PROBLEMI DI PREVISIONE

Si vuole prevedere il valore di Y_{n+1} per un insieme di valori X osservati.

Supponiamo però per X i valori

$$C = \{1 \quad X_{2,n+1} \quad X_{3,n+1} \; \ldots \; X_{k,n+1}\}$$

E' possibile fare una previsione puntuale o stimare un intervallo di previsioni.

$$Y_{n+1} = \beta_1 + \beta_2 X_{2,n+1} + \ldots + \beta_k X_{k,n+1} + u_{n+1}$$

$$= \underset{1\times k}{C'} \underset{k\times 1}{\beta} + \underset{1\times 1}{u_{n+1}}$$

$$E[Y_{n+1}] = C'\beta$$

Utilizzando le proprietà BLUE di $\hat{\beta}$ avremo il

PREVISORE PUNTUALE

$$\hat{Y}_{n+1} = C'\hat{\beta}$$ sarà BLUFF

Best Linear Unbiased Forecasting Function

APPLICAZIONE

$$Y = \beta_1 + \beta_2 X + u$$

Voglio prevedere Y da X_0. Per calcolare l'intervallo devo determinare

$$(X'X) = \begin{bmatrix} n & \sum X \\ \sum X & \sum X^2 \end{bmatrix}$$

$$C = \{1 \quad X_0\}$$

$$C'(X'X)^{-1}C = \frac{\sum X^2 - 2X_0 \sum X + u X_0^2}{u\sum X^2 - \left(\sum X\right)^2}$$

Infatti

$$\{1 \quad X_0\} \frac{1}{n\sum X^2 - \left(\sum X\right)^2} \begin{bmatrix} \sum X^2 & -\sum X \\ -\sum X & n \end{bmatrix} \begin{pmatrix} 1 \\ X_0 \end{pmatrix} =$$

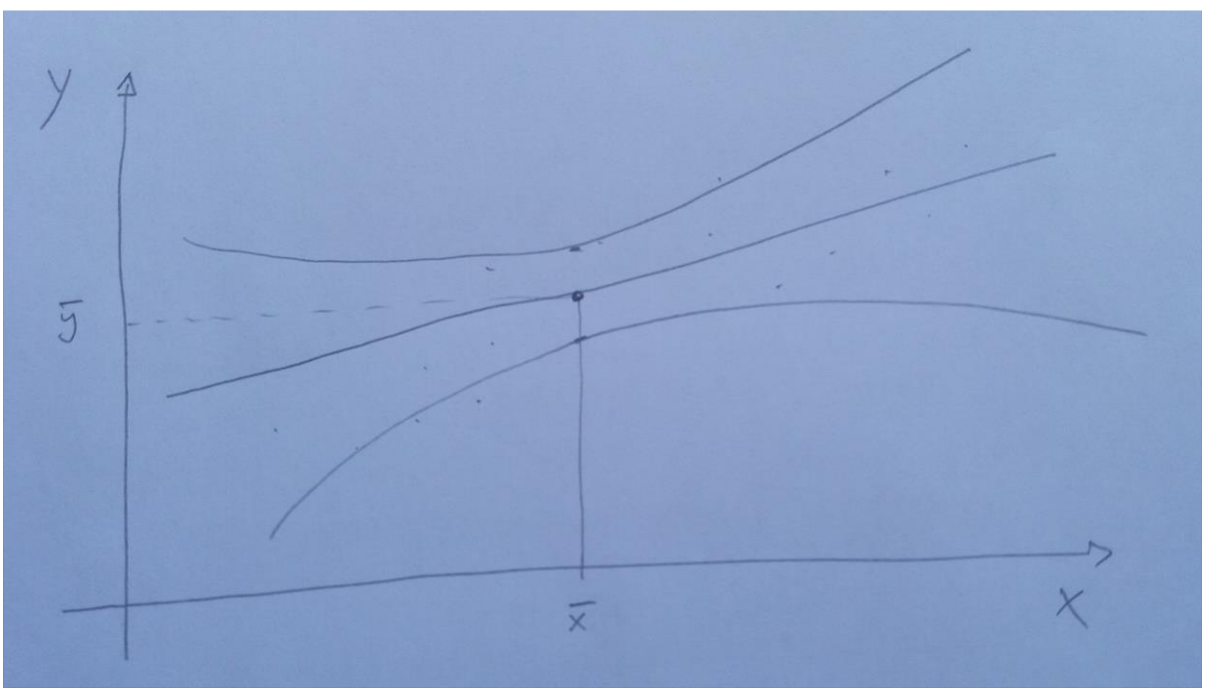
$$= \frac{1}{n\sum X^2 - \left(\sum X\right)^2} \left[\sum X^2 - X_0 \sum X \, , X_0 n - \sum X\right] \begin{pmatrix} 1 \\ X_0 \end{pmatrix} =$$

