Università di Pavia

Econometria

2008-2009

Esercizi 4
Soluzioni

Maggio, 2009

Istruzioni: I commenti devono essere concisi!

1. Dato il modello di regressione lineare, con K regressori

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

con $E(\boldsymbol{\epsilon}) = \mathbf{0}$ e $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\mathbf{I}_N$

(a) Mostrate che l'R_c^2 si può scrivere come:

$$R_c^2 = \frac{\|\mathbf{M}_\iota \mathbf{X}\widehat{\boldsymbol{\beta}}\|^2}{\|\mathbf{M}_\iota \mathbf{y}\|^2} = \frac{\|\mathbf{P}_X \mathbf{M}_\iota \mathbf{y}\|^2}{\|\mathbf{M}_\iota \mathbf{y}\|^2}$$

(b) Mostrate che

$$Var(\widehat{\beta}_i|\mathbf{X}) = \frac{\sigma^2}{\sum_{t=1}^N (x_{ti} - \overline{x}_i)^2 \left(1 - \frac{\sum_{t=1}^N (\widehat{x}_{ti} - \overline{x}_i)^2}{\sum_{t=1}^N (x_{ti} - \overline{x}_i)^2}\right)}$$

dove $\frac{\sum_{t=1}^N (\widehat{x}_{ti} - \overline{x}_i)^2}{\sum_{t=1}^N (x_{ti} - \overline{x}_i)^2}$ è l'R_c^2 di una regressione del generico regressore $\mathbf{x}_i$ sugli altri $K-1$ inclusa la costante. (Suggerimento: Per ottenere $Var(\widehat{\beta}_i|\mathbf{X})$ partite dall'espressione dello stimatore OLS di β_i ottenuto applicando il teorema di FWL).

Soluzione

(a) Dato che

$$\begin{aligned} \mathbf{M}_\iota \mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{M}_\iota \mathbf{P}_{\mathbf{X}}\mathbf{y} \\ &= \mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{P}_\iota \mathbf{P}_{\mathbf{X}}\mathbf{y} \end{aligned}$$

ma $\mathbf{P}_\iota \mathbf{P}_{\mathbf{X}} = \mathbf{P}_\iota = \mathbf{P}_{\mathbf{X}}\mathbf{P}_\iota$, perchè

$$\boldsymbol{\iota} = \mathbf{X}\mathbf{e}_1$$

dove $\mathbf{e}_1 = (1, 0, \ldots, 0)'$.

$$\begin{aligned}\mathbf{P}_\iota &= \mathbf{X}\mathbf{e}_1(\mathbf{e}_1'\mathbf{X}'\mathbf{X}\mathbf{e}_1)^{-1}\mathbf{e}_1'\mathbf{X}' \\ \mathbf{P}_\iota\mathbf{P}_\mathbf{X} &= \mathbf{X}\mathbf{e}_1(\mathbf{e}_1'\mathbf{X}'\mathbf{X}\mathbf{e}_1)^{-1}\mathbf{e}_1'\mathbf{X}'\mathbf{P}_\mathbf{X} = \mathbf{P}_\iota \\ \mathbf{P}_\mathbf{X}\mathbf{P}_\iota &= \mathbf{P}_\mathbf{X}\mathbf{X}\mathbf{e}_1(\mathbf{e}_1'\mathbf{X}'\mathbf{X}\mathbf{e}_1)^{-1}\mathbf{e}_1'\mathbf{X}' = \mathbf{P}_\iota\end{aligned}$$

quindi

$$\mathbf{M}_\iota\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{P}_\mathbf{X}\mathbf{M}_\iota\mathbf{y}$$

(b) Lo stimatore OLS dell'i-esimo β:

$$\widehat{\beta}_i = (\mathbf{x}_i'\mathbf{M}_{K-i}\mathbf{x}_i)^{-1}\mathbf{x}_i'\mathbf{M}_{K-i}\mathbf{y}$$

dove $K - i$ rappresenta tutte le colonne di $\mathbf{X}$ ad eccezione dell'i-esima. La varianza di β_i è:

$$Var(\widehat{\beta}_i|\mathbf{X}) = \sigma^2(\mathbf{x}_i'\mathbf{M}_{K-i}\mathbf{x}_i)^{-1}$$

$$\begin{aligned}\mathbf{x}_i'\mathbf{M}_{K-i}\mathbf{x}_i &= \mathbf{x}_i'\mathbf{x}_i - \mathbf{x}_i'\mathbf{P}_{K-i}\mathbf{x}_i + N\overline{x}_i^2 - N\overline{x}_i^2 \\ &= \left(\sum_{t=1}^{N} x_{ti}^2 - N\overline{x}_i^2\right) - \mathbf{x}_i'\mathbf{P}_{K-i}\mathbf{x}_i + N\overline{x}_i^2 \\ &= \sum_{t=1}^{N}(x_{ti} - \overline{x}_i)^2 - (\widehat{\mathbf{x}}_i'\widehat{\mathbf{x}}_i - N\overline{x}_i^2)\end{aligned}$$

