Stima del modello di regressione lineare

- Consideremo i modelli:
  - generalizzato
  - ad errori eteroschedastici
  - classico

- Stimatori:
  - OLS (minimi quadrati ordinari)
  - MM (metodo dei momenti)
  - GLS (minimi quadrati generalizzati)
  - ML (massima verosimiglianza)

Il contesto è quello del modello lineare generalizzato:

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\underset{k\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

$$E[\underset{n\times 1}{\varepsilon}|X] = \underset{n\times 1}{0}$$

$$V[\varepsilon|X] = E[\varepsilon\varepsilon'|X] = \underset{n\times n}{\Sigma}$$ def. positiva, non stocastica

Come primo obiettivo, vogliamo stimare β (k coeff. di regressione)

Ipotesi: $\boxed{\beta \in \Theta_\beta \subseteq \mathbb{R}^K}$

$y_t = \beta x_t + \varepsilon_t$

$\beta \geq 0$

Θ_β, lo "spazio dei parametri" è contenuto in $\mathbb{R}^K$ ed è noto a priori.

Indichiamo con $\tilde{\beta}$ un generico stimatore di β.

$\tilde{\beta}$ è una funzione che "mappa" i nostri dati (y, X) nello spazio dei parametri –

$\tilde{\beta}$ è quindi una funzione di y e X che ci restituisce un punto dello spazio dei parametri Θ_β.

Stimatore di minimi quadrati ordinari (OLS)

$\tilde{\beta} \implies$ residuo: $\tilde{\varepsilon}_i = y_i - x_i' \tilde{\beta}$ $\quad i = 1, 2, \dots, n$

In notazione compatta: $\tilde{\varepsilon} := y - X\tilde{\beta}$

$\widehat{E[y|X]} = X\tilde{\beta}$

Il residuo è una funzione di $\tilde{\beta}$: $\tilde{\varepsilon} = \varepsilon(\tilde{\beta})$

Principio dei minimi quadrati: lo stimatore OLS è il punto $\hat{\beta}$ dello spazio dei parametri che minimizza la somma dei quadrati dei residui

La somma dei quadrati dei residui è

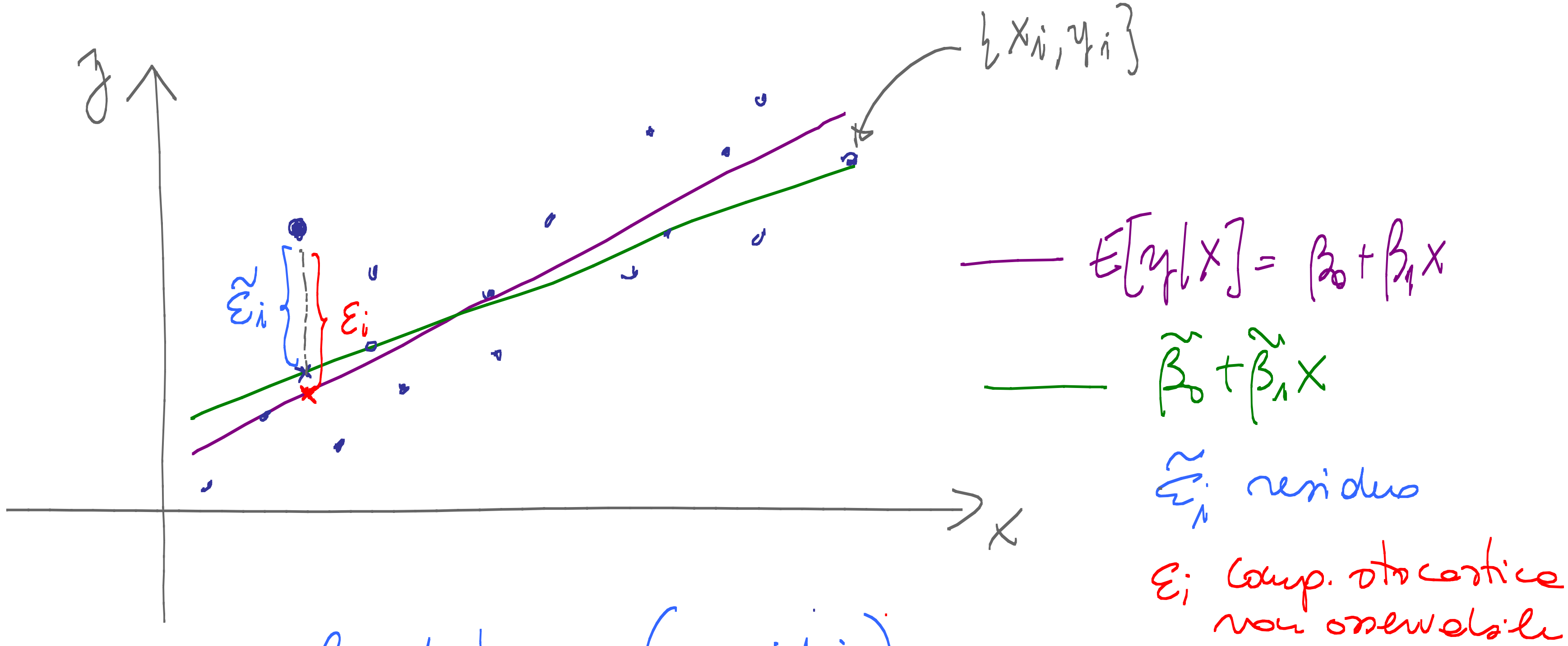
$$Q(\tilde{\beta}) := \sum_{i=1}^{n} \tilde{\varepsilon}_i^2 = \sum_{i=1}^{n} \varepsilon_i(\tilde{\beta})^2$$

che posso anche scrivere come

$$Q(\tilde{\beta}) = \tilde{\varepsilon}'\tilde{\varepsilon} = \varepsilon(\tilde{\beta})'\varepsilon(\tilde{\beta})$$

Lo stimatore OLS è il punto di minimo di $Q(\tilde{\beta})$, con $\tilde{\beta} \in \Theta_\beta$:

$$\boxed{\hat{\beta} := \arg\min_{\tilde{\beta} \in \Theta_\beta} Q(\tilde{\beta})}$$



OLS minimizza le distanze (residui) tra i punti empirici e la retta ipotizzata

La "metrica" che viene usata è la somma dei quadrati dei residui

nota: L'uso dei quadrati non è obbligatorio. Ad esempio lo stimatore LAD (least absolute deviations) minimizza

$$Q_{LAD}(\tilde{\beta}) := \sum_{i=1}^{n} |\tilde{\varepsilon}_i|$$

ed è diverso dallo stimatore OLS.