$$= \frac{\left(\sum X^2 - X_0 \sum X\right) + \left(X_0^2 n - X_0 \sum X\right)}{n\sum X^2 - \left(\sum X\right)^2} =$$

$$= \frac{\sum X^2 - 2X_0 \sum X + nX_0^2}{n\sum X^2 - \left(\sum X\right)^2} = \frac{1}{n} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum X^2}$$

Lo standard error di *Y0* sarà

$$s.e.(\hat{Y}_0) = s\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$



A parità di dati osservati l'intervallo sarà tanto più largo quanto più X_0 è distante da $\bar{X}$

CENNI SULLE VARIABILI DUMMY
(Variabili di comodo)

Fino ad ora abbiamo assunto che nella equazione generale $Y = X\beta + u$

Le variabili X siano variabili cardinali date dalla teoria economica.

E' possibile introdurre variabili cosiddette "di comodo" che riescano a rappresentare diversi fattori :

- EFFETTI TEMPORALI
- EFFETTI SPAZIALI
- VARIABILI QUALITATIVE

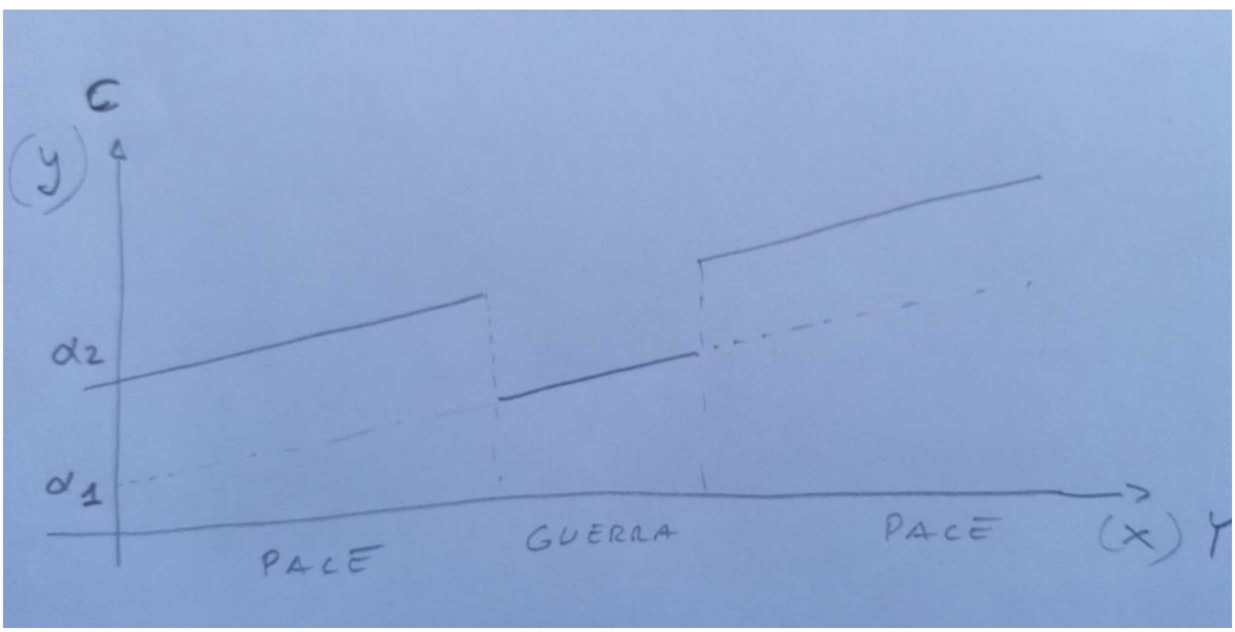
È possibile che un modello economico possa subire mutamenti strutturali :

FUNZIONE DI CONSUMO

$$\boxed{\begin{array}{l} C = \alpha_1 + \beta Y + u \\ C = \alpha_2 + \beta Y + u \end{array}}$$

Tempo di guerra

Tempo di pace



Si ipotizza comunque che la propensione marginale al consumo $\dfrac{\partial C}{\partial Y} = \beta$ rimanga invariata in entrambi i periodi

Invece di considerare i due modelli separatamente (stime meno precise) vengono uniti in una sola relazione

$$\boxed{C = \alpha_1 X_1 + \alpha_2 X_2 + \beta Y + u}$$

Dove X_1 e X_2 sono variabili dummy :

$$X_1 = \begin{cases} 1 & \textit{anni di guerra} \\ 0 & \textit{anni di pace} \end{cases}$$

$$X_2 = \begin{cases} 0 & \textit{anni di guerra} \\ 1 & \textit{anni di pace} \end{cases}$$

La matrice β dei coefficienti sarà $\beta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix}$

e la matrice dei dati

$$X = \begin{pmatrix} X_1 & X_2 & Y \end{pmatrix} = \begin{bmatrix} 0 & 1 & Y_1 \\ 0 & 1 & Y_2 \\ . & . & Y_3 \\ 0 & 1 & . \\ 1 & 0 & . \\ 1 & 0 & . \\ . & . & . \\ 1 & 0 & . \\ 0 & 1 & . \\ . & . & . \\ 0 & 1 & Y_n \end{bmatrix}$$

La trappola delle variabili di comodo

Quando utilizziamo le variabili dummy è necessariob fare attenzione a come viene costruito il modello, per non rendere la matrice (X'X) singolare .

Infatti se nel modello precedente lasciavamo una intercetta : $C=\alpha_0+\alpha_1 X_1+\alpha_2 X_2+\beta Y+u$

$$X=\begin{bmatrix} 1 & 0 & 1 & Y_1 \\ 1 & 0 & 1 & Y_2 \\ . & . & . & . \\ 1 & 0 & 1 & . \\ 1 & 1 & 0 & . \\ 1 & 1 & 0 & . \\ . & . & . & . \\ 1 & 1 & 0 & . \\ 1 & 0 & 1 & . \\ . & . & . & . \\ 1 & 0 & 1 & Y_n \end{bmatrix}$$

$$\boxed{1\times X_0-1\times X_1-1\times X_2+0\times Y=0}$$

Abbiamo che le 4 colonne di X sono linearmente dipendenti $rank(X)=rank(X'X)=3\neq k$

(X'X) non è invertibile

Volendo utilizzare una regressione con intercetta si utilizzerà così solo una dummy :

$$\boxed{C = \gamma_1 + \gamma_2 X_2 + \beta Y + u}$$

$$X_2 = \begin{cases} 0 & anni\ di\ guerra \\ 1 & anni\ di\ pace \end{cases}$$

β = PMC in entrambi i periodi

$\alpha_1 = \gamma_1$ = intercetta anni di guerra

$\alpha_2 = \gamma_1 + \gamma_2$ = intercetta anni di pace

$\alpha_1 - \alpha_2 = \gamma_2$ = differenza tra l'intercetta del periodo guerra e pace

- Cambiamento di coefficiente angolare

$$\boxed{C = \alpha + \beta_1 Y + (\beta_2 - \beta_1) X_2 Y + u}$$

$$X_2 = \begin{cases} 0 & anni\ di\ guerra & C = \alpha + \beta_1 Y + u \\ 1 & anni\ di\ pace & C = \alpha + \beta_2 Y + u \end{cases}$$

$\beta_2 - \beta_1$ = differenza propensione marginale al consumo nei due periodi

APPLICAZIONE (p.255 Maddala)

$$Y = \beta_1 + \beta_2 \text{ SVA} + u$$

Y = km / litro

SVA = Stima Vita Auto in anni

$$\underset{(1.753)}{\hat{Y} = 7.952} + \underset{(0.061)}{0.693}\, SVA$$

$$\bar{R}^2 = 0.74$$

$$Y = \beta_1 + \beta_2 W + \beta_3 \, {}^{S}\!/_{A} + \beta_4 \, {}^{G}\!/_{D} + \beta_5 SVA + u$$

W = peso in Kg

$$ {}^{S}\!/_{A} = \begin{cases} 0 & cambio\ standard \\ 1 & cambio\ automatico \end{cases}$$

$$ {}^{G}\!/_{D} = \begin{cases} 0 & gas \\ 1 & diesel \end{cases}$$

$$\hat{Y} = \underset{(5.349)}{22.008} - \underset{(0.001)}{0.002} W - \underset{(0.708)}{2.760} \, {}^{S}\!/_{A} + \underset{(1.413)}{3.28} \, {}^{G}\!/_{D} + \underset{(0.097)}{0.415} SVA$$

$$\bar{R}^2 = 0.82$$

MULTICOLLINEARITA'

Quando tra due o più variabili esplicative vi è perfetta collinearità o multicollinearità, la matrice ($X'X$) non è più a rango pieno e le stime OLS non possono essere calcolate.