$\widehat{\mathbf{x}}_i = \mathbf{P}_{K-i}\mathbf{x}_i$ si ottiene quindi

$$\mathbf{x}_i'\mathbf{M}_{K-i}\mathbf{x}_i = \sum_{t=1}^{N}(x_{ti} - \overline{x}_i)^2\left[1 - \frac{\sum_{t=1}^{N}(\widehat{x}_{ti} - \overline{x}_i)^2}{\sum_{t=1}^{N}(x_{ti} - \overline{x}_i)^2}\right]$$

2. Considerate un modello di regressione con solo due variabili esplicative $\mathbf{x}_1$ and $\mathbf{x}_2$ entrambe centrate:

$$\mathbf{y} = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \boldsymbol{\epsilon}.$$

dove $\mathbf{x}_1$ and $\mathbf{x}_2$ sono non stocastici e $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N)$. Con $\widehat{\rho}$ indichiamo il coefficiente di correlazione campionario tra $\mathbf{x}_1$ e $\mathbf{x}_2$. Poichè entrambi sono centrati, la correlazione campionaria è

$$\widehat{\rho} \equiv \frac{\sum_{t=1}^{N} x_{1t}x_{2t}}{\left((\sum_{t=1}^{T} x_{1t}^2)(\sum_{t=1}^{T} x_{2t}^2)\right)^{1/2}}$$

(a) Mostrate che la correlazione tra le stime OLS $\widehat{\beta}_1$ e $\widehat{\beta}_2$ è uguale a $-\widehat{\rho}$.

(b) Quali, se ve ne sono, delle assunzioni di questo modello possono essere abbandonate senza cambiare il risultato?

Soluzione

(a) La correlazione campionaria può essere scritta come:

$$\widehat{\rho} = \frac{\mathbf{x}_1'\mathbf{x}_2}{(\mathbf{x}_1'\mathbf{x}_1)^{1/2}(\mathbf{x}_2'\mathbf{x}_2)^{1/2}}$$

La correlazione fra $\widehat{\beta}_1$ e $\widehat{\beta}_2$ è

$$\frac{Cov(\widehat{\beta}_1, \widehat{\beta}_2)}{(Var(\widehat{\beta}_1)^{1/2})(Var(\widehat{\beta}_2)^{1/2})}$$

$$Var(\widehat{\beta}_1) = \sigma^2(\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{x}_1)^{-1}$$
$$Var(\widehat{\beta}_2) = \sigma^2(\mathbf{x}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2)^{-1}$$

$$\begin{aligned}
\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{x}_1 &= \mathbf{x}_1'(\mathbf{I}_N - \mathbf{x}_2(\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2')\mathbf{x}_1 \\
&= \mathbf{x}_1'\mathbf{x}_1 - \mathbf{x}_1'\mathbf{x}_2(\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{x}_1 \\
&= \left(1 - \frac{(\mathbf{x}_1'\mathbf{x}_2)^2}{(\mathbf{x}_1'\mathbf{x}_1)(\mathbf{x}_2'\mathbf{x}_2)}\right)\mathbf{x}_1'\mathbf{x}_1 \\
&= (1 - \widehat{\rho}^2)\mathbf{x}_1'\mathbf{x}_1
\end{aligned}$$

Per lo stesso argomento con i pedici invertiti troviamo che

$$\mathbf{x}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2 = (1 - \widehat{\rho}^2)\mathbf{x}_2'\mathbf{x}_2$$

Possiamo concludere che

$$Var(\widehat{\beta}_1) = \sigma^2(1 - \widehat{\rho}^2)\mathbf{x}_1'\mathbf{x}_1 \qquad Var(\widehat{\beta}_2) = \sigma^2(1 - \widehat{\rho}^2)\mathbf{x}_2'\mathbf{x}_2$$

Ora la covarianza covarianