzione di Newey West) deve essere almeno valida la seguente ipotesi: A7) $E(\underline{x}_i \varepsilon_i) = 0$. Nei casi in cui non è possibile assumere né l'ipotesi A2) né l'ipotesi A7) si deve ricorrere alle variabili strumentali. Infatti il modello per essere ritenuto idoneo deve rappresentare il valore atteso condizionato della variabile dipendente (o endogena) date le variabili indipendenti (o esogene). Cioè

Se vale $y_t = a + bx_t + \varepsilon_t \rightarrow E(y_t | x_t) = a + bx_t + E(\varepsilon_t | x_t) = a + bx_t \rightarrow$ cioè deve essere $E(\varepsilon_t | x_t) = 0$

Alcuni esempi in cui non vale la condizione A7) sono dati dalla presenza della variabile dipendente ritardata fra le esplicative quando il termine d'errore è autocorrelato, di errori di misura nei regressori che contribuiscono quindi a generare il termine di errore totale, di simultaneità o endogeneità nei regressori (ho due variabili che sono determinate congiuntamente nel modello, ovvero flussi circolari in economia).

STIMATORE GENERALIZZATO DELLE VARIABILI STRUMENTALI

$$y_i = \underline{x}_i' \underline{\beta} + \varepsilon_i \quad (\rightarrow y = X\underline{\beta} + \underline{\varepsilon}), \quad \text{con } \underline{x}_i \text{ vettore di dimensione } K, \quad X \text{ matrice } K * N$$

$$\text{Stimatore OLS basato sulle } K \text{ condizione dei momenti} \rightarrow E(\varepsilon_i \underline{x}_i) = E[(y_i - \underline{x}_i' \underline{\beta}) \underline{x}_i] = 0$$

$$\text{Supponiamo siano presenti } R \text{ strumenti} \rightarrow E(\varepsilon_i \underline{z}_i) = E[(y_i - \underline{x}_i' \underline{\beta}) \underline{z}_i] = 0$$

$$\left\{ \begin{array}{l} R = K \rightarrow \hat{\underline{\beta}}_{IV} = \left(\sum_{i=1}^N \underline{z}_i \underline{x}_i' \right)^{-1} \sum_{i=1}^N \underline{z}_i y_i \rightarrow \hat{\underline{\beta}}_{IV} = (Z'X)^{-1} Z'y \\ R > K \rightarrow \text{Ho più strumenti che Regressori} \rightarrow \text{anzichè tralasciare strumenti minimizzo nuova forma quadratica} \\ R < K \rightarrow \underline{\beta} \text{ è sottoidendificato, inversa non esiste e condizioni } 1^{\text{o}} \text{ ordine verificate da infinite soluzioni} \end{array} \right.$$

STRUMENTI DEBOLI

Le proprietà dello stimatore IV possono essere molto precarie, e lo stimatore gravemente distorto, se la correlazione fra gli strumenti e il regressore endogeno è troppo bassa. La distribuzione dello stimatore non è più asintoticamente normale e quindi non si possono più utilizzare gli usuali test di ipotesi. Per testare se lo strumento è debole bisogna effettuare una regressione tra la variabile ipotizzata endogena, le variabili esogene e lo strumento e poi verificare la significatività del coefficiente della variabile strumentale.

PRESENZA DI UNA VARIABILE ENDOGENA

$$y_i = \underline{x}'_i \underline{\beta} + \varepsilon_i = \underline{x}'_{1i} \underline{\beta}_1 + x_{2i} \beta_2 + \varepsilon_i$$

con $E(\varepsilon_i x_{2i}) \neq 0 \rightarrow$ quando vale x_{2i} è detta endogena rispetto all'effetto causale β_t

Con l'assunzione di endogeneità la procedura OLS fornisce uno stimatore distorto e non consistente dei parametri del modello. Per derivare uno stimatore consistente dobbiamo assicurarci che il modello sia statisticamente identificato ovvero che esistano condizioni dei momenti sufficienti a stimare i parametri del modello.

$$\text{Se } E(\varepsilon_i x_{2i}) = 0 \rightarrow \begin{matrix} \text{Condizioni} \\ \text{dei momenti} \end{matrix} \rightarrow \begin{cases} E(\varepsilon_i \underline{x}_{1i}) = 0 \\ E(\varepsilon_i x_{2i}) = 0 \end{cases}$$

Se $E(\varepsilon_i x_{2i}) \neq 0$ la seconda condizione cessa di valere. In questo caso il numero di condizioni non è più pari al numero di parametri da stimare ed è necessario aggiungere almeno una condizione dei momenti. Questa ultima viene derivata dalla disponibilità di una variabile strumentale o strumento: ovvero una variabile, indicabile con z_{2i} , che è possibile assumere incorrelata con l'errore ε_i del modello, ma correlata con il regressore endogeno x_{2i} . Allora se è possibile trovare una siffatta variabile posso calcolare lo stimatore delle variabili strumentali (IV=Instrumental Variablese) $\hat{\beta}_{IV}$:

$$\begin{aligned} E(\varepsilon_i z_{2i}) = 0 \rightarrow \begin{matrix} \text{Variabili} \\ \text{strumentali} \end{matrix} \rightarrow \begin{cases} \frac{1}{N} \sum_{i=1}^N (y_i - \underline{x}'_{1i} \hat{\beta}_{1,IV} + x_{2i} \beta_2) \underline{x}_{1i} = 0 \\ \frac{1}{N} \sum_{i=1}^N (y_i - \underline{x}'_{1i} \hat{\beta}_{1,IV} + x_{2i} \beta_2) z_{2i} = 0 \end{cases} \rightarrow \hat{\beta}_{IV} = \left(\sum_{i=1}^N \underline{z}_i \underline{x}'_i \right)^{-1} \sum_{i=1}^N \underline{z}_i y_i \rightarrow \begin{cases} \underline{x}'_i = (\underline{x}'_{1i}, x_{2i}) \\ \underline{z}_i = (\underline{x}'_{1i}, z_{2i}) \end{cases} \\ \text{Se } \text{plim} \frac{1}{N} \sum_{i=1}^N \underline{z}_i \underline{x}'_i = \Sigma_{zx} \text{ è finita e invertibile} \rightarrow \hat{\beta}_{IV} \text{ consistente} \end{aligned}$$

$$\hat{V}(\hat{\beta}_{IV}) = \hat{\sigma}^2 \left[\left(\sum_{i=1}^N \underline{x}_i \underline{z}'_i \right) \left(\sum_{i=1}^N \underline{z}_i \underline{z}'_i \right)^{-1} \left(\sum_{i=1}^N \underline{z}_i \underline{x}'_i \right) \right]^{-1}, \quad \text{con } \hat{\sigma}^2 = \frac{1}{N-K} \sum_{i=1}^N (y_i - \underline{x}'_i \hat{\beta}_{IV})^2$$

TEST DI ENDOGENEITA' DI HAUSMAN

Si basa sul confronto tra lo stimatore dei minimi quadrati e quello delle variabili strumentali: se la differenza tra i due stimatori è piccola allora la variabile supposta endogena è esogena.

Stimo la regressione ausiliaria inserendo nella regressione ordinaria una variabile che corrisponde ai residui della regressione calcolata tra la variabile ipotizzata endogena e la variabile esogena più lo strumento:

$$y_i = \underline{x}'_{1i} \underline{\beta}_1 + x_{2i} \beta_2 + \hat{v}_i \gamma + \varepsilon_i,$$

Queste operazioni riproducono (con standard error non corretti) lo stimatore IV di $\underline{\beta}$, ma generano anche una stima di γ che se è uguale a zero significa che x_{2i} è esogena. Se l'ipotesi viene accettata x_{2i} è esogena e posso usare OLS, infatti in questo caso stime OLS e stima IV coinciderebbero.

5 MOLTIPLICATORI DI LAGRANGE E ALTRI TEST VEROSIMIGLIANZA

MOLTIPLICATORI DI LAGRANGE

$\mathcal{L}(\theta; \underline{X})$ Funzione di Verosimiglianza, $V(\theta) = \log \mathcal{L}(\theta; \underline{X})$, $V'(\theta) = \frac{\partial \log \mathcal{L}(\theta; \underline{X})}{\partial \theta}$ Funzione Score

$$I_n(\theta) = -E[V_n''(\theta)] = E[(V_n'(\theta))^2] = \text{Informazione Fisher} \Rightarrow$$

$$\text{Sotto } H_0: \theta = \theta_0 \rightarrow S_n(\underline{X}) = \frac{[V'(\theta_0)]^2}{I_n(\theta_0)} \xrightarrow{d} \chi^2_{m=\text{numero parametri}}$$

FORMULAZIONE ALTERNATIVA

$$\mathbf{V}'(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}_1)}{\partial \theta_1} & \dots & \frac{\partial \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}_n)}{\partial \theta_m} \\ \dots & \dots & \dots \\ \frac{\partial \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}_n)}{\partial \theta_1} & \dots & \frac{\partial \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{x}_n)}{\partial \theta_m} \end{pmatrix}^T \text{ con } m \text{ numero parametri sia vincolati che non, } \underline{i} = \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix}$$

$$\Rightarrow \underline{i} = \underline{\beta} \mathbf{V}'(\boldsymbol{\theta}) \rightarrow \xi_{LM} = NR^2 \sim \chi^2_{j=\text{\#vincoli}}$$

ESEMPIO

Nel caso stessi testando l'eteroschedasticità, ipotizzando $V(\varepsilon_i) = \sigma_i^2 = \sigma^2 h\{\underline{z}_i' \underline{\alpha}\}$ con $H_0: \underline{\alpha} = 0$, calcolo $V'(\theta)$ da $\mathcal{L}(\underline{\beta}, \underline{\alpha}; \underline{x})$ sia per parametri vincolati che non.