Derivazione dello stimatore OLS

$\tilde{\beta}, \quad \tilde{\varepsilon} := y - X\tilde{\beta}$

$Q(\tilde{\beta}) := \tilde{\varepsilon}'\tilde{\varepsilon} = (y - X\tilde{\beta})'(y - X\tilde{\beta})$

$(AB)' = B'A'$

Consideriamo il caso in cui $\Theta_\beta \equiv \mathbb{R}^K$

Otteniamo la condizione del primo ordine, $\frac{\partial}{\partial\tilde{\beta}} Q(\tilde{\beta}) = 0$

$$= \begin{pmatrix} \partial Q(\tilde{\beta})/\partial\beta_1 \\ \vdots \\ \partial Q(\tilde{\beta})/\partial\beta_K \end{pmatrix}$$

Poiché $Q(\tilde{\beta}) = (y' - \tilde{\beta}'X')(y - X\tilde{\beta})$

$$= \underbrace{y'y}_{1\times 1} - \underbrace{y'X\tilde{\beta}}_{\underbrace{1\times n \; n\times K \; K\times 1}_{1\times 1}} - \underbrace{\tilde{\beta}'X'y}_{1\times 1} + \underbrace{\tilde{\beta}'X'X\tilde{\beta}}_{1\times 1}$$

$(y'X\tilde{\beta})' = \tilde{\beta}'X'y$

$$= y'y - 2\tilde{\beta}'X'y + \tilde{\beta}'X'X\tilde{\beta}$$

$$\boxed{\begin{gathered} \frac{\partial}{\partial a} a'B = B \\ \frac{\partial}{\partial a} a'Ba = 2Ba \end{gathered}}$$

otteniamo:

$$\frac{\partial Q(\tilde{\beta})}{\partial\tilde{\beta}} = 0 - 2\frac{\partial}{\partial\tilde{\beta}}(\tilde{\beta}'X'y) + \frac{\partial}{\partial\tilde{\beta}} \tilde{\beta}'X'X\tilde{\beta}$$

sistema di K equazioni in K incognite

$$= -2\underbrace{X'y}_{K\times 1} + 2\underbrace{X'X\tilde{\beta}}_{K\times 1} = \underset{K\times 1}{0}$$

Se $X'X$ è invertibile (ovvero se $\underset{n\times K}{X}$ ha rango massimo $= K$) allora possiamo scrivere

$$X'y = X'X\tilde{\beta}$$

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

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

$$\frac{\partial^2 Q(\beta)}{\partial\beta\,\partial\beta'} = \begin{pmatrix} \frac{\partial^2 Q(\beta)}{(\partial\beta_1)^2} & \frac{\partial^2 Q(\beta)}{\partial\beta_1\partial\beta_2} & \cdots \\ \vdots & \vdots & \end{pmatrix}$$

Corrisponde ad un punto di minimo:

$$\frac{\partial^2 Q(\beta)}{\partial\beta\,\partial\beta'} = 2X'X$$

(fare i passaggi)

dove $X'X$ è positiva definita ($a'X'Xa > 0$, per ogni $a \neq 0$) in quanto rango$(X) = K$.

$\tilde{\beta}$ è un punto di minimo

E' inoltre unico, in quanto la funzione $Q(\tilde{\beta})$ è convessa.

Lo stimatore $\hat{\beta} = (X'X)^{-1}X'y$ è lo stimatore OLS.

<u>nota</u> ricordiamo che

$$\underset{n \times k}{X} = \begin{bmatrix} — x_1' — \\ — x_2' — \\ \vdots \\ — x_n' \end{bmatrix} \quad \text{dove} \quad x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix}$$

e, analogamente,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Quindi:

$$X'X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — x_1' — \\ — x_2' — \\ \vdots \\ — x_n' — \end{bmatrix} = x_1x_1' + x_2x_2' + \dots + x_nx_n' = \sum_{i=1}^{n} x_i x_i'$$

$$X'y = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

Possiamo quindi scrivere

$$\hat\beta = (X'X)^{-1} X'y = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} x_i y_i \right)$$

$$= \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1}}_{S_{xx}} \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} x_i y_i \right)}_{S_{xy}} = S_{xx}^{-1} S_{xy}$$

S_{xx} è la matrice dei momenti secondi di x (nel campione)
S_{xy} " " " " " " incrociati tra x e y

Se x e y fossero espressi come scarti dalle rispettive medie,
S_{xx} sarebbe la matrice di varianze e covarianze di x_i $i=1\dots n$
e S_{xy} " " " " covarianze tra x_i e y_i, $i=1\dots n$

Nel caso scalare, $\hat\beta = \dfrac{\frac{1}{n}\sum x_i y_i}{\frac{1}{n}\sum x_i^2}$

Stimatore OLS: residui, valori stimati, proiezioni

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

Il modello teorico è $y = X\beta + \varepsilon$, ovvero $E[y|X] = X\beta$.

I residui OLS sono $\hat{\varepsilon} := y - X\hat{\beta}$ (differenza tra i valori empirici y e la retta di regressione stimata)

I valori stimati sono $\hat{y} := X\hat{\beta}$. Rappresentano la nostra stima del valore atteso condizionato $E[y|X]$:

$$\hat{y} := X\hat{\beta} = \hat{E}[y|X]$$

Notiamo che

$$y = \hat{y} + \hat{\varepsilon}$$

$$y = \hat{y} + \hat{\varepsilon}$$
$$(y = X\beta + \varepsilon)$$

Consideriamo il valore atteso stimato $\hat{y}$. Abbiamo:

$$\hat{y} := X\hat{\beta} = \underset{n\times k}{X}\underbrace{(\underset{k\times n}{X'}\underset{n\times k}{X})^{-1}}_{k\times k}\underset{k\times n}{X'}y = \underset{n\times n}{P_X}\cdot \underset{n\times 1}{y}, \quad \underset{n\times n}{P_X} := \underset{n\times k}{X}(\underset{k\times k}{X'X})^{-1}\underset{k\times n}{X'}$$

La matrice P_X prende il nome di "matrice di proiezione": proietta nel sottospazio (dello spazio $\mathbb{R}^n$) generato da X.