Si può però facilmente fare una sostituzione di variabile

Es :

$$\boxed{\begin{aligned} Y &= \alpha + \beta_1 X_1 + \beta_2 X_2 + u \\ X_2 &= \gamma X_1 + \theta \\ Y &= \alpha + \beta_1 X_1 + \beta_2 \theta + \beta_2 \gamma X_1 + u \end{aligned}}$$

$$= \lambda_1 + \lambda_2 X_1 + u$$

$$\lambda_1 = \alpha + \beta_2 \theta$$

$$\lambda_2 = \beta_1 + \gamma \beta_2$$

Il problema della multicollinearità esiste quindi quando due o più regressori sono quasi-collineari ovvero quando il coefficiente di correlazione tra i regressori è alto .

- MODELLO A 3 VARIABILI

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$Y = \beta_2 X_2 + \beta_3 X_3 + u - \bar{u}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \left(X'X\right)^{-1} X' y$$

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

$$= \frac{\sigma^2}{\sum X_2^2 \sum X_3^2 - \left(\sum X_2 X_3\right)^2} \begin{pmatrix} \sum X_3^2 & -\sum X_2 X_3 \\ -\sum X_2 X_3 & \sum X_2^2 \end{pmatrix}$$

$$V\left(\hat{\beta}_2\right)=\frac{\sigma^2\sum X_3^2}{\sum X_2^2\sum X_3^2-\left(\sum X_2X_3\right)^2}=$$

$$=\frac{\sigma^2\sum X_3^2}{\sum X_2^2\sum X_3^2\,\frac{\left(\sum X_2^2\sum X_3^2-\left(\sum X_2X_3\right)^2\right)}{\sum X_2^2\sum X_3^2}}$$

$$=\frac{\sigma^2}{\sum X_2^2\left(1-r_{23}^2\right)}$$

$$V\left(\hat{\beta}_3\right)=\frac{\sigma^2}{\sum X_3^2\left(1-r_{23}^2\right)}$$

È facile vedere che valori molto alti di r_{23}^2 rendono le stime OLS molto imprecise.

Inoltre piccole variazioni nella matrice dei dati provocano o possono provocare grandi variazioni nella stima dei parametri.

ESEMPIO-APPLICAZIONE: instabilità delle stime

$$Y = \beta_2 X_2 + \beta_3 X_3 + u - \bar{u}$$

Dati : $\sum X_{2i}^2 = 200$ $\quad \sum X_{2i} X_{3i} = 150$

$\sum X_{3i}^2 = 113$ $\quad \sum X_{2i} Y_i = 350$

$\sum X_{3i} Y_i = 263$

$$\hat{\beta}_2 = \frac{\sum X_3^2 \sum X_2 Y - \sum X_2 X_3 \sum X_3 Y}{\sum X_2^2 \sum X_3^2 - \left(\sum X_2 X_3\right)^2} =$$

$$\frac{113 \times 350 - 150 \times 263}{200 \times 113 - 150^2} = \frac{39550 - 39450}{22600 - 22500} = \frac{100}{100} = 1$$

$$\hat{\beta}_3 = \frac{52600 - 52500}{22600 - 22500} = \frac{100}{100} = 1$$

$$\boxed{r_{X_2 X_3}^2 = \frac{\left(\sum X_2 X_3\right)^2}{\sum X_2^2 \sum X_3^2} = \frac{150^2}{200 \times 113} = 0.995}$$

Togliendo solo una osservazione:

$$\sum X_2^2 = 199 \qquad \sum X_2 X_3 = 149$$

$$\sum X_3^2 = 112 \qquad \sum X_2 Y = 327.5$$

$$\sum X_3 Y = 261.5$$

$$\hat{\beta}_2 = \frac{112 \times 347.5 - 149 \times 261.5}{199 \times 112 - 149^2} = \frac{-43.5}{87} = -\frac{1}{2}$$

$$\hat{\beta}_3 = \frac{199 \times 261.5 - 149 \times 347.5}{199 \times 112 - 149^2} = \frac{261}{87} = 3$$

Si modificano molto le stime

ETEROSCHEDASTICITA'

Avevamo ipotizzato che

$$E[uu'] = \sigma^2 I$$

tale assunzione è in molte situazioni non valida

dobbiamo quindi riformulare il problema nella forma

$$E[u] = 0$$

$$E[uu'] = \sigma^2 \Omega$$

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

$$y = X\beta + u$$

$$E\left[\hat{\beta}\right] = \beta + (X'X)^{-1} X' E(u) = \beta$$

Sono ancora corretti ma non efficienti (ovvero non sono necessariamente a varianza minima)

$$V\left(\hat{\beta}\right) = (X'X)^{-1} X' E(uu') X (X'X)^{-1}$$
$$= (X'X)^{-1} X' \Omega X \; (X'X)^{-1} \sigma^2$$

GOLDFELD – QUANDT TEST

- Si ordinano le osservazioni secondo la variabile X_j che si ipotizza sia la causa dell'eteroschedasticità
- Si divide il campione in tre parti di numerosità n_1 n_2 n_3 .
- Dopo la stima OLS nei tre sottocampioni si calcola $e_1' e_1$ e $e_3' e_3$

$$F = \frac{e_1' e_1}{e_3' e_3} \qquad F_{n_1-k,\, n_3-k}$$

Sotto H_0 : omoschedasticità : (il valore di F è piccolo)

$$F_{empirico} > F_{teorico} \Rightarrow Rifiuto \quad H_0$$

RIMEDI

1. σ_i $i = 1 , \dots , n$ siano valori noti.

si applicano i MINIMI QUADRATI PESATI (WLS)

ovvero si applica OLS al modello trasformato

$$y_i^* = \frac{y_i}{\sigma_i} \quad ; \quad x_{ij}^* = \frac{x_{ij}}{\sigma_i} \quad ; \quad \varepsilon_i^* = \frac{\varepsilon_i}{\sigma_i}$$

Ovvero $$y_i^* = \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \dots + \beta_k x_{ik}^* + \varepsilon_i^*$$

Dove $$Var(\varepsilon_i^*) = Var\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2} Var(\varepsilon_i) = \frac{\sigma_i^2}{\sigma_i^2} = 1$$

2. $\exists$ relazione tra la componente stocastica e uno dei regressori

$$y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

Es. $$Var\ \varepsilon_i = C\, x_{i2}^2$$

Trasformiamo il modello

$$y_i^* = \frac{y_i}{x_{i2}} \quad ; \quad x_{ij}^* = \frac{x_{ij}}{x_{i2}} \quad ; \quad \varepsilon_i^* = \frac{\varepsilon_i}{x_{i2}}$$

$$\Rightarrow \frac{y_i}{x_{i2}} = \beta_1 \frac{1}{x_{i2}} + \beta_2 + \ldots + \beta_k \frac{x_{ik}}{x_{i2}} + \frac{\varepsilon_i}{x_{i2}}$$

Dove $\quad Var(\varepsilon_i^*) = Var\left(\frac{\varepsilon_i}{x_{i2}}\right) = \frac{1}{x_{i2}^2} Var(\varepsilon_i) = C$

Applico OLS

ESERCIZIO

La stima di un modello lineare sulla base dei valori del Reddito e del Consumo di 30 famiglie americane fornisce i seguenti valori :

$$\hat{C} = \underset{(3.29)}{1480} + \underset{(29.37)}{0.788}\,y \qquad R^2 = 0.97$$

La stima dello stesso modello sulle prime 12 e sulle ultime 12 osservazioni fornisce i seguenti valori:

$$\hat{C} = \underset{(0.74)}{846.7} + \underset{(9.91)}{0.837}\,y \qquad R^2 = 0.91$$

$$RSS = 1069000$$

$$\hat{C} = \underset{(0.79)}{2306.7} + \underset{(5.00)}{0.747}\,y \qquad R^2 = 0.71$$

$$RSS = 3344000$$

Verificare l’ipotesi di presenza di eteroschedasticità ed in caso affermativo indicare la procedura di correzione.

$$F = \frac{3344000}{1069000} = 3.12 \qquad F_{10,10} = 1.83$$

C’è presenza di eteroschedasticità