6 MODELLO LOGIT E PROBIT

Se la variabile endogena è binaria utilizzando l'usuale modello di regressione lineare si incorre in errori. I modelli binari descrivono direttamente la probabilità di $y_i = 1$. Poniamo:

$$E[Y|X = x] = 1 * P(y_i = 1|x_i) + 0 * P(y_i = 0|x_i) = 1 * P(y_i = 1|x_i) = G(x'_i \beta), \quad \text{con } G(x'_i \beta) \text{ che è probabilità}$$

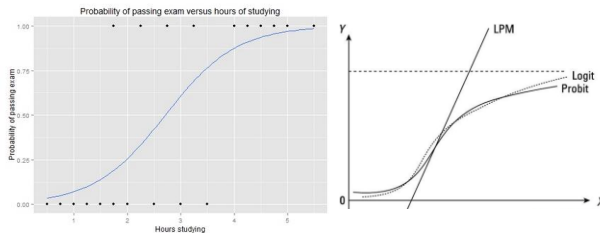
$$G(x'_i \beta) = F(x'_i \beta) \rightarrow \begin{cases} \text{PROBIT} \rightarrow \text{NORMALE STANDARD} \rightarrow F(x'_i \beta) = \Phi(x'_i \beta) = \int_{-\infty}^w \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x'_i \beta^2\right) dt \\ \text{LOGIT} \rightarrow \text{LOGISTICA STANDARD} \rightarrow F(x'_i \beta) = L(x'_i \beta) = \frac{e^{x'_i \beta}}{1 + e^{x'_i \beta}} \\ \text{MODELLO PROBABILITA' LINEARE} \rightarrow \text{UNIFORME IN } [0, 1] \rightarrow F(x'_i \beta) = \begin{cases} 0, & w < 0 \\ x'_i \beta, & 0 \leq w \leq 1 \\ 1, & w > 1 \end{cases} \end{cases}$$

$$I \text{ parametri } \beta: \begin{cases} \text{PROBIT} \rightarrow \text{NORMALE STANDARD} \rightarrow \frac{\partial \Phi(x'_i \beta)}{\partial x_{ik}} = \phi(x'_i \beta) \beta_k, \quad \text{con } \phi \text{ densità normale} \\ \text{LOGIT} \rightarrow \text{LOGISTICA STANDARD} \rightarrow \frac{\partial L(x'_i \beta)}{\partial x_{ik}} = e^{x'_i \beta} / (1 + e^{x'_i \beta})^2 \beta_k \\ \text{MODELLO PROBABILITA' LINEARE} \rightarrow \text{UNIFORME IN } [0, 1] \rightarrow \frac{\partial x'_i \beta}{\partial x_{ik}} = \beta_k, \quad \text{oppure } 0 \end{cases}$$

I parametri β sono difficili da interpretare: l'effetto di una variazione di x_{ik} dipende dal valore di tutte le x_i . Il segno dell'effetto di una variazione di x_{ik} coincide con il segno del coefficiente corrispondente β_k .

La stima di Y , nel caso Logit e Probit, esprime la probabilità che la variabile di risposta sia pari ad 1 dato l'insieme dei predittori X .

Il termine di errore $\varepsilon = P(y_i = 1|x_i) - G(x'_i \beta)$ riflette una forma di eteroschedasticità condizionale.



Come si vede dalle figure accanto (la prima un modello logit su ore di studio e passaggio o meno dell'esame) i modelli logit, probit e linear probability model come output forniscono la probabilità che la variabile aleatoria assuma valori pari a 0 o 1.

INTERPRETAZIONE MODELLO LATENTE

Il modello di scelta binaria può essere derivato a partire da alcune ipotesi comportamentali strutturali. Ad esempio considerando il caso di una donna sposata che decida se avere o meno un'occupazione retribuita. Si chiama variabile latente perché la variabile dicotomica assumerà valore pari a 0 o 1 a seconda di un'altra variabile che non è osservata e quindi è chiamata latente.

$$y_i^* = \text{espressione differenza di utilità tra fare o meno qualcosa}$$
$$y_i^* = \underline{x}_i' \underline{\beta} + \varepsilon_i, \quad y_i^* \text{ non osservata,} \quad \text{modello latente}$$

La nostra ipotesi è che un individuo scelga di lavorare se la differenza di utilità supera una certa soglia oppure no.

$$y_i = 1 \text{ (occupazione)} \Leftrightarrow y_i^* > 0, \quad y_i = 0 \text{ (non occupazione)} \Leftrightarrow y_i^* < 0$$
$$P(y_i = 1) = P(y_i^* > 0) = P(\underline{x}_i' \underline{\beta} + \varepsilon_i > 0) = P(-\varepsilon_i \leq \underline{x}_i' \underline{\beta}) = F(\underline{x}_i' \underline{\beta})$$

Dove F indica la funzione di ripartizione di $-\varepsilon_i$ o, nel caso in cui la F sia simmetrica, la funzione di ripartizione di ε_i . Specificando una distribuzione per l'errore specifico il modello. Es. $\varepsilon_i \sim N(0,1) \rightarrow \text{Modello PROBIT}$

INTERPRETAZIONE COME MODELLI LINEARI GENERALIZZATI

I modelli Probit e Logit sono casi particolari di modelli GLS (Modelli Lineari Generalizzati): in questi modelli ciascun valore della variabile dipendente Y si assume venga generato da una particolare variabile casuale della famiglia esponenziale. Si assume che $E(Y) = \mu = g^{-1}(X\beta) \rightarrow X\beta = g(\mu)$. Si dimostra che per avere il modello logit devo avere $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$, per avere il Probit $g(\mu) = \Phi^{-1}(\mu)$ ricordando che $\mu = E[Y|X = x] = P(y_i = 1|\underline{x}_i) = p \text{ successo}$.

LEGAME TRA MODELLO LOGISTICO E ODD-RATIO

Poiché nel modello logistico vale $E[Y|X = x] = P(y_i = 1|\underline{x}_i) = \frac{e^{\underline{x}_i' \underline{\beta}}}{1 + e^{\underline{x}_i' \underline{\beta}}}$ con delle semplici manipolazioni algebriche dimostro che (supponendo $P(y_i = 1|\underline{x}_i) = p = \text{probabilità di successo}$) che vale anche $\log\left(\frac{p}{1-p}\right) = \underline{x}_i' \underline{\beta}$.

STIMA

Concentrare l'attenzione sulla forma della funzione di massima verosimiglianza

$$L(\underline{\beta}) = \prod_{i=1}^N P(y_i = 1|\underline{x}_i; \underline{\beta})^{y_i} P(y_i = 0|\underline{x}_i; \underline{\beta})^{1-y_i} \rightarrow \log L(\underline{\beta}) = \sum_{i=1}^N y_i \log(F(\underline{x}_i' \underline{\beta})) + \sum_{i=1}^N (1 - y_i) \log(1 - F(\underline{x}_i' \underline{\beta}))$$

Lavoro con la funzione di log-verosimiglianza e sostituisco $P(y_i = 1|\underline{x}_i; \underline{\beta}) = F(\underline{x}_i' \underline{\beta})$.

STIMA PER I LOGIT

Nel modello logit si dimostra che la frequenza prevista (ovvero la probabilità con $\underline{\beta}$ stimato) è paria alla frequenza osservata (numero di osservazioni per cui è 1 diviso il numero di osservazioni totali). Risultati validi approssimativamente anche per PROBIT.

ADATTAMENTO AI DATI

Misuro l'accuratezza con il quale il modello approssima le osservazioni, come l'indice R^2 in regressione lineare

$\log L_1$ = valore massimo log - verosimiglianza, $\log L_0$ = valore massimo con tutti i parametri tranne intercetta = 0

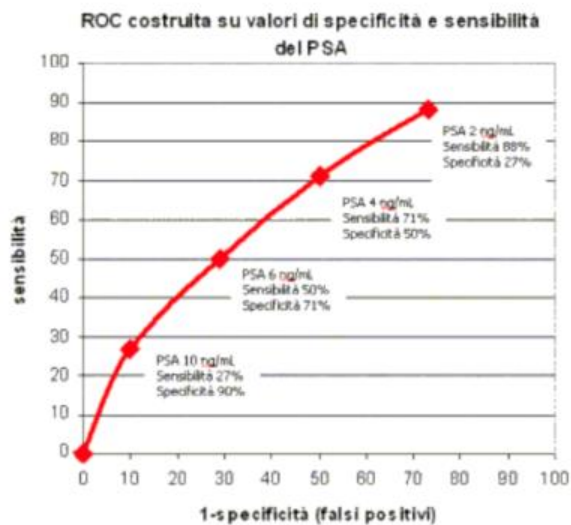
$\log L_1 \geq \log L_0 \rightarrow$ Tanto maggiore è la differenza \rightarrow tanto migliore la specificazione del modello

$$PSEUDO R^2 = 1 - \frac{\log L_1}{\log L_0}, \quad \text{che essendo logaritmi di probabilità } 0 \geq \log L_1 \geq \log L_0$$

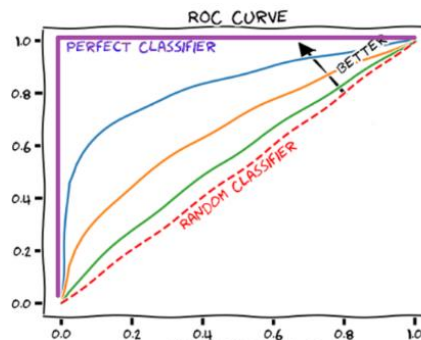
Per prevedere $y_i = 1$ o $y_i = 0$ ci si basa sulle probabilità stimate calcolate con il modello, definite da $F(\underline{x}_i' \underline{\beta})$. In generale si prevede $y_i = 1$ se $F(\underline{x}_i' \underline{\hat{\beta}}) > \frac{1}{2}$ che, con una distribuzione simmetrica attorno allo zero $F(0) = \frac{1}{2}$, ciò corrisponde a $\underline{x}_i' \underline{\hat{\beta}} > 0$. Per valutare l'adattamento ai dati costruiamo innanzitutto la matrice di confusione, dove ogni colonna della matrice rappresenta i valori predetti, mentre ogni riga rappresenta i valori reali. Una misura di adattamento si ottiene confrontando la percentuale di previsioni errate (ovvero i valori predetti 1 che sono 0 e i valori predetti 0 che sono 1) con la percentuale di previsioni errate di un modello che contiene solo l'intercetta (che prevede 1 per tutte le osservazioni se $\hat{p} = \frac{\#Unità=1}{\#UnitàTotali} > \frac{1}{2}$ e 0 altrimenti).