Proprietà di P_X

(1) $\text{rango}(P_X) = K$

(2) "idempotenza": $P_X \cdot P_X = P_X$

fare dimostrazione

nota: il rango di una matrice idempotente è uguale alla sua traccia.

Consideriamo ora i residui OLS, $\hat{\varepsilon} = y - X\hat{\beta} = y - \hat{y}$

Abbiamo:

$$\hat{\varepsilon} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = y - P_X \cdot y$$

$$= (I_n - P_X)\cdot y = M_X \cdot y, \quad M_X = I_n - P_X = I_n - X(X'X)^{-1}X'$$

M_X è un'altra matrice di proiezione: proietta nel sottospazio ortogonale a quello generato da X

$\boxed{\text{Alcune proprietà:}}$ ① M_X è idempotente $(M_X \cdot M_X = M_X)$ e simmetrica $(M_X' = M_X)$
(provare per esercizio)

② M_X e P_X sono ortogonali tra loro:

$$M_X \cdot P_X = (I_n - P_X) \cdot P_X = P_X - P_X P_X = P_X - P_X = 0$$

Una conseguenza della ② è che $\hat{\varepsilon}$ e $\hat{y}$ sono tra loro ortogonali:

$$\hat{\varepsilon}' \hat{y} = (M_X \cdot y)' P_X \cdot y = y' M_X \cdot P_X \cdot y = 0 \quad \text{poiché } M_X P_X = 0$$

Abbiamo usato la simmetria di M_X

Analogamente $\hat{\varepsilon}$ è ortogonale a X:

$$\hat{\varepsilon}' X = (M_X \cdot y)' X = y' M_X \cdot X = 0 \quad \text{in quanto } M_X X = 0:$$

$$M_X \cdot X = (I_n - X(X'X)^{-1} X') X = X - X \cancel{(X'X)^{-1} X'X} = 0$$

(nota: $P_X \cdot X = X$ (provare))

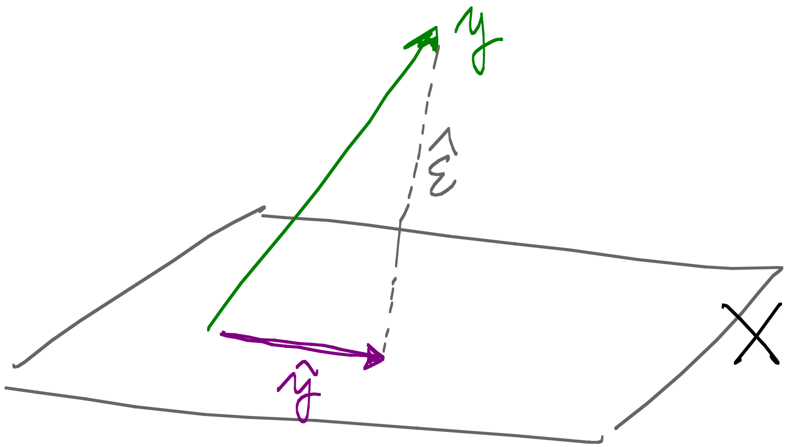
③ Se $y = X\beta + \varepsilon$, allora:

$$\hat{\varepsilon} = M_X \cdot y = M_X(X\beta + \varepsilon) = \underbrace{M_X X}_{0}\beta + M_X \cdot \varepsilon = 0 + M_X \varepsilon$$

Quindi

$$M_X \cdot y = M_X \varepsilon$$

Interpretazione grafica



$\longrightarrow$ $\hat{y} = P_X \cdot y$

------ $\hat{\varepsilon} = M_X \cdot y$

Note: scomposizione della devianza

$$\sum_{i=1}^{n} y_i^2 = y'y = y' I_n y = y'(M_X + P_X) y$$

$$= y' M_X y + y' P_X y$$

$$= \hat{\varepsilon}'\hat{\varepsilon} + y' X (X'X)^{-1} X' y$$

$$= \hat{\varepsilon}'\hat{\varepsilon} + y' X (X'X)^{-1} (X'X) (X'X)^{-1} X' y$$

$$= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\beta}' (X'X) \hat{\beta}$$

$$= \hat{\varepsilon}'\hat{\varepsilon} + (X\hat{\beta})' (X\hat{\beta})$$

$$y'y = \hat{\varepsilon}'\hat{\varepsilon} + \hat{y}'\hat{y}$$

somme dei quadrati totale — somme dei quadrati residua — somme dei quadrati stimata

$$\left(y'y - n\bar{y}^2\right) = \hat{\varepsilon}'\hat{\varepsilon} + \left(\hat{y}'\hat{y} - n\bar{y}^2\right)$$

Devianza Totale — Devianza residua — Devianza Spiegata

Proprietà dello stimatore OLS

Ricordiamo le ipotesi: $y = X\beta + \varepsilon$, $E(\varepsilon|X) = 0$, $V(\varepsilon|X) = \Sigma$, def. posit.

$\text{rango}(X) = K$

Proprietà ①: correttezza

$$E(\hat{\beta}|X) = \beta$$

dimostrazione: $E(\hat{\beta}|X) = E\left((X'X)^{-1}X'y \,|\, X\right)$

$$= (X'X)^{-1}X' \underline{E[y|X]} = \cancel{(X'X)^{-1}}\cancel{X'}X\beta = \beta$$

dimo alternativa: $E(\hat{\beta}|X) = (X'X)^{-1}X'\underbrace{E[y|X]} = (X'X)^{-1}X'X\beta = \beta$

$$E[X\beta + \varepsilon|X] = X\beta + \underbrace{E[\varepsilon|X]}_{=0}$$

② linearità

Uno stimatore si dice lineare se può essere scritto nella forma $\tilde{\beta} = A_X \cdot y$, dove A_X non dipende da y (è una combinazione lineare di y)

$$\hat{\beta} = (X'X)^{-1}X'y = A_X \cdot y, \qquad A_X = (X'X)^{-1}X'$$

$V(aX) = a^2 X$

$V(AX) = AV(X)A'$

③ Varianza dello stimatore OLS:

$$V(\hat{\beta}|X) = \underbrace{(X'X)^{-1}}\,X'\Sigma X\,\underbrace{(X'X)^{-1}}$$

dimostrazione:

$$V(\hat{\beta}|X) = V(\underbrace{(X'X)^{-1}X'}y/X)$$

$$= (X'X)^{-1}X'\,\underbrace{V(y|X)}\,X(X'X)^{-1}$$

$$= \underbrace{(X'X)^{-1}}X'\Sigma\,X\underbrace{(X'X)^{-1}}$$

note ①. La varianza di $\hat{\beta}$ dipende da Σ, che non è nota

② Supponiamo che $\Sigma := \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$.
In questo caso:

$$X'\Sigma X = [x_1, x_2, \dots, x_n] \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix} = \sum_{i=1}^{n} x_i x_i' \cdot \sigma_i^2$$

Quindi,

$$V(\hat{\beta}|X) = (X'X)^{-1} X'\Sigma X (X'X)^{-1}$$

$$= \left(\sum x_i x_i'\right)^{-1} \left(\sum x_i x_i' \sigma_i^2\right) \left(\sum x_i x_i'\right)^{-1}$$

$$= \frac{1}{n} \left(\frac{1}{n} \sum x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \sigma_i^2\right) \left(\frac{1}{n} \sum x_i x_i'\right)^{-1}$$

dipende da $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ (non note in generale)

③ Supponiamo che $\Sigma = \sigma^2 I_n$ (modello lineare classico)
Abbiamo:

$$V(\hat{\beta}|X) = (X'X)^{-1} X'\Sigma X (X'X)^{-1} = (X'X)^{-1} X'(\sigma^2 I_n) X (X'X)^{-1}$$

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

che dipende solo da σ^2.
Inoltre:

$$V(\hat{\beta}|X) = \sigma^2 (X'X)^{-1} = \sigma^2 \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} = \frac{\sigma^2}{n} \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i'\right)^{-1}$$

Riassunto su OLS:

Lo stimatore OLS risolve il problema di minimo

$$\hat{\beta} := \arg\min_{\tilde{\beta}\in\Theta_\beta} \varepsilon(\tilde{\beta})'\varepsilon(\tilde{\beta}), \quad \varepsilon(\tilde{\beta}) := y - X\tilde{\beta}$$

- La soluzione è:

$$\hat{\beta} = (X'X)^{-1}X'y = S_{xx}^{-1} S_{xy}, \quad S_{xx} = \frac{1}{n}X'X = \frac{1}{n}\sum x_i x_i'$$

$$S_{xy} = \frac{1}{n}X'y = \frac{1}{n}\sum x_i y_i$$

- Lo stimatore OLS scompone y in due parti ortogonali tra loro, $\hat{y}$ (valore atteso stimato) e $\hat{\varepsilon}$ (residuo)

- $\hat{\beta}$ è uno stimatore lineare (comb. lineare di y)

- Sotto le ipotesi del modello lineare generalizzato,

$$E(\hat{\beta}|X) = \beta \quad \text{(correttezza)}$$

$$V(\hat{\beta}|X) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

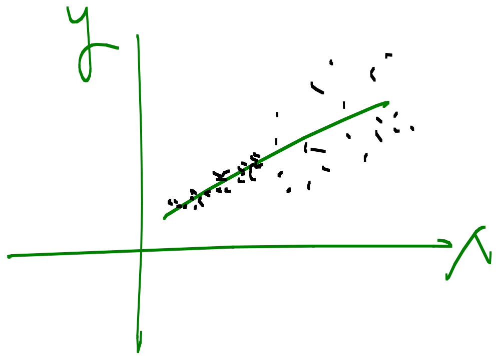
Se $\Sigma = \sigma^2 I_n$ (ML classico), allora $V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

Stimatore GLS (generalized least squares)

OLS risolve un problema di minimo dove tutte le osservazioni hanno lo stesso peso:

$$\varepsilon(\tilde{\beta})'\varepsilon(\tilde{\beta}) = \sum_{i=1}^{n} (y_i - x_i'\tilde{\beta})^2 = \min$$

GLS pesa le osservazioni in funzione di quanta informazione sulla retta di regressione incognita sono in grado di portare.



Formalmente:

$$\hat{\beta}_{GLS} := \arg\min_{\tilde{\beta}\in\Theta_\beta} \underbrace{\underbrace{\varepsilon(\tilde{\beta})'}_{1\times n}\ \underbrace{\Sigma^{-1}}_{n\times n}\ \underbrace{\varepsilon(\tilde{\beta})}_{n\times 1}}_{1\times 1}$$

Indicando con σ^{ij} il generico elemento di Σ^{-1}, lo stimatore GLS può essere anche scritto come

$$\hat{\beta}_{GLS} := \arg\min_{\tilde{\beta}\in\Theta_\beta} \sum_{i=1}^{n}\sum_{j=1}^{n} (y_i - x_i'\tilde{\beta})(y_j - x_j'\tilde{\beta})\cdot\sigma^{ij} = \arg\min_{\tilde{\beta}\in\Theta_\beta} \sum_i\sum_j \tilde{\varepsilon}_i\tilde{\varepsilon}_j\sigma^{ij}$$

nota: Se $\Sigma := \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$ (ML eteroschedastico)
allora

$$\Sigma^{-1} = \text{diag}\left\{\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \dots, \frac{1}{\sigma_n^2}\right\}.$$

Quindi,

$$\sigma^{ij} = \begin{cases} 0 & \text{se } i \neq j \\ 1/\sigma_i^2 & \text{se } i = j \end{cases}$$

e

$$\hat{\beta}_{GLS} = \underset{\tilde{\beta} \in \Theta_\beta}{\arg\min} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} (y_i - \tilde{\beta}' x_i)^2$$

↳ ogni osservazione viene pesata con l'inversa di σ_i^2
(σ_i^2 grande $\to$ poco peso)

nota: Se $\Sigma = \sigma^2 I_n$ (ML classico), allora $\Sigma^{-1} = \frac{1}{\sigma^2} I_n$ e

$$\hat{\beta}_{GLS} := \arg\min \varepsilon(\tilde{\beta})' \Sigma^{-1} \varepsilon(\tilde{\beta}) = \arg\min \frac{1}{\sigma^2} \varepsilon(\tilde{\beta})' \varepsilon(\tilde{\beta})$$

$$= \arg\min \varepsilon(\tilde{\beta})' \varepsilon(\tilde{\beta}) = \hat{\beta}_{OLS}$$

Deriviamo $\hat{\beta}_{GLS}$:

$$Q_\Sigma(\tilde{\beta}) = \underbrace{\varepsilon(\tilde{\beta})}'\,\Sigma^{-1}\,\varepsilon(\tilde{\beta})$$

Somma "generalizzata" dei quadrati dei residui

Corrisponde a:

$$Q_\Sigma(\tilde{\beta}) = \overbrace{(y - X\tilde{\beta})}'\,\Sigma^{-1}(y - X\tilde{\beta}) = (y - X\tilde{\beta})'(\Sigma^{-1}y - \Sigma^{-1}X\tilde{\beta})$$

$$= y'\Sigma^{-1}y - \tilde{\beta}'X'\Sigma^{-1}y - y'\Sigma^{-1}X\tilde{\beta} + \tilde{\beta}'X'\Sigma^{-1}X\tilde{\beta}$$

$$= y'\Sigma^{-1}y - 2\tilde{\beta}'X'\Sigma^{-1}y + \tilde{\beta}'X'\Sigma^{-1}X\tilde{\beta}$$

La derivata prima è:

$$\frac{\partial}{\partial\tilde{\beta}} Q_\Sigma(\tilde{\beta}) = 0 - 2X'\Sigma^{-1}y + 2X'\Sigma^{-1}X\tilde{\beta} = 0$$

$$\boxed{\frac{\partial}{\partial x} x'A = A}$$

$$\boxed{\frac{\partial}{\partial x} x'Ax = 2Ax}$$

$$X'\Sigma^{-1}X\tilde{\beta} = X'\Sigma^{-1}y$$

Ora se : rango(X) = max = K
rango (Σ) = max = n

allora $X'\Sigma^{-1}X$ ha rango max = K ed è quindi invertibile

Infatti: se rango Σ = max, allora rango Σ^{-1} è max.
Se rango $X = k$ = max, allora rango $X'\Sigma^{-1}X$ è massimo
Quindi $X'\Sigma^{-1}X$ è invertibile

In questo caso, premoltiplicando per $(X'\Sigma^{-1}X)^{-1}$, otteniamo

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

Lo stimatore OLS corrisponde a $\hat{\beta}_{GLS}$ per $\Sigma = I_n$

$\boxed{\text{Proprietà}}$ Se valgono le ipotesi del ML generalizzato:

1 Correttezza: $E(\hat{\beta}_{GLS}|X) = \beta$

<u>dim</u>
$$E[\hat{\beta}_{GLS}|X] = E[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y|X]$$
$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}E[y|X]$$
$$= \cancel{(X'\Sigma^{-1}X)^{-1}}\cancel{X'\Sigma^{-1}X}\beta = \beta$$

$V(Ax) = A V(x) A'$

② La varianza (condizionata) di $\hat{\beta}_{GLS}$ è

$$V[\hat{\beta}_{GLS} | X] = (X'\Sigma^{-1}X)^{-1}$$

dim:
$$V[\hat{\beta}_{GLS}|X] = V[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y|X]$$
$$= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} V(y|X)\, \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}$$
$$= \cancel{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}}\cancel{\Sigma}\,\cancel{\Sigma^{-1}}X(X'\Sigma^{-1}X)^{-1}$$
$$= (X'\Sigma^{-1}X)^{-1}$$

note: $V(\hat{\beta}_{GLS}|X) = (X'\Sigma^{-1}X)^{-1} \neq (X'X)^{-1}X'\Sigma X(X'X)^{-1} = V(\hat{\beta}_{OLS}|X)$

vedremo che $V(\hat{\beta}_{GLS}|X) < V(\hat{\beta}_{OLS}|X)$

note: se $\Sigma = \sigma^2 I_n$, $V(\hat{\beta}_{GLS}|X) = (X'(\sigma^2 I_n)^{-1}X)^{-1}$
$$= ((\sigma^2)^{-1}X'X)^{-1} = \sigma^2(X'X)^{-1} = V(\hat{\beta}_{OLS}|X)$$

③ Teorema di Aitken

Sotto le ipotesi effettuate [ovvero: $E[y|X] = X\beta$, $V[y|X] = \Sigma$ non stocastica, $\text{rango}(X) = \max = k$, $\text{rango}(\Sigma) = \max = n$] lo stimatore GLS è efficiente (= a minima varianza) nella classe degli stimatori lineari e corretti

(In breve, $\hat{\beta}_{GLS}$ è BLUE — best linear unbiased estimator)

dimostrazione.

Abbiamo già visto che $\hat{\beta}_{GLS}$ è lineare e corretto.
Consideriamo un generico stimatore lineare:

$$\underset{k\times 1}{\tilde{\beta}} = \underset{k\times n}{A}\ \underset{n\times 1}{y}$$

(ad esempio, nel caso di $\hat{\beta}_{GLS}$, $\hat{\beta}_{GLS} = A \cdot y$, per $A = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$)

Il valore atteso di $\tilde{\beta}$ è

$$E[\tilde{\beta}|X] = E[Ay|X] = A\,E[y|X] = A \cdot X\beta$$

Quindi, affinché $\tilde{\beta}$ sia corretto, A deve soddisfare $\underset{K\times n}{A}\cdot\underset{n\times K}{X} = I_K$.
Calcoliamo la varianza di $\tilde{\beta}$:

$$V[\tilde{\beta}|X] = V[A\cdot y|X] = A\,V(y|X)\,A' = A\Sigma A'$$

Vogliamo dimostrare che $V[\tilde{\beta}|X] = A\Sigma A' \geq (X'\Sigma^{-1}X)^{-1} = V[\hat{\beta}_{GLS}|X]$

Riscriviamo A nel seguente modo:

$$A = \left[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} + D\right]$$

Notiamo che la condizione di correttezza $A\cdot X = I_K$ implica che

$$A\cdot X = \left[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} + D\right]\cdot X = I_K + DX = I_K$$

