Università di Pavia

Econometria

2008-2009

Esercizi 3

Maggio, 2009

1. Sia $\widehat{\boldsymbol{\beta}}$ il vettore delle stime OLS di una regressione di $\mathbf{y}$ su $\mathbf{X}$ e $\mathbf{c}$ un qualunque vettore $(K \times 1)$ diverso da $\widehat{\boldsymbol{\beta}}$. Provate che la differenza nella somma dei residui al quadrato sia
$$(\mathbf{y}-\mathbf{Xc})'(\mathbf{y}-\mathbf{Xc}) - (\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) = (\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}})$$
e che $(\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}}) > 0$.

   Soluzione
   Scrivete
$$\mathbf{c} = \widehat{\boldsymbol{\beta}} + (\mathbf{c}-\widehat{\boldsymbol{\beta}})$$
   da cui
$$(\mathbf{y}-\mathbf{Xc}) = \mathbf{y} - \mathbf{X}\left[\widehat{\boldsymbol{\beta}} + (\mathbf{c}-\widehat{\boldsymbol{\beta}})\right] = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}})$$
$$(\mathbf{y}-\mathbf{Xc})'(\mathbf{y}-\mathbf{Xc}) = (\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) + (\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}}) + 2(\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})$$
   ma
$$\mathbf{X}'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{0}$$
   così il terzo termine è nullo. Possiamo scrivere
$$(\mathbf{y}-\mathbf{Xc})'(\mathbf{y}-\mathbf{Xc}) = (\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) + (\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} + (\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}})$$
$$(\mathbf{y}-\mathbf{Xc})'(\mathbf{y}-\mathbf{Xc}) - \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = (\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}}).$$
   Infine, $(\mathbf{c}-\widehat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\mathbf{c}-\widehat{\boldsymbol{\beta}}) = \mathbf{a}'\mathbf{a} > 0$, data $\mathbf{X}$ di rango K per $\mathbf{a} \neq 0$.

2. Considerate la regressione OLS di $\mathbf{y}$ su K variabili, $\mathbf{X}$ (inclusa la costante). Considerate un insieme alternativo di regressori, $\mathbf{Z} = \mathbf{XP}$, dove $\mathbf{P}$ è una matrice non singolare. Così ogni colonna di $\mathbf{Z}$ è una combinazione delle colonne di $\mathbf{X}$.

   (a) Provate che i vettori dei residui nella regressione di $\mathbf{y}$ su $\mathbf{X}$ e di $\mathbf{y}$ su $\mathbf{Z}$ sono identici.

   (b) Come possiamo rappresentare il cambio di unità di misura delle variabili indipendenti?

   (c) Cambia il fit della regressione?

   Soluzione

(a) Consideriamo la regressione di $\mathbf{y}$ su $\mathbf{X}$, i residui sono:

$$\mathbf{M}_X\mathbf{y} = [\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}$$

la regressione alternativa di $\mathbf{y}$ su $\mathbf{Z} = \mathbf{XP}$ produce i seguenti residui:

$$\begin{aligned}\mathbf{M}_Z\mathbf{y} &= \left[\mathbf{I}_N - \mathbf{XP}(\mathbf{P}'\mathbf{X}'\mathbf{XP})^{-1}\mathbf{P}'\mathbf{X}'\right]\mathbf{y} \\ &= \left[\mathbf{I}_N - \mathbf{XPP}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{P}')^{-1}\mathbf{P}'\mathbf{X}'\right]\mathbf{y} \\ &= \mathbf{M}_X\mathbf{y}\end{aligned}$$

con residui uguali dobbiamo avere la stessa regressione, cioè stime identiche $\boldsymbol{\beta}$.

(b) Semplicemente con $\mathbf{XD}$ dove $\mathbf{D}$ è una matrice nonsingolare diagonale, l'elemento k-esimo sulla diagonale principale rappresenta il fattore di scala applicato alla colonna k-esima di $\mathbf{X}$.

(c) Dal risultato precedente risulta che il fit non può essere diverso.

3. Supponete che il modello che abbia generato i dati sia:

$$y_t = \beta x_t + \epsilon_t \quad t = 1, \ldots, N$$

con $\epsilon_t \sim WN(0, \sigma^2)$ ed x_t non stocastico.

Il modello non ha costante. Confrontate la varianza dello stimatore OLS di β calcolato senza il termine costante nella regressione con quella che si avrebbe se il modello includesse la costante.

Soluzione

Lo stimatore OLS senza costante è:

$$\widehat{\beta} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}$$

Lo stimatore OLS del coefficiente di x_t del modello con costante è

$$\widehat{\beta_2} = \frac{\sum_{t=1}^{N}(x_t - \overline{x})(y_t - \overline{y})}{(x_t - \overline{x})^2}$$

La varianza di questo stimatore è

$$Var[\widehat{\beta_2}] = \sigma^2 / \left(\sum_{t=1}^{N}(x_t - \overline{x})^2\right)$$

la varianza di $\widehat{\beta}$ è

$$Var[\widehat{\beta}] = \sigma^2/(\mathbf{x}'\mathbf{x})$$

Inoltre,

$$\sum_{t=1}^{N}(x_t - \overline{x})^2 = \mathbf{x}'\mathbf{x} - N\overline{x}^2$$

Per comparare le due varianze costruiamo il rapporto

$$\begin{aligned}\frac{Var[\widehat{\beta}]}{Var[\widehat{\beta_2}]} &= [\mathbf{x}'\mathbf{x} - N\overline{x}^2]/\mathbf{x}'\mathbf{x} \\ &= 1 - N\overline{x}^2/\mathbf{x}'\mathbf{x} \\ &= 1 - N\overline{x}^2/[N\overline{x}^2 + \sum_t (x_t - \overline{x})^2] \\ &\leq 1\end{aligned}$$

Conclusione: introdurre il termine costante quando non è necessario aumenta la varianza dello stimatore OLS di β se la media del regressore è non nulla.

4. Considera il modello

$$y_t = \beta_1 + \beta_2 x_{t2} + \epsilon_t \quad t = 1, ..., N \tag{1}$$

dove $\mathbf{X}$ e ϵ sono indipendenti. Scrivi un codice in GRETL che permetta di valutare attraverso un esercizio Monte Carlo, con S simulazioni, la correttezza degli stimatori OLS di β_1 e β_2.
Considera i seguenti scenari alternativi per definire il setup del Monte Carlo:

- $N = 50, 100, 500$;
- $S = 100, 1000, 5000$;
- $\epsilon_t \sim iidN(0, 1)$ e $x_{t2} \sim iidN(0, 2)$. Utilizza un *seed* progressivo che parta da 500 per generare gli ϵ_t, mentre considera sempre la stessa $\mathbf{X}$ in ciascuna replica del Monte Carlo;
- ϵ_t distribuito come t-Student con media zero ed 1 grado di libertà e $x_{t2} \sim N(0, 2)$. Anche in questo caso, utilizza un *seed* progressivo che parta da 500 per generare ϵ, mentre considera la stessa $\mathbf{X}$ in ciascuna replica del Monte Carlo;
- $\beta_1 = 0$ e $\beta_2 = 2$;
- $\beta_1 = 2$ e $\beta_2 = 0.5$.

Quali conclusioni si possono trarre sulla consistenza degli OLS? Come cambiano i risultati asintotici rispetto ai diversi scenari presi in considerazione?