$$R_p = 1 - \frac{wr_1 \left(= \frac{n_{01} + n_{10}}{N} \right)}{wr_0 \left(= \begin{cases} 1 - \hat{p}, & \text{se } \hat{p} > 0,5 \\ \hat{p}, & \text{se } \hat{p} \leq 0,5 \end{cases} \right)}, \quad \text{con } \begin{cases} wr_1 = \% \text{ previsioni errate modello completo} \\ wr_0 = \% \text{ previsioni errate modello solo intercetta} \end{cases}$$

		\hat{y}_i		Totale
		0	1	
y_i	0	n_{00}	n_{01}	N_0
	1	n_{10}	n_{11}	N_1
Totale		n_0	n_1	N



		predetti		
		n'	p'	
Valori Reali	n	Veri negativi	Falsi positivi	N
	p	Falsi negativi	Veri positivi	P
totale		N'	P'	



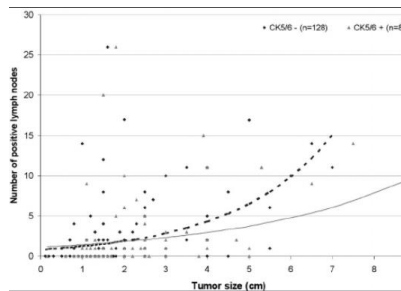
CURVA DI ROC

La curva di ROC è una curva disegnata su un diagramma cartesiano dove in ascissa viene posizionata $1 - \text{Specificità}$, dove $\text{Specificità} = \frac{\#VeriNegativi}{\#VeriNegativi + \#FalsiPositivi}$ e sull'asse delle ordinate la $\text{Sensibilità} = \frac{\#VeriPositivi}{\#VeriPositivi + \#FalsiNegativi}$. In altri termini con la

Sensibilità si indica la proporzione di pazienti con test positivo tra tutti quelli che hanno la malattia e con la Specificità la proporzione di pazienti con test negativo tra tutti quelli che sono sani (1-Specificità quindi indicherà la proporzione di pazienti con un test positivo tra tutti quelli che sono sani). Può anche essere pensato come un diagramma della potenza in funzione dell'errore di I tipo. Una curva di ROC è il grafico dell'insieme delle coppie (1-Specificità, Sensibilità) al variare di un parametro del classificatore. Per esempio, in un classificatore a soglia, si calcola la frazione di veri positivi e quella di falsi positivi per ogni possibile valore della

soglia; tutti i punti così ottenuti nello spazio FP-TP descrivono la curva di ROC. Il valore dell'accuratezza (ovvero della capacità del classificatore di discernere) si calcola attraverso l'area sottesa alla curva (maggiore è migliore è l'accuratezza).

MODELLI PER DATI DI CONTEGGIO: MODELLO DI POISSON



La regressione di poisson è appropriata quando la variabile dipendente è un contatore (es: numero chiamate in un ora).

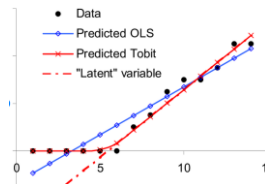
$$E(y_i|x_i) = e^{\underline{x}_i' \underline{\beta}} = \lambda_i \rightarrow \log(E(y_i|x_i)) = \underline{x}_i' \underline{\beta} \text{ (modello log - lineare)}$$

$$E(y_i|x_i) = \lambda_i \rightarrow P(y_i = y | \underline{x}_i) = \frac{e^{-\lambda_i} (\lambda_i)^y}{y!}, \quad \text{con } y = 0, 1, 2, \dots$$

Difetto è che implica automaticamente, $media = varianza = \lambda_i$, $V(y_i | \underline{x}_i) = e^{\underline{x}_i' \underline{\beta}} \rightarrow EQUIDISPERSIONE$

In molte applicazioni pratiche i dati rifiutano l'ipotesi di uguaglianza di media e varianza condizionata, per ovviare a questa problematica si utilizza l'approccio di quasi massima verosimiglianza, modificando la matrice di covarianza asintotica dello stimatore di massima verosimiglianza. L'approccio di quasi massima verosimiglianza è basato sulla funzione di verosimiglianza sbagliata, ma si può dimostrare la consistenza, dimostrando che $E(V'(\theta)) = 0$, ovvero che $\sum_{i=1}^N [y_i - e^{\underline{x}_i' \underline{\beta}}] \underline{x}_i = 0$. Posso interpretare la stima di massima verosimiglianza come una condizione sui momenti. Con questo approccio non posso calcolare le probabilità condizionate perché di fatto non sto assumendo Poisson.

MODELLO TOBIT



Variabile dipendente continua, ma intervallo formato dai suoi possibili valori è limitato. Ad esempio il consumo di sigaretta in funzione del reddito per alcuni individui potrebbe risultare negativo che non ha senso economico.

$$\begin{cases} y_i^* = \underline{x}_i' \underline{\beta} + \varepsilon_i, & \varepsilon_i \sim NID(0, \sigma^2) \\ \rightarrow y_i = y_i^*, & \text{se } y_i^* > 0 \\ = 0, & \text{se } y_i^* \leq 0 \end{cases} \rightarrow \text{TOBIT STANDARD, MODELLO REGRESSIONE CENSURATA}$$

Modello regressione censurata -> modello di regressione classico in cui tutti i valori negativi sono convertiti in zero. Viene innanzitutto calcolata la probabilità che il modello assuma valori negativi (ovvero che il modello venga troncato):

$$P(y_i = 0) = P(y_i^* \leq 0) = P(\varepsilon_i \leq -\underline{x}_i' \underline{\beta}) = P\left(\frac{\varepsilon_i}{\sigma} \leq -\underline{x}_i' \underline{\beta} / \sigma\right) = \Phi\left(-\underline{x}_i' \underline{\beta} / \sigma\right) = 1 - \Phi\left(\underline{x}_i' \underline{\beta} / \sigma\right)$$

7 SERIE STORICHE

Serie storica di osservazioni su qualche variabile, come per esempio il tasso di disoccupazione indicandola con y_1, \dots, y_T . Queste osservazioni sono viste come realizzazioni di variabili casuali descritte da qualche processo stocastico.

PROCESSO STOCASTICO

Dato uno spazio di probabilità (Ω, \mathcal{F}, P) e uno spazio misurabile $(\mathcal{B}, \mathbb{R})$, un processo stocastico è una collezione di variabili aleatorie $\{Y(t, \omega): t \in T, \omega \in \Omega\}$.

SERIE STORICA

Una serie storica è una realizzazione finita e parziale del processo stocastico.

PERSISTENZA ED ERGODICITA'

La **persistenza** (caratteristica per cui la serie storica ha memoria di sé che sostanzialmente è espressa dall'autocorrelazione) è quella caratteristica della serie storica per la quale l'osservazione al tempo t è più simile all'osservazione al tempo $t-1$ rispetto ad una osservazione ancora più lontana. Un processo è **ergodico**, euristicamente, se eventi "molto" lontani fra loro possono essere considerati "virtualmente" indipendenti; la ergodicità quindi limita la persistenza. In una serie storica non ergodica la persistenza è così forte che non sarebbe possibile fare inferenza.

PROCESSO STAZIONARIO

Processo STAZIONARIO in senso stretto $\rightarrow y_1, \dots, y_T$ stessa distribuzione

Detto in maniera più formale, sia $\{Y_t\}$ un processo stocastico e $F_Y(y_{t_1+\tau}, \dots, y_{t_k+\tau})$ rappresenti la funzione cumulativa della distribuzione congiunta di $\{Y_t\}$ negli istanti $t_1 + \tau, \dots, t_k + \tau$. Allora si dice che $\{Y_t\}$ è un processo fortemente stazionario se, per ogni k , per ogni τ , e per ogni t_1 vale: $F_Y(y_{t_1+\tau}, \dots, y_{t_k+\tau}) = F_Y(y_{t_1}, \dots, y_{t_k})$.

$$\text{Processo STAZIONARIO in senso debole} \rightarrow \begin{cases} E(y_t) = \mu < \infty \\ V(y_t) = E[(y_t - \mu)^2] = \gamma_0 < \infty \forall t \\ \text{Cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \quad k = 1, 2, \dots \end{cases}$$

Il processo stazionario implica che media, varianza e covarianza sono indipendenti dal tempo, con covarianze che dipendono solo dalla lunghezza dell'intervallo che separa le due osservazioni.

$$\text{PROCESSO DI MARKOV} \rightarrow P(Y_{t+1}|Y_t, \dots, Y_1) = P(Y_{t+1}|Y_t)$$

MODELLO ARMA

Parametrizzazione efficiente di un processo stocastico stazionario debole, realizzato sulla base di una serie storica.

WHITE NOISE

Un white noise è un processo stocastico composto da variabili aleatorie di media nulla, varianza costante, incorrelate tra loro (assenza di persistenza): $E(WN) = 0, Var(WN) = \sigma^2, Cov(WN_t, WN_{t-k}) = 0 \forall k$.

. A rigore questo non significa che esse siano indipendenti; se si parla di White Noise Gaussiano, ovvero di un white noise in cui la distribuzione congiunta di tutte le coppie $(\varepsilon_t, \varepsilon_{t+k})$ sia una normale bivariata, allora si (indipendenza e incorrelazione di distribuzione marginali normali di una congiunta normale multivariata sono la stessa cosa).

TEOREMA DI WOLD

Ogni processo stazionario debole può essere scritto come la somma di un processo MA infinito più una serie deterministica.

$$y_t = \sum_{j=0}^{+\infty} \alpha_j \varepsilon_{t-j} + \eta_t, \quad \sum_{j=0}^{+\infty} \alpha_j^2 < +\infty, \quad \eta_t = \text{comb. lin. di seno e coseno}$$

AUTOCORRELAZIONE E AUTOCORRELAZIONE PARZIALE

$$\text{AUTOCORRELAZIONI, ACF (AutoCorrelationFunction)} \rightarrow \rho_k = \frac{Cov(y_t, y_{t-k})}{V(y_t)} = \frac{\gamma_k}{\gamma_0}$$

$$\text{PACF (PartialAutoCorrelationFunction)} \rightarrow P_k = Corr(Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1})$$

Nel contesto delle serie storiche, buona parte della correlazione tra Y_t e Y_{t-k} può essere dovuta alla correlazione che tali variabili hanno con $Y_{t-1}, \dots, Y_{t-k+1}$. Un possibile modo per tener conto di ciò è considerare la funzione di autocorrelazione parziale, che misura l'autocorrelazione tra Y_t e Y_{t-k} al netto delle variabili intermedie.

Posso interpretare l'autocorrelazione di ordine k di una serie storica come il coefficiente del termine k-esimo in una sua scrittura in forma AR. Posso scrivere MA come AR infinito e infatti MA hanno PACF che tendono a zero per $n \rightarrow \infty$.

OPERATORE RITARDO

$$L = \text{Lag (o B = Backshift)}: L c y_t = c y_{t-1}; L^p c y_t = c y_{t-p}; L^{-1} c y_t = c y_{t+1}; L^0 = 1;$$

$$(1 - \theta L)^{-1} = \sum_{j=0}^{+\infty} \theta^j L^j; \theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_p L^p,$$

$$(1 + \alpha L)^{-1} = \sum_{j=0}^{+\infty} (-\alpha)^j L^j; \alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$$

RANDOM WALK

y_t è random walk se Δy_t è un white noise ($y_t = y_{t-1} + \varepsilon_t$). Non condivide con i processi stazionari la caratteristica di essere mean reverting (ovvero la caratteristica a muoversi preferenzialmente verso il suo valore atteso).

GENESI DELL'ARMA

Supponiamo che il processo sia stazionario, grazie al teorema di Wold, esso può essere scritto come un $MA(\infty)$. Supponiamo che, per la proprietà per cui $\sum_{j=0}^{+\infty} \alpha_j^2 < +\infty$, i coefficienti tenderanno ad annullarsi, o comunque a diventare trascurabili. Supponiamo di conoscere le autocovarianze, che tenderanno anche esse ad annullarsi da un certo punto in poi (o a diventare trascurabili).

È possibile identificare i parametri di una eventuale rappresentazione $MA(q)$? No, (a meno di supporre il WN gaussiano) perché dati q coefficienti di autocorrelazione si dimostra che esistono 2^q processi $MA(q)$.

L'identificazione è possibile soltanto introducendo un'ipotesi ulteriore, ovvero che il processo sia invertibile, cioè scrivibile in funzione della sua storia passata, ovvero che sia possibile scrivere il processo come un $AR(\infty)$.

Tuttavia anche in questo caso i pesi del modello $AR(\infty)$ tenderanno a zero, o meglio sarà possibile ipotizzarli inferiori ad un prefissato valore positivo da un certo momento in poi (il passato influenza sempre meno il presente) e non è restrittivo porli uguale a zero in modo da ottenere un $AR(p)$. Un $AR(p)$ è stazionario se le soluzioni del polinomio caratteristico sono fuori dal cerchio unitario.

Infine, si deve citare un importante risultato che, in modo informale, dice che se io devo approssimare un $MA(\infty)$ con il minor numero di parametri possibile sarà sempre più efficiente cercare questa approssimazione combinando tra loro le rappresentazione MA e AR piuttosto che concentrarsi solo sugli MA o gli AR .

DEFINIZIONI ARMA(p, q) TRADIZIONALI

$$\begin{aligned} \text{ARMA}(p, q) \rightarrow {}_c y_t &= \theta_1 {}_c y_{t-1} + \dots + \theta_p {}_c y_{t-p} + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q} + \varepsilon_t, & \text{RISCRITTURA con } L \rightarrow \theta(L) {}_c y_t &= \alpha(L) \varepsilon_t \\ & \text{RISCRITTURA come } MA(\infty) \rightarrow {}_c y_t = \theta(L)^{-1} \alpha(L) \varepsilon_t, & \theta(L) \text{ invertibile} \\ & \text{RISCRITTURA come } AR(\infty) \rightarrow \alpha(L)^{-1} \theta(L) {}_c y_t = \varepsilon_t, & \alpha(L) \text{ invertibile} \\ \text{AR}(p) \rightarrow {}_c y_t &= \theta_1 {}_c y_{t-1} + \dots + \theta_p {}_c y_{t-p} + \varepsilon_t, & \text{RISCRITTURA con } L \rightarrow \theta(L) {}_c y_t &= \varepsilon_t \\ \text{MA}(q) \rightarrow {}_c y_t &= \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q}, & \text{RISCRITTURA con } L \rightarrow {}_c y_t &= \alpha(L) \varepsilon_t \end{aligned}$$

AR(1)

$$\begin{aligned} y_t &= \delta + \theta y_{t-1} + \varepsilon_t, & \varepsilon_t \text{ WN, } |\theta| < 1, \text{ stazionario debole,} & \text{RISCRITTURA } {}_c y_t = \theta {}_c y_{t-1} + \varepsilon_t, & {}_c y_t = y_t - \mu \\ \text{STAZIONARIO} \rightarrow \mu &= E[y_t] = \frac{\delta}{1 - \theta}, & V(y_t) = \frac{\sigma^2}{1 - \theta^2}, & \text{Cov}(y_t, y_{t-k}) = \theta^k \frac{\sigma^2}{1 - \theta^2} \rightarrow \text{non dipende da } t \end{aligned}$$

$$\left\{ \begin{aligned} & \text{RISCRITTURA } MA(\infty) \text{ con } L \rightarrow {}_c y_t = \theta {}_c y_{t-1} + \varepsilon_t \rightarrow (1 - \theta L) {}_c y_t = \varepsilon_t \rightarrow (1 - \theta L)(1 - \theta L)^{-1} {}_c y_t = (1 - \theta L)^{-1} \varepsilon_t \\ & {}_c y_t = \sum_{j=0}^{+\infty} \theta^j L^j \varepsilon_t \rightarrow {}_c y_t = \sum_{j=0}^{+\infty} \theta^j \varepsilon_{t-j} \rightarrow MA(\infty) \end{aligned} \right.$$

MA(1)

$$y_t = \mu + \alpha \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ WN}, \quad \text{RISCRITTURA } {}_c y_t = \alpha \varepsilon_{t-1} + \varepsilon_t, \quad {}_c y_t = y_t - \mu$$

$$\text{STAZIONARIO} \rightarrow \mu = E[y_t], \quad V(y_t) = (1 + \alpha^2)\sigma^2, \quad \text{Cov}(y_t, y_{t-1}) = \alpha\sigma^2, \quad \text{Cov}(y_t, y_{t-k}) = 0, k = 2, 3, \dots$$

Processo a media mobile \rightarrow osservazioni lontane più di 2 periodi sono incorrelate

$$\left\{ \begin{array}{l} \text{RISCRITTURA AR}(\infty) \text{ con } L \rightarrow {}_c y_t = \varepsilon_t + \alpha \varepsilon_{t-1} \rightarrow {}_c y_t = (1 + \alpha L)\varepsilon_t \rightarrow (1 + \alpha L)^{-1} {}_c y_t = (1 + \alpha L)^{-1} \varepsilon_t \\ \varepsilon_t = \sum_{j=0}^{+\infty} (-\alpha)^j {}_c y_{t-j} \rightarrow {}_c y_t = \alpha \sum_{j=0}^{+\infty} (-\alpha)^j {}_c y_{t-j-1} + \varepsilon_t \rightarrow \text{AR}(\infty) \end{array} \right.$$

PREVISIONE

Il predittore ottimo è $y_{T+h|T} = E(y_{T+h} | I_T (= Y_T, \dots, Y_{-\infty}))$. Le previsioni di un passo in avanti su processi del tipo $y_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q} + \varepsilon_t$ si fanno così: $\hat{y}_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q} + 0$ ovvero utilizzando i valori effettivamente osservati delle y_t e i valori degli errori di previsione passati al posto delle ε_{t-1} . L'errore nella previsione $C_h = E[({}_c y_{T+h} - {}_c y_{T+h|T})^2] = E[({}_c y_{T+h} - E({}_c y_{T+h} | I_T))^2] = V({}_c y_{T+h} | I_T)$. Per un processo MA(∞) vale che $C_h = E[({}_c y_{T+h} - {}_c y_{T+h|T})^2] = \sigma^2 \sum_{j=0}^{h-1} \alpha_j^2$.

INVERTIBILITA' POLINOMI RITARDO

$$\left\{ \begin{array}{l} \text{INVERTIBILITA' } 1 - \theta L \rightarrow \text{polinomio 1ordine invertibile se } |\theta| < 1 \\ \text{INVERTIBILITA' } 1 - \theta L - \theta_2 L^2 \rightarrow \text{polinomio 2ordine invertibile se } \exists \phi_1, \phi_2: 1 - \theta L - \theta_2 L^2 = (1 - \phi_1 L)(1 - \phi_2 L) \\ \rightarrow \phi_1, \phi_2: \begin{cases} \phi_1 + \phi_2 = \theta_1 \\ -\phi_1 * \phi_2 = \theta_2 \end{cases} \rightarrow \text{CONDIZIONI INVERTIBILITA' } |\phi_1| < 1, |\phi_2| < 1 \end{array} \right.$$

$$\text{EQUAZIONE CARATTERISTICA} \rightarrow 1 - \theta z - \theta_2 z^2 = (1 + \phi_1 z)(1 + \phi_2 z) = 0 \rightarrow z_i = \frac{1}{\phi_i} \rightarrow |\phi_i| < 1 \Rightarrow |z_i| > 1$$

Le radici caratteristiche z_i devono essere maggiori di 1 in modulo.

$$\text{RADICE UNITARIA} \rightarrow |z_i| = 1 \rightarrow \sum_{j=1}^p \theta_j = 1 \rightarrow \text{Se la somma supera 1 non è invertibile}$$

Per processi MA, stazionari per definizione, l'invertibilità del polinomio è importante per la stima e per la previsione.

Per processi AR il polinomio è invertibile se e solo se è stazionario.

STAZIONARIETA' E RADICI UNITARIE

Tabella 2.1: Processo stocastico, stazionario e invertibile

Processo	Stazionario	Invertibile
AR(p)	$\phi(B) = 0$ $ B_i > 1$	Sempre
MA(q)	Sempre	$\theta(B) = 0$ $ B_i > 1$
ARMA(p, q)	$\phi(B) = 0$ $ B_i > 1$	$\theta(B) = 0$ $ B_i > 1$

Una serie che diventa stazionaria in seguito al passaggio alle differenze prime è detta integrata di ordine uno, e indicata con I(1). Se $\Delta_c y_t$ è rappresentata da un modello ARMA(p, q) stazionario, diremo che y_t segue un modello autoregressivo integrato a media mobile (ARIMA, Auto Regressive Integrated Moving Average) di ordine p,1,q ovvero un ARIMA(p, 1, q).

$\left\{ \begin{array}{l} \text{SERIE INTEGRATA DI ORDINE 1 (I(1))} \rightarrow \text{serie diventa stazionaria dopo un passaggio alle differenze prime} \\ \text{SERIE INTEGRATA DI ORDINE 2 (I(2))} \rightarrow \Delta(\Delta_c y_t) \text{ è ARMA (p, q)} \end{array} \right.$
 $\left\{ \begin{array}{l} \text{I(0)} \rightarrow \text{STAZIONARIA} \rightarrow \text{Ad esempio una serie AR(1), oppure un WN} \rightarrow \text{memoria limitata e shock economico si ferma} \\ \text{I(1)} \rightarrow \text{Ad esempio AR(1) con } \theta = 1 \rightarrow \text{memoria illimitata e shock economico è permanente} \end{array} \right.$
 $\Delta y_t = \delta + \theta y_{t-1} + \gamma t + \varepsilon_t, \quad |\theta| < 1, \gamma \neq 0 \rightarrow \text{PROCESSO STAZIONARIO ATTORNO AD UN TREND}$
 $\Delta y_t = \delta + \varepsilon_t \rightarrow \text{RANDOM WALK CON DRIFT}$

TEST DI DICKEY AND FULLER

$$y_t = \delta + \theta y_{t-1} + \varepsilon_t, \quad \text{stimo } \theta \text{ con OLS e verifico stazionarietà con Dickey and Fuller } DF = \frac{\hat{\theta} - 1}{\text{se}(\hat{\theta})}$$

Anche se apparentemente la statistica test di DF è la classica statistica test del test t, in realtà (poiché l'ipotesi nulla è di non stazionarietà) la distribuzione della statistica DF non è normale ed è stata tabulata in apposite tabelle.

TEST KPSS

Nel test KPSS l'ipotesi nulla è di stazionarietà attorno ad un trend mentre l'esistenza di una radice unitaria costituisce l'alternativa. Se y_t fosse stazionario attorno ad un trend deterministico allora una regressione tipo $y_t = \beta_0 + \beta_1 t + u_t$ dovrebbe produrre residui I(0). Fatta la regressione, si prendono i residui OLS e si cumulano, producendo una nuova serie $S_t = \frac{1}{T} \sum_{s=1}^t \hat{u}_s$. La somma dei quadrati di S_t opportunamente normalizzata converge in distribuzione ad una variabile casuale che è sempre la stessa per qualunque processo stazionario. Se y_t è non stazionario la statistica diverge.

STIMA DI MODELLI ARMA MINIMI QUADRATI

Per i modelli AR la stima dei minimi quadrati fornisce stime consistenti in quanto il termine d'errore White Noise è incorrelato con qualsiasi variabile osservata in t-1 o a qualche data precedente.

MINIMI QUADRATI AR(p) $\rightarrow y_t = \delta + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t \rightarrow (\varepsilon_t \text{ WN}; E[y_{t-j}\varepsilon_t] = 0) \rightarrow \text{OLS classico}$

Per i modelli MA la stima dei minimi quadrati non si può applicare senza prima trasformare il modello in un modello AR, in quanto altrimenti si rischierebbe di minimizzare un termine di errore che dipende da un altro termine di errore

MINIMI QUADRATI MA(1) $\rightarrow y_t = \mu + \varepsilon_t + \alpha_1 \varepsilon_{t-1} \rightarrow \text{minimizzo } \varepsilon_t = y_t - \mu - \alpha_1 \varepsilon_{t-1} \rightarrow \text{non conosco } \varepsilon_{t-1} \rightarrow \text{AR}(\infty)$

STIMA DI MODELLI ARMA LOG-VEROSIMIGLIANZA

Un approccio alternativo alla stima dei modelli ARMA è rappresentato dalla massima verosimiglianza. Devo formulare un'ipotesi sulla distribuzione di ε_t , che di solito è quella di normalità.

SCELTA TRA PIU' MODELLI - AIC, BIC

Per un modello ARMA(p,q) calcolo l'AIC e il BIC al variare di p e q, scegliendo quello più basso e privilegiando il BIC perché se $T \rightarrow \infty$ esso seleziona il vero modello. Nel seguito $\hat{\sigma}^2$ rappresenta la varianza stimata di ε_t

$$\text{AIC} = \log \hat{\sigma}^2 + 2 \frac{(p+q+1)}{T}, \quad \text{BIC} = \log \hat{\sigma}^2 + \frac{(p+q+1)}{T} \log T$$

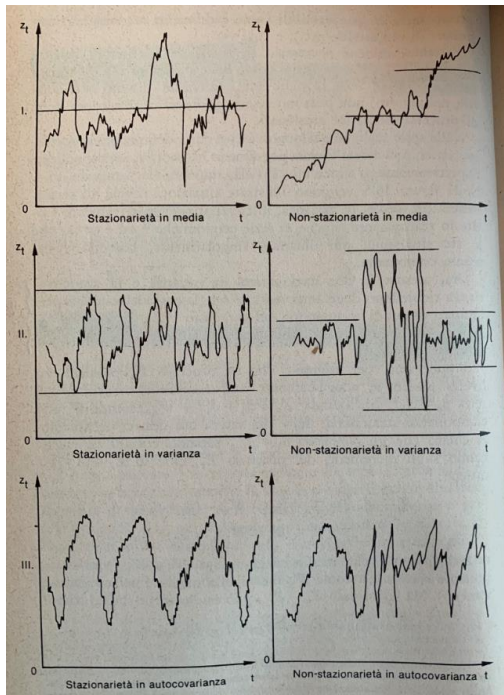
Trade-off diversi di "adattamento", misurato dal valore della verosimiglianza (della varianza degli errori calcolati su \hat{y}_t misurata con la verosimiglianza), e "parsimonia" misurata dal numero di parametri liberi, $p+q+1$.

PROCESSI SARMA

Un processo *SARMA* serve per descrivere la stagionalità: in pratica è analogo ad un processo ARMA con dei "buchi". Ad esempio ipotizzando una stagionalità di 12 mesi – serie storica delle presenze alberghiere di Riccione – ed ipotizzando di aver inizialmente scelto un ARMA(1,0), invece che utilizzare un ARMA(12,0), per tenere conto che il dato di Agosto di un anno assomiglia a quello di Agosto dell'anno precedente, utilizzo un *SARMA*(1,0) \times (1,0)₁₂ ovvero aggiungendo al mio processo ARMA un operatore ritardo, sia lato AR che lato MA, pari a 12 mesi. Bisogna fare attenzione che il polinomio caratteristico si modifica: $\theta(L) * B(L) = (1 - \theta L)(1 - BL^s) = 1 - \theta L - BL^s + \theta BL^{s+1}$ con s =sottoperiodo.

PROCEDURA DI BOX E JENKINS

La procedura di Box e Jenkins consta di 5 fasi fondamentali:



ANALISI PRELIMINARE

Essa è finalizzata a capire se la serie storica è stazionaria o no. Per fare ciò posso guardare il grafico oppure la funzione di autocorrelazione.

- Se la serie storica non è stazionaria in media essa si può differenziare (se differenzio una volta lavoro su incrementi assoluti, profit e loss per serie economiche, e sono ancora in grado di fornire un senso economico alla differenziazione).
- Se la serie storica non è stazionaria in varianza posso eseguire una trasformazione di Box e Cox

$$y_t(\lambda) = \begin{cases} \frac{y_t - 1}{\lambda}, & \lambda \neq 0 \\ \ln y_t, & \lambda = 0 \end{cases}$$

λ viene scelto con una apposita procedura. La cosa più favorevole è che differenzio una volta e che $\lambda = 0$ perché in questo caso riesco a dare un senso economico $\Delta \ln y_t$ in quanto, in questo caso, starei lavorando sui log-return, ovvero starei modellizzando le varianze percentuali.

- Nel caso di outlier esistono procedure apposite che si chiamano "intervention analysis"
- Devo infine verificare che i residui sono gaussiani perché utilizzerò questa informazione nella stima.