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

da cui:

$$DX = 0$$

Torniamo alla varianza di $\tilde{\beta}$:

$$V[\tilde{\beta}|X] = A\Sigma A' = \left[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} + D\right]\cdot\Sigma\left[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} + D\right]'$$

$$= \left[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} + D\right]\Sigma\left[\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1} + D'\right]$$

$$= (X'\Sigma^{-1}X)^{-1}X'\cancel{\Sigma^{-1}}\cancel{\Sigma}\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}$$
$$+ (X'\Sigma^{-1}X)^{-1}X'\cancel{\Sigma^{-1}}\cancel{\Sigma}D' + D\cancel{\Sigma}\cancel{\Sigma^{-1}}X(X'\Sigma^{-1}X)^{-1}$$
$$+ D\Sigma D'$$

$$= (X'\Sigma^{-1}X)^{-1} + 0 + 0 + D\Sigma D' \qquad (\text{in quanto } DX = 0)$$

$$= V[\hat{\beta}_{GLS}|X] + D\Sigma D'.$$

Quindi:

$$V[\tilde{\beta}|X] - V[\hat{\beta}_{GLS}|X] = D\Sigma D' \geq 0 \qquad (\text{semidefinita positiva})$$

L'uguaglianza a 0 si ottiene se $D = 0$. Ma in questo caso $\tilde{\beta} = \hat{\beta}_{GLS}$. $\square$

<u>nota importante</u>: Se $\Sigma = \sigma^2 I_n$ (modello lineare classico)

il teorema di Aitken diventa il seguente:

Se $E[y|X] = X\beta$, $V[y|X] = \sigma^2 I_n$, rango$(X) = \max = K$

allora lo stimatore $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ è lo stimatore a minima varianza nella classe degli stimatori lineari e corretti (B.L.U.E)

Questo è il noto teorema Gauss-Markov.

Esercizio: dimostrare Gauss-Markov facendo gli stessi passaggi del teorema di Aitken.

Una derivazione alternativa dello stimatore GLS.

Il teorema Gauss-Markov dice che, se $E[y|X] = X\beta$ e $V[y|X] = \sigma^2 I_n$, allora OLS è lo stimatore "migliore"

Supponiamo però che $V[y|X] = \Sigma \neq \sigma^2 I_n$

Come possiamo ottenere lo stimatore BLU?

Possibile idea: trasformare la regressione $y = X\beta + \varepsilon$ in un'altra regressione che soddisfi G.M

Come si fa:

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\,\underset{k\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

Trasformiamo il modello usando una "matrice di trasformazione" T, di dimensione $n\times n$

$$\underbrace{T\cdot y}_{} = \underbrace{TX}_{}\beta + \underbrace{T\varepsilon}_{}$$

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

dove $\underset{n\times 1}{\tilde{y}} := \underset{n\times n}{T}\,\underset{n\times 1}{y}$, $\tilde{X} := TX$, $\tilde{\varepsilon} := T\varepsilon$.

Notiamo che: $E[\tilde{y}|\tilde{X}] = E[\tilde{X}\beta + \tilde{\varepsilon}|X] = \tilde{X}\beta + E[\tilde{\varepsilon}|X]$

$$= \tilde{X}\beta + E[T\cdot\varepsilon|X] =$$

$$= \tilde{X}\beta + T\cdot\underbrace{E[\varepsilon|X]}_{0} = \tilde{X}\beta$$

La "nuova" componente stocastica soddisfa

$$E[\tilde{\varepsilon}|X] = 0$$

e la sua varianza è:

$$V[\tilde{\varepsilon}|X] = V[T\cdot\varepsilon|X] = T\cdot V[\varepsilon|X]\cdot T' = T\Sigma T'$$

Vogliamo T in modo che $T\cdot\Sigma\cdot T' = I_n$.

Per trovare T, scomponiamo Σ nel seguente modo:

$$\underset{n\times n}{\Sigma} = \underset{n\times n}{Q}\cdot\underset{n\times n}{\Lambda}\,\underset{n\times n}{Q'}$$

dove Λ è la matrice (diagonale) degli autovalori e Q è la matrice (ortonormale) degli autovettori,

$$Q'Q = QQ' = I_n$$

Notiamo che $\Sigma^{-1} = Q\Lambda^{-1}Q'$

(infatti $\Sigma\Sigma^{-1} = Q\Lambda Q'Q\Lambda^{-1}Q = I_n$)

e prendiamo $T := \Lambda^{-1/2}\cdot Q'$

Abbiamo che: $V[\tilde{\varepsilon}|X] = T\Sigma T' = \Lambda^{-1/2}\cancel{Q'}\,\cancel{Q}\Lambda\cancel{Q'}\,Q\Lambda^{-1/2}$

$$= \Lambda^{-1/2} \cdot \Lambda \Lambda^{-1/2} = I_n$$

Il modello trasformato usando $T = \Lambda^{-1/2} Q'$ soddisfa G-M
Lo stimatore BLU è quindi

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

$$= (X'T'TX)^{-1} X'T'Ty$$

$$= (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y$$

$\tilde{X} := TX$
$\tilde{y} := Ty$

in quanto: $T'T = Q\Lambda^{-1/2} \cdot \Lambda^{-1/2} Q' = Q\Lambda^{-1}Q' = \Sigma^{-1}$

Notiamo infine che

$$\hat{\beta} = \arg\min_{\tilde{\beta}} (\tilde{y} - \tilde{X}\tilde{\beta})'(\tilde{y} - \tilde{X}\tilde{\beta})$$

$$= \arg\min (Ty - TX\tilde{\beta})'(Ty - TX\tilde{\beta})$$

$$= \arg\min (y'T' - \tilde{\beta}'X'T')(Ty - TX\tilde{\beta})$$

$$= \arg\min \left(y' - \tilde{\beta}' X' \right) T' T \left(y - X\tilde{\beta} \right)$$

$$= \arg\min \left(y - X\tilde{\beta} \right)' \Sigma^{-1} \left(y - X\tilde{\beta} \right) =: \hat{\beta}_{GLS}$$

Ricapitolando

modello lineare generalizzato $\quad y = X\beta + \varepsilon, \; E[\varepsilon / X] = 0, \; V[\varepsilon / X] = \Sigma$ d.p.

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

- corretto lineare
- non efficiente

$$V[\hat{\beta}_{OLS} | X] = (X'X)^{-1} X' \Sigma X (X'X)^{-1}$$ non è "minimo"

stimatore GLS $\quad \hat{\beta}_{GLS} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$

- corretto lineare
- efficiente (B.L.U.E.)

$$V[\hat{\beta}_{GLS} | X] = (X' \Sigma^{-1} X)^{-1}$$ è "minimo"

Purtroppo, lo stimatore GLS non è sempre calcolabile: dipende infatti da Σ.

Si parla di stimatore "virtuale". In inglese: "INFEASIBLE"

Esistono alcuni casi in cui $\hat{\beta}_{GLS}$ può essere calcolato

① $\Sigma = \sigma^2 I_n$ (modello lineare classico).

In questo caso, $\hat{\beta}_{GLS} = \hat{\beta}_{OLS} = (X'X)^{-1}X'y$

② modello lineare con eteroschedasticità (esempio)

$$y_i = x_i'\beta + \varepsilon_i \qquad E[\varepsilon_i|X] = 0$$

$$\sigma_i^2 = V[\varepsilon_i|X] = h \cdot z_i^2 \qquad (h > 0)$$

$$Cov[\varepsilon_i, \varepsilon_j|X] = 0 \text{ se } i \neq j$$

(ipotesi che la varianza di ε_i dipende da una variabile z_i)

In forma compatta: $y = X\beta + \varepsilon$, $E[\varepsilon|X] = 0$

$$V[\varepsilon|X] = V\left[\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{pmatrix}\Big| X\right] = \begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{bmatrix} = \begin{bmatrix}h z_1^2 & & \\ & h z_2^2 & \\ & & \ddots \\ & & & h z_n^2\end{bmatrix} =$$

$$= h\begin{bmatrix}z_1^2 & & \\ & z_2^2 & \\ & & \ddots \\ & & & z_n^2\end{bmatrix} = \Sigma = h\cdot \mathrm{diag}\{z_1^2, \dots, z_n^2\}$$

Calcoliamo lo stimatore GLS:

$$X = \begin{bmatrix}- x_1' -\\ \vdots \\ - x_n' -\end{bmatrix} \quad n\times k \qquad X' = [x_1 \cdots x_n] \quad k\times n \qquad X'X$$

$$\hat\beta_{GLS} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} y$$

$$= \left[\cancel{h}^{-1}(x_1\ x_2 \dots x_n)\begin{pmatrix}z_1^2 & & \\ & z_2^2 & \\ & & \ddots \\ & & & z_n^2\end{pmatrix}^{-1}\begin{pmatrix}x_1'\\ x_2'\\ \vdots\\ x_n'\end{pmatrix}\right]^{-1}\left[\cancel{h}^{-1}(x_1\ x_2 \dots x_n)\begin{pmatrix}z_1^2 & & \\ & z_2^2 & \\ & & \ddots \\ & & & z_n^2\end{pmatrix}^{-1}\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}\right]$$

$$= \left(\sum_{i=1}^n \frac{1}{z_i^2} x_i x_i'\right)^{-1}\left(\sum_{i=1}^n \frac{1}{z_i^2} x_i y_i\right) \qquad \left(\sum \frac{x_i x_i'}{z_i^2}\right)^{-1}\left(\sum \frac{x_i y_i}{z_i^2}\right)$$

In questo caso, $\hat\beta_{GLS}$ è calcolabile (FEASIBLE) e corrisponde allo stimatore dei minimi quadrati ponderati.

Il peso di ogni unità statistica è $\frac{1}{z_i^2}$ (inversa della var. di ε_i)

<u>Nota/esercizio</u>: in questo caso, la matrice di trasformazione T è

$$T := \begin{bmatrix} 1/z_1 & & & \\ & 1/z_2 & & \\ & & \ddots & \\ & & & 1/z_n \end{bmatrix}$$

e il modello trasformato è $(Ty) = (TX)\beta + (T\varepsilon)$
dove

$$T \cdot y = \begin{bmatrix} y_1/z_1 \\ y_2/z_2 \\ \vdots \\ y_n/z_n \end{bmatrix} \qquad TX = \begin{bmatrix} x_1'/z_1 \\ x_2'/z_2 \\ \vdots \\ x_n'/z_n \end{bmatrix}$$

$$y_i = \beta' x_i + \underbrace{\varepsilon_i}_{(0,\, h \cdot z_i^2)}$$

$$\frac{y_i}{z_i} = \beta' \frac{x_i}{z_i} + \underbrace{\frac{\varepsilon_i}{z_i}}_{(0,\, h)}$$

che corrisponde alla regressione

$$\left(\frac{y_i}{z_i}\right) = \left(\frac{x_i}{z_i}\right)' \beta + \tilde{\varepsilon}_i \quad , \quad i = 1, 2, \dots, n$$

Infatti:

$$\hat{\beta} = \left[\sum_{i=1}^{n} \left(\frac{x_i}{z_i} \frac{x_i'}{z_i}\right)\right]^{-1} \sum_{i=1}^{n} \left(\frac{x_i}{z_i}\right)\left(\frac{y_i}{z_i}\right) = \left(\sum_1^n \frac{x_i x_i'}{z_i^2}\right)^{-1} \sum_1^n \left(\frac{x_i y_i}{z_i^2}\right)$$

③ Dati aggregati

x_{ij} = reddito del j-esimo individuo della regione i $j = 1, 2, \dots, n_i$ = numero di individui nella regione i-esima

$i = 1, \dots, N$ = num. regioni

y_{ij} = consumo j-esimo individuo della regione i

Supponiamo che valga la relazione a livello individuale:

$$y_{ij} = \beta x_{ij} + \varepsilon_{ij} \quad , \quad E[\varepsilon_{ij} | X] = 0$$
$$E[\varepsilon_{ij}^2 | X] = \sigma^2$$
$$E[\varepsilon_{ij} \varepsilon_{hk} | X] = 0 \, , \quad ij \neq hk$$

Il campione osservato è invece costituito da dati aggregati (medie)

$$y_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \qquad \text{(consumo medio regione } i\text{)}$$

$$x_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \qquad \text{(reddito medio regione } i\text{)}$$

e costruiamo il modello

$$y_i = \beta x_i + \varepsilon_i \quad , \quad i = 1, 2, \dots, N$$

Quali sono le proprietà di ε_i ?