Tabella 2.2: ACF e PACF per ciascun processo analizzato

Processo stocastico, stazionario e invertibile	ACF	PACF
$AR(p)$	Non si annulla mai ma decade verso lo 0	È nulla per $k > p$
$MA(q)$	È nulla per $k > q$	Non si annulla mai ma decade verso lo 0
$ARMA(p, q)$	Non si annulla mai ma decade verso lo 0 per i lags $k > p - q$	Non si annulla mai ma decade verso lo 0 per i lags $k > q - p$

IDENTIFICAZIONE

Una volta stabilito che la mia serie è stazionaria devo identificare p, q .

1. Metodo dell'analogia. Vedi tabella a fianco. Si presta bene se il processo è solo MA o solo AR.
2. Metodo di identificazione automatica. Si prova a cercare una funzione che si azzeri in corrispondenza di p, q in un processo $ARMA(p, q)$. Ad esempio metodo dell'angolo o metodo R&S Array.
3. Identificazione automatica. Per stimare i parametri p e q di un modello ARMA si utilizza AIC e BIC

STIMA

Si può utilizzare la massima verosimiglianza o il metodo dei minimi quadrati con le accortezze riportate sopra.

CONTROLLO

Bisogna verificare che i residui del modello si comportano come White Noise. Se ho utilizzato la massima verosimiglianza essi dovrebbero comportarsi come un WN gaussiano. Per fare ciò si possono utilizzare diversi approcci:

1. Analisi grafica – Guardo il grafico dei residui: essi dovrebbero disporsi in maniera casuale.
2. Autocorrelazione dei residui – Deve essere nulla: trattando la serie dei residui e_t come una serie storica a sé stante, calcolo la funzione di autocorrelazione empirica. Per ogni valore di k è possibile utilizzare $\hat{\rho}_k$ per verificare se esso è diverso da zero (un white noise ha autocorrelazione tutti uguali a zero).
3. Test di Ljung Box – Verifica l'ipotesi nulla (i dati sono indipendentemente distribuiti) contro alternativa generica essendo questo un test di portmanteau (l'ipotesi alternativa non è bene specificata). La regione di rifiuto è:

$$Q = T(T + 2) \sum_{k=1}^h \frac{\hat{\rho}_k}{T - k} > \chi^2_{1-\alpha, h}, \quad h = \text{numero di lag testati}$$

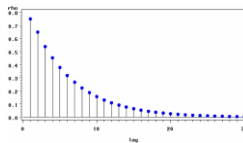
4. Test di normalità di Jarque e Bera – Si testa se i residui sono distribuiti normalmente. Nel caso di gaussianità infatti l'incorrelazione dei residui implica anche la loro indipendenza. Come statistica test si utilizza la statistica di JB basata sulla simmetria e la curtosi: $JB = \frac{T}{6} * \left[\text{skewness}^2 + \frac{1}{4}(\text{curtosi} - 3)^2 \right]$

Se i residui si comportano come un $ARMA(r, s)$ vuol dire che il modello giusto è un modello $ARMA(p + r, q + s)$ e quindi devo verificare nuovamente che i residui sono un WN dopo aver stimato un modello $ARMA(p + r, q + s)$. Dopo tante eventuale reiterazione, quando alla fine i residui sono un WN, ho la struttura corretta del modello.

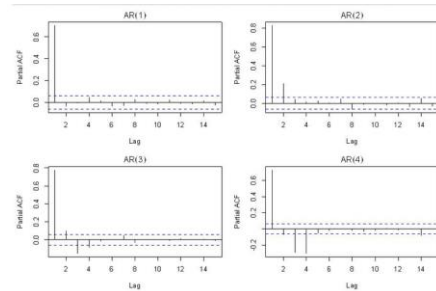
PREVISIONE

Più l'orizzonte di previsione è lontano più la previsione stima semplicemente la media del processo. Ad esempio in un processo $AR(1)$ con media zero, la funzione di previsione per $t + k$, ipotizzando di essere nell'istante t , è θ^k che, essendo minore di uno in modulo per la stazionarietà, riporta la previsione alla media di lungo periodo.

MODELLO AR(P) – ACF, PACF

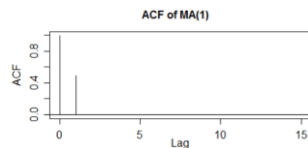


Nei modelli AR(p) la ACF tende a zero in maniera decrescente.
Bisogna fare attenzione perché il grafico dell'ACF parte dal punto 0.



L'autocorrelazione parziale (il grafico parte dal punto 1) è diversa da zero ("esce" fuori dalla banda entro la quale si trovano gli errori casuali) per un numero di volte iniziali (a sinistra del grafico) pari all'ordine del processo. Nella figura accanto AR(1) ha una funzione di autocorrelazione parziale solo una volta diversa da zero, così come AR(2) ha una PACF solo due volte diversa da zero etc.

MODELLI MA(q) – ACF, PACF



Nei modelli MA(q) la ACF è diversa da zero per un numero di volte iniziali pari all'ordine del processo. Attenzione, partendo da 0, il grafico a sinistra ha solo una barretta verticale, cioè un solo valore diverso da zero.



Nei modelli MA(q) la PACF tende a zero in maniera decrescente (Attenzione che decresce esponenzialmente in modulo: ovvero decresce intorno allo zero sia in alto che in basso).

MODELLI ARMA (p,q)

ACF decresce esponenzialmente
PACF decresce esponenzialmente in modulo

ARCH/GARCH

Clustering della volatilità - a periodi di alta volatilità seguono periodi di alta volatilità e viceversa

ARCH (Auto Regressive Conditional Heteroskedasticity) ovvero eteroschedasticità condizionale autoregressiva. La varianza dei termini del termine d'errore alla data t dipende dal quadrato dei termini d'errore nei periodi precedenti. Potrei essere in un mondo ARMA ed interessarmi della varianza condizionata.

Dato un processo $\{\varepsilon_t\}_t$ per i rendimenti di un titolo, si ipotizza $\varepsilon_t = \sigma_t z_t$ con $z_t \sim N(0,1)$ e σ_t^2 segue un processo AR:

$$ARCH(1) \rightarrow \sigma_t^2 = V(\varepsilon_t | I_{t-1}) = E(\varepsilon_t^2 | I_{t-1}) = \omega + \alpha \varepsilon_{t-1}^2, \quad \omega \geq 0, \alpha \geq 0 \text{ per garantire } \sigma_t^2 \geq 0$$

Questa specificazione non implica che il processo di ε_t sia non stazionario, ma stabilisce solo che i quadrati delle innovazioni $\varepsilon_t^2, \varepsilon_{t-1}^2$ sono correlati. Il modello ARCH(1) stabilisce che se nel periodo $t-1$ si verifica uno shock elevato, anche ε_t tende ad avere un valore (assoluto) elevato. In altre parole, quando ε_{t-1}^2 è alto, anche la varianza dell'innovazione successiva, ε_t , è alta.

La varianza non condizionale invece non dipende da t :

$$\sigma_t^2 = E(\varepsilon_t^2) = \omega + \alpha E(\varepsilon_{t-1}^2), \quad \text{a cui è associata la soluzione stazionaria } \sigma^2 = \frac{\omega}{1-\alpha}, \quad 0 < \alpha < 1$$

$$ARCH(p) \rightarrow \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2, \quad \text{shock prima di } p \text{ non hanno effetto}$$

Presenza di errori ARCH nel modello non comporta perdita di validità di OLS. Per verificare la presenza di eteroschedasticità condizionale autoregressiva di ordine p (analogamente a Breusch-Pagan) stimo regressione al quadrato e_t^2 sui residui $e_{t-1}^2, \dots, e_{t-p}^2$ più una costante e calcolo prodotto TR^2 . Sotto ipotesi nulla di omoschedasticità $\alpha_1 = \dots = \alpha_p = 0$ si comporta come Chi-Quadrato con p gradi libertà. In altre parole, verificare l'omoschedasticità rispetto all'alternativa che gli errori seguano un processo ARCH(p) è molto semplice.

$$GARCH(p, q) \rightarrow \sigma_t^2 = \omega + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

$$GARCH(1,1) \rightarrow \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$GARCH(1,1) \rightarrow (v_t = \varepsilon_t^2 - \sigma_t^2) \rightarrow \varepsilon_t^2 = \omega + (\alpha + \beta) \varepsilon_{t-1}^2 + v_t - \beta v_{t-1} \rightarrow ARMA(1,1)$$

$$GARCH(1,1) \rightarrow \text{sostituendo ricorsivamente} \rightarrow \sigma_t^2 = \frac{\omega}{1-\beta} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} \varepsilon_{t-j}^2 \rightarrow ARCH(\infty)$$

SCOMPOSIZIONE DI BEVERIDGE E NELSON

Qualsiasi processo $I(1)$ può essere pensato come la somma di un random walk e di un processo $I(0)$. Questa scomposizione è nota come scomposizione BN e può essere derivata da una proprietà dei polinomi.

8 SERIE STORICHE MULTIVARIATE

Se in una serie storica la storia di una seconda variabile x_t può aiutare a prevedere i valori futuri di y_t abbiamo un modello autoregressivo a ritardi distribuiti semplici (sistema dinamico). Ad esempio prezzo del petrolio passato può aiutare a prevedere il consumo di carburanti). Per poter applicare le procedure classiche di stima o verifica di ipotesi in un modello dinamico in serie storiche è necessario che le diverse variabili siano stazionarie. L'uso di variabili non stazionarie non comporta necessariamente la perdita di validità degli stimatori classici, se le due variabili sono cointegrate. Infatti, se due variabili sono cointegrate possono essere scritte con un modello di correzione di errore, ovvero si può assumere tra le variabili una relazione dinamica.

FUNZIONE DI RISPOSTA DI IMPULSO

La funzione di risposta di impulso è data semplicemente dai coefficienti della rappresentazione MA del processo (interpretabile come l'effetto che lo shock avvenuto i periodi addietro ha sul valore attuale di y): $b_i = \frac{\partial y_t}{\partial \varepsilon_{t-i}} = \frac{\partial y_{t+i}}{\partial \varepsilon_t}$

Mentre nei modelli MA leggiamo la funzione di risposta di impulso come uno shock casuale, accaduto in $t - s$ e che influenza quello che succede a t , nei processi VMA poiché più variabili $\underline{y}_t = (y_{1t}, \dots, y_{pt})'$ sono influenzate da uno stesso set di errori $\underline{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{qt})$ essi sono interpretabili come "shock strutturali", interpretabili sulla base delle covarianze.

COINTEGRAZIONE

Se y_t, x_t sono serie integrate di ordine 1 $I(1)$ ed esiste un β : $z_t = y_t - \beta x_t \in I(0)$ allora y_t, x_t si dicono cointegrate.

Lo stimatore OLS b di β in questo caso è superconsistente, cioè converge ad un tasso molto più elevato al parametro β rispetto alla teoria asintotica classica. Infatti se il modello è $y_t = a - bx_t + e_t$ la serie per tutti i $b \neq \beta$ sarà $I(1)$ e avrà varianza elevata mentre per $b = \beta$ la varianza sarà bassa. Poiché l'approccio OLS si basa sulla minimizzazione della varianza questa proprietà aiuta ad individuare la stima.

Poiché y_t, x_t sono $I(1)$ e dominate dalle rispettive tendenze di lungo periodo, affinché z_t sia $I(0)$ le tendenze di lungo periodo di y_t, x_t si devono elidere fra loro, ovvero le due variabili devono condividere lo stesso trend stocastico.

RELAZIONE SPURIA

Se y_t, x_t sono due Random Walk $I(1)$ $y_t = y_{t-1} + \varepsilon_t, x_t = x_{t-1} + \varepsilon_t$ allora la relazione $y_t = \alpha - \beta x_t + \varepsilon_t$ è una Regressione Spuria. Il termine di errore in questo caso è $I(1)$, in particolare è autocorrelato. y_t, x_t contengono entrambe un trend stocastico e lo stimatore OLS tende a rilevare la presenza di una correlazione statisticamente significativa.

GRANGER CASUALITA'

X_t, Y_t siano due serie storiche. X_t è Granger Causale Y_t se $E(Y_t | Y_{t-1}, Y_{t-2}, \dots, X_{t-1}, X_{t-2}) \neq E(Y_t | Y_{t-1}, Y_{t-2}, \dots)$.

DISTINZIONE TRA COINTEGRAZIONE E RELAZIONE SPURIA

Se voglio sapere se fra y_t, x_t esiste una relazione di cointegrazione o una di regressione spuria. Supponiamo di sapere che entrambe sono $I(1)$ allora scrivo $y_t = \alpha + \beta x_t + \varepsilon_t$ e verifico se il termine d'errore è $I(0)$ (cointegrazione) oppure $I(1)$ (relazione spuria).

Stimo la regressione sui residui $\Delta e_t = \gamma_0 + \gamma_1 e_{t-1} + u_t$ e verifico se $\gamma_1 = 0$ (caso di radice unitaria).

Posso utilizzare il **Test di Durbin-Watson di Regressione di Cointegrazione** (CRDW) basato sulla classica statistica di Durbin Watson. Infatti asintoticamente la presenza di una radice unitaria in ε_t corrisponde ad un valore nullo della statistica dw: $2 - 2\hat{\rho}$ con $\hat{\rho} = \hat{\gamma}_1$.

I test di cointegrazione verificano la presenza di una radice unitaria nei residui di una regressione. Ciò implica che l'ipotesi nulla di radice unitaria corrisponde all'assenza di cointegrazione. Un mancato rifiuto della presenza di una radice unitaria nei residui OLS, dunque, implica che non è possibile escludere che y_t, x_t non siano cointegrate.

MODELLO AUTOREGRESSIVO A RITARDI DISTRIBUITI SEMPLICE

x_t, y_t stazionario, ad esempio $y_t = \text{fatturato}, x_t = \text{investimento pubblicitario}$

$$y_t = \delta + \theta y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN \text{ indipendente da } x_t, x_{t-1}, y_t, |\theta| < 1$$

Per stimare i parametri ricorro ai minimi quadrati ordinari

$$\frac{\partial y_t}{\partial x_t} = \phi_0 \text{ moltiplicatore di impatto}, \quad \frac{\partial y_{t+1}}{\partial x_t} = \theta \frac{\partial y_t}{\partial x_t} + \phi_1 = \theta \phi_0 + \phi_1, \quad \frac{\partial y_{t+2}}{\partial x_t} = \theta(\theta \phi_0 + \phi_1)$$

$$\phi_0 + (\theta \phi_0 + \phi_1) + \theta(\theta \phi_0 + \phi_1) + \dots = \phi_0 + (1 + \theta + \theta^2 + \dots) * (\theta \phi_0 + \phi_1) = \frac{\phi_0 + \phi_1}{1 - \theta}, \quad \text{moltiplicatore lungo periodo}$$

MODELLO A CORREZIONE DELL'ERRORE SEMPLICE

Se sottraggo y_{t-1} ad entrambi i membri a $(y_t = \delta + \theta y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \varepsilon_t)$

$$\Delta y_t = \delta - (1 - \theta)y_{t-1} + \phi_0 \Delta x_t + (\phi_0 + \phi_1)x_t + \varepsilon_t \Leftrightarrow \Delta y_t = \phi_0 \Delta x_t - (1 - \theta)(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

In questo modello la variazione di y è determinata dalla variazione contemporanea di x più un termine di correzione dell'errore, con $(1 - \theta)$ chiamata velocità di aggiustamento.

TEOREMA DI RAPPRESENTAZIONE DI GRANGER

Se alcune variabili sono cointegrate, esiste una rappresentazione a correzione dell'errore valida dei dati.

y_t, x_t sono $I(1) \rightarrow$ vettore di cointegrazione $(1 - \beta)'$

$$\theta(L)\Delta y_t = \delta + \phi(L)\Delta x_t - \gamma z_{t-1} + \alpha(L)\varepsilon_t, \quad z_t = y_t - \beta x_t$$

La dimostrazione del teorema di Granger si basa sul fatto che se x_t, y_t sono $I(1)$, ma sono fra loro cointegrate deve esistere un qualche meccanismo che riporta a zero l'errore rispetto all'equilibrio.

TEOREMA DI RAPPRESENTAZIONE INVERSO $\rightarrow y_t, x_t$ sono $I(1)$ e hanno una rappresentazione a correzione dell'errore allora esse sono cointegrate.

PROCESSI VAR

Un processo stocastico multivariato è un processo stocastico i cui elementi non sono variabili casuali semplici, ma multiple. Ad esempio se pensiamo alla rilevazione giornaliera del tasso di cambio euro/dollaro come alla realizzazione di un processo univariato, possiamo pensare alla rilevazione giornaliera dei tassi di cambio euro/dollaro, euro/yen, euro/sterlina come alla realizzazione di un processo multivariato.

I processi VAR costituiscono la generalizzazione multivariata dei processi AR. Un processo VAR(p) può essere scritto:

$$\begin{cases} \text{Var}(p) \rightarrow \underline{y}_t = \underline{\delta} + \theta_1 \underline{y}_{t-1} + \dots + \theta_p \underline{y}_{t-p} + \underline{\varepsilon}_t, & \text{con } \theta_i = \text{matrice } K \times K, \quad \underline{y}_t = (y_{1t}, \dots, y_{pt})' \\ \text{Se indico con } \theta(L) = \text{polinomio matriciale} = I_k - \theta_1 L - \dots - \theta_p L^p \rightarrow \theta(L) \underline{y}_t = \underline{\delta} + \underline{\varepsilon}_t \end{cases}$$

Un VAR descrive l'evoluzione dinamica di un insieme di variabili a partire dalla loro storia passata congiunta.

Se consideriamo 2 variabili y_{1t}, y_{2t} il var consiste in due equazioni.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad \varepsilon_{1t}, \varepsilon_{2t} \text{ sono WN indipendenti da } y_1, y_2$$

Un processo VAR(p) è stazionario se le soluzioni dell'equazione $|\theta(L)| = 0$ sono tutti maggiori di 1 in modulo. Un processo VAR(1) è stazionario se gli autovalori della matrice sono tutti quanti minori di 1. Si può scrivere un processo VAR di ordine p come un processo VAR(1), cambiando la variabile in una ausiliaria, che è chiamata rappresentazione **companion form**. A questo punto si può verificare la stazionarietà o con gli autovalori o con le radici del polinomio.

I parametri di un VAR si possono stimare in maniera consistente con una serie di regressioni OLS.

Le stime di un VAR possono essere utilizzate per:

- **Previsione** – come i processi ARMA tranne che il set informativo sul quale effettuare il condizionamento comprende il passato di più serie;
- **Analisi di causalità** - In generale, le relazioni di causa-effetto sono molto complesse da stabilire in un'analisi empirica di dati economici. Se osserviamo un'alta correlazione fra due variabili X e Y, possiamo dire tutt'al più che quelle due variabili presentano una spiccata tendenza a muoversi insieme. Tuttavia esiste una definizione di causalità che offre la possibilità di determinare il senso del nesso causa-effetto su basi puramente statistiche: la causa precede sempre l'effetto (Granger Causalità). Per vedere se in un processo VAR(2) scritto come prima ci sono variabili Granger Casuali, ad esempio, verifico se nella prima delle due equazioni $y_{1t} = \delta_1 + \theta_{11} * y_{1,t-1} + \theta_{12} * y_{2,t-2} + \varepsilon_{1t}$ si può rifiutare o meno l'ipotesi nulla $\theta_{12} = 0$. In verità la nozione di Granger causalità è abusata perché il concetto logico di causa-effetto prescinde da ciò che accade nel tempo fisico. In particolare, è possibile che la causa si manifesti solo dopo l'effetto, quando questo è influenzato dalle aspettative: il fatto che gli acquisti natalizi vengano fatti prima di Natale non ci autorizza a dire che la celebrazione del Natale il 25 dicembre sia causata dall'aumento di vendite nei negozi.
- **Scrittura processo VAR in VMA** (modello a media mobile multivariato) – Se il polinomio di ritardo matriciale è invertibile si ha la stessa procedura per la scrittura dei AR in MA.

9 DATI PANEL

Osservazioni ripetute relative alle stesse unità (individui, famiglie, imprese) relative ad un certo numero di periodi (dati sia in cross-section che in serie storica). I dati Panel sono sia in grado di descrivere perché le singole unità si comportano in modo diverso, sia di rappresentare il meccanismo che induce una certa unità a comportarsi in maniera diversa da una data all'altra a causa di un passato diverso.

MODELLO GENERALE

$y_{it} = \underline{x}_{it}' \underline{\beta}_{it} + \varepsilon_{it}$ con $\underline{\beta}_{it}$ che misura l'effetto parziale di x_{it} nel periodo t per l'unità i . Questo modello risulta essere troppo generale. Ipotesizzo perciò che $\underline{\beta}_{it}$ sia costante per tutti gli i e per tutti i t con l'eccezione del termine d'intercetta.

$$y_{it} = a_i + \underline{x}_{it}' \underline{\beta} + \varepsilon_{it}, \quad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2), \text{ tutte le } x_{it} \text{ indipendenti da } \varepsilon_{it}, \underline{x}_{it} \text{ non ha costante}$$

MODELLO AD EFFETTI FISSI

Il modello ad effetti fissi è una regressione lineare in cui i termini di intercetta variano da individuo a individuo:

$$y_{it} = a_i + \underline{x}_{it}' \underline{\beta} + \varepsilon_{it}$$

I termini di intercetta catturano gli effetti delle variabili specifiche dell'individuo i -esimo e costanti nel tempo. Cioè gli a_i sono ipotizzati essere N parametri costanti ignoti. Ogni individuo ha come caratteristica individuale una media: tale media non è ad ogni misurazione un errore casuale che dipende dall'individuo che sto misurando, ma è parte delle caratteristiche dell'individuo e serve a spiegare la differenza tra un individuo e l'altro.

STIMATORE DEI MINIMI QUADRATI CON VARIABILI DUMMY- LSDV (Least Square Dummy Variables)

$$OLS \text{ normale} \rightarrow y_{it} = \sum_{j=1}^N a_j d_{ij} + \underline{x}_{it}' \underline{\beta} + \varepsilon_{it}, \quad \text{con } d_{ij} = 1 \text{ se } i = j \text{ variabile dummy per ciascuna unità } i$$

TEST PULLED VS FIXES – testo l'annullarsi contemporaneo di tutte quante le a_i .

Lo stimatore LSDV non è efficiente.

STIMATORE WITHIN

Lo stimatore within, che coincide con lo stimatore LSDV, stima "concentrandosi" tra la varianza dentro i gruppi.

$$y_{it} = a_i + \underline{x}_{it}' \underline{\beta} + \varepsilon_{it} \rightarrow \begin{cases} \bar{y}_i = a_i + \bar{\underline{x}}_i' \underline{\beta} + \bar{\varepsilon}_i, \text{ con } \bar{y}_i = T^{-1} \sum_t y_{it} \rightarrow y_{it} - \bar{y}_i = (\underline{x}_{it} - \bar{\underline{x}}_i)' \underline{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i) \\ \hat{\beta}_{FE} = \left(\sum_{i=1}^N \sum_{t=1}^T (\underline{x}_{it} - \bar{\underline{x}}_i)(\underline{x}_{it} - \bar{\underline{x}}_i)' \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T (\underline{x}_{it} - \bar{\underline{x}}_i)(y_{it} - \bar{y}_i)' \end{cases}$$

Effetti fissi si preferiscono quando il numero di unità è modesto, e se queste presentano un carattere strettamente specifico. Ovvero lo stimatore ad effetti fissi è preferibile quando è importante poter identificare le singole unità individuali. Lo stimatore ad effetti fissi elimina le variabili costanti nel tempo.

MODELLO A EFFETTI CASUALI

Gli effetti individuali sono considerati come estrazioni indipendenti da una variabile casuale con media μ e varianza σ_a^2 . Allora il termine di errore consiste in due componenti: una fissa rispetto al tempo (errore specifico per ogni individuo) e una residuale ed incorrelata dal tempo (un errore per ogni misurazione in generale)

$$y_{it} = \mu + \underline{x'_{it}} \underline{\beta} + \varepsilon_{it} + a_i, \quad \varepsilon_{it} \sim IID(0, \sigma_\varepsilon^2), a_i \sim IID(0, \sigma_a^2) \text{ tutte le } x_{it} \text{ indipendenti da } \varepsilon_{it}, a_i \text{ a loro volta indi.}$$

Dove $\varepsilon_{it} + a_i$ viene considerato un termine d'errore composto da due elementi: una componente specifica all'individuo, che non varia nel tempo, e una componente residuale, che è ipotizzata essere incorrelata nel tempo.

Si preferisce il modello ad effetti casuali quando commetto un errore diverso da individuo ad individuo, pur mantenendo tale errore costante nel tempo; ad esempio, immaginando che la variabile dipendente sia il PIL mondiale e che essa sia messa in relazione con il PIL delle società più grandi del mondo APPLE, MICROSOFT (che fanno il ruolo di individui) si ha un modello a effetti casuali quando si può ipotizzare che per misurare il PIL delle società tecnologiche produco un errore diverso da società a società, specifico della società di cui sto misurando il PIL.

STIMATORE GLS

La struttura a componenti d'errore del modello ad effetti casuali implica che il termine d'errore composito $\varepsilon_{it} + a_i$ presenta una forma di autocorrelazione particolare. Di conseguenza gli standard error dello stimatore OLS calcolato automaticamente non sono più corretti ed è possibile costruire uno stimatore più preciso GLS.

Lo stimatore ad effetti casuali (EGLS) è una media ponderata matriciale dello stimatore between e lo stimatore within, dove i pesi dipendono dalle varianze relative dei due stimatori. È efficiente rispetto sia a within che between.

$$\hat{\beta}_{RE} = \Delta \hat{\beta}_B + (I_k - \Delta) \hat{\beta}_{FE}$$

STIMATORE BETWEEN

Sfrutta la dimensione between, ovvero la differenza tra la media dell'individuo i e la media complessiva. Lo stimatore between trascura l'informazione in serie storica contenuta nel campione. Non è efficiente.

$$y_{it} = \mu + \underline{x'_{it}} \underline{\beta} + \varepsilon_{it} + a_i \rightarrow \begin{cases} \bar{y} = \mu + \bar{x}' \underline{\beta} + \bar{\varepsilon} + a_i \rightarrow (\bar{y}_i - \bar{y}) = (\bar{x}'_i - \bar{x}) \underline{\beta} + (a_i - \bar{a} + \bar{\varepsilon}_i + -\bar{\varepsilon}) \\ \hat{\beta}_B = \left[\sum_{t=1}^T (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right]^{-1} \sum_{t=1}^T (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}) \end{cases}$$

STIMATORE OLS

Ignora l'autocorrelazione del modello ad effetti casuali (c'è autocorrelazione la componente dell'errore a_i resta costante nel tempo e quindi gli errori sono correlati nel tempo) e stima in maniera consistente, ma con standard error sbagliati e senza più la proprietà dell'efficienza.

TEST DI HAUSMAN

L'ipotesi che x_{it} e a_i siano incorrelati fra loro può essere verificata con il test di Hausman. Se ci dovesse essere correlazione tra x_{it} e a_i si preferisce lo stimatore within perché in quello between si dovrebbe usare una variabile strumentale. Per verificare se c'è correlazione tra i due termini il test di Hausman confronta se lo stimatore a effetti fissi (consistente a prescindere) è significativamente diverso dallo stimatore a effetti casuali.

ADATTAMENTO AI DATI

Possono essere calcolate diversi R^2 (between, within, effetti casuali, OLS) come coefficienti di correlazione:

$$\begin{cases} R^2_{between} = Corr^2 \left\{ (\bar{x}_i - \bar{x}) \hat{\beta}_B, (\bar{y}_i - \bar{y}) \right\} \\ R^2_{within} = Corr^2 \left\{ (x_{it} - \bar{x}_i) \hat{\beta}_{FE}, (y_{it} - \bar{y}_i) \right\} \\ R^2_{overall} = Corr^2 \left\{ x_{it} \hat{\beta}_B, y_{it} \right\} \end{cases}$$

I tre R^2 possono essere calcolati per qualsiasi stimatore. Se calcoliamo i tre R^2 per lo stimatore ad effetti casuali, essi saranno maggiori rispetto agli R^2 calcolati per effetti fissi, between e OLS. Quindi, più che per confrontare modelli diversi, gli R^2 sono utili per scegliere fra specificazioni alternative del modello.

INTERPRETAZIONE STIMATORE WITHIN COME STIMATORE VARIABILI STRUMENTALI

Dati panel forniscono variabili strumentali "interne" che è possibile utilizzare nel caso di regressori endogeni soggetti a un errore di misura.

Alcune trasformazioni delle variabili originarie sono incorrelate con il termine d'errore del modello, e correlate con le variabili esplicative stesse.

Ad esempio lo stimatore ad effetti fissi può essere riscritto in un modo in cui è possibile interpretarlo come uno stimatore delle variabili strumentali di β nel modello ad effetti casuali: $y_{it} = \mu + x'_{it} \beta + \varepsilon_{it} + a_i$ dove per ciascuna variabile esplicativa si considera lo scarto del proprio valore rispetto alla media individuale.

Ovvero lo strumento per x_{it1} è $x_{it1} - \bar{x}_{it1}$.

Per costruzione $E[(x_{it1} - \bar{x}_{it1})a_i] = 0$,

$E[(x_{it1} - \bar{x}_{it1})\varepsilon_{it}] = 0$ che è implicito nell'esogeneità stretta di x_{it1}