Ovviamente $\varepsilon_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij}$, quindi

$$E[\varepsilon_i | X] = E\left[\frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij} \, \Big| X \right] = \frac{1}{n_i} \sum_{j=1}^{n_i} E[\varepsilon_{ij} | X] = 0$$

$$V[\varepsilon_i | X] = V\left[\frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij} \, \Big| X \right] = \frac{1}{n_i^2} \sum_{j=1}^{n_i} \underbrace{V[\varepsilon_{ij} | X]}_{\sigma^2} = \frac{\sigma^2}{n_i}$$

Quindi, siamo nell'ambito di un modello lineare con errori eteroschedastici

$$E[y_i | X] = \beta x_i$$

$$V[y_i | X] = \sigma_i^2 = \frac{\sigma^2}{n_i}$$

Se conosco le numerosità m_i, $i=1,\dots,N$
posso usare GLS:

$$y_i = \beta x_i + \varepsilon_i \qquad \text{(modello di partenza)}$$

$$E(\varepsilon_i|X)=0, \quad V(\varepsilon_i|X) = \sigma_i^2 = \frac{\sigma^2}{m_i}$$

Passo al modello trasformato dividendo per $\sigma_i = \frac{\sigma}{\sqrt{m_i}}$

$$\frac{y_i}{\sigma_i} = \beta \frac{x_i}{\sigma_i} + \varepsilon_i/\sigma_i$$

$$\frac{y_i}{\sqrt{\cancel{\sigma^2}/m_i}} = \beta \frac{x_i}{\sqrt{\cancel{\sigma^2}/m_i}} + \tilde{\varepsilon}_i \qquad \text{(modello trasformato)}$$

$$\tilde{y}_i = \beta \tilde{x}_i + \tilde{\varepsilon}_i$$

dove $\tilde{y}_i = y_i/\sqrt{\sigma^2/m_i}$ e $\tilde{x}_i = x_i/\sqrt{\sigma^2/m_i}$, $i=1,2\dots,N$

Inoltre, $\tilde{\varepsilon}_i$ soddisfa le condizioni del Modello lin. ordinario.

Lo stimatore GLS corrisponde allo stimatore OLS sul modello trasformato:

$$\hat{\beta}_{GLS} = \left(\sum_1^N \tilde{x}_i \tilde{x}_i'\right)^{-1} \sum_1^N \tilde{x}_i \tilde{y}_i$$

$$= \left(\sum_1^N x_i^2 \Big/ \left(\frac{\sigma^2}{n_i}\right)\right)^{-1} \left(\sum_1^N \frac{x_i y_i}{\left(\frac{\sigma}{\sqrt{n_i}}\right)\left(\frac{\sigma}{\sqrt{n_i}}\right)}\right) = \left(\sum_1^N n_i x_i^2\right)^{-1} \sum_1^N n_i x_i y_i$$

che è diverso da $\hat{\beta}_{OLS} = \left(\sum_1^N x_i^2\right)^{-1} \left(\sum_1^N x_i y_i\right)$.

(4) Break strutturale nella varianza

La varianza di ε_i cambia in corrispondenza di una particolare osservazione:

$$E[\varepsilon_i^2 | X] = \begin{cases} \sigma_A^2 & \text{se } i = 1, 2, \dots, n^* \\ \sigma_B^2 & \text{se } i = n^*+1, \dots, n \end{cases}$$

Esempio: great moderation

Se gli errori sono "incorrelati" $\left(E[\varepsilon_i \varepsilon_j | X] = 0, \ i \neq j \right)$ allora:

$$V[\varepsilon|X] = E[\varepsilon\varepsilon'|X] = \begin{bmatrix} \sigma_A^2 & & & & \\ & \ddots & & & \\ & & \sigma_A^2 & & \\ & & & \sigma_B^2 & \\ & & & & \ddots \\ & & & & & \sigma_B^2 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n^* \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n-n^* \end{array}$$

Lo stimatore GLS è:

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} y$$

$$= \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i y_i \right), \qquad \sigma_i^2 = \begin{cases} \sigma_A^2 & i = 1 \dots n^* \\ \sigma_B^2 & i = n^*+1, \dots n \end{cases}$$

$$= \left(\sum_{i=1}^{n^*} \frac{1}{\sigma_A^2} x_i x_i' + \sum_{i=n^*+1}^{n} \frac{1}{\sigma_B^2} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n^*} \frac{1}{\sigma_A^2} x_i y_i + \sum_{i=n^*+1}^{n} \frac{1}{\sigma_B^2} x_i y_i \right)$$

$$= \left(\frac{1}{\sigma_A^2} \sum_{i=1}^{n^*} x_i x_i' + \frac{1}{\sigma_B^2} \sum_{i=n^*+1}^{n} x_i x_i' \right)^{-1} \left(\frac{1}{\sigma_A^2} \sum_{i=1}^{n^*} x_i y_i + \frac{1}{\sigma_B^2} \sum_{i=n^*+1}^{n} x_i y_i \right)$$

Lo stimatore non è (ancora) calcolabile: dipende da σ_A^2 e σ_B^2.

Possiamo stimare σ_A^2 e σ_B^2 usando i residui OLS:

$$\hat{\varepsilon}_i = y_i - x_i' \hat{\beta}_{OLS} \quad , \quad i = 1, 2, \ldots, n$$

come

$$\hat{\sigma}_A^2 = \frac{1}{n^*} \sum_{i=1}^{n^*} \hat{\varepsilon}_i^2 \qquad \text{oppure} \qquad \frac{1}{n^* - k} \sum_{i=1}^{n^*} \hat{\varepsilon}_i^2$$

$$\hat{\sigma}_B^2 = \frac{1}{n - n^*} \sum_{i=n^*+1}^{n} \hat{\varepsilon}_i^2 \qquad \qquad \frac{1}{n - n^* - k} \sum_{i=1}^{n^*} \hat{\varepsilon}_i^2$$

Lo stimatore GLS "feasible" (FGLS) diventa:

$$\hat{\beta}_{FGLS} = \left(\frac{1}{\hat{\sigma}_A^2} \sum x_i x_i' + \frac{1}{\hat{\sigma}_B^2} \sum x_i x_i' \right)^{-1} \left(\frac{1}{\hat{\sigma}_A^2} \sum x_i y_i + \frac{1}{\hat{\sigma}_B^2} \sum x_i y_i \right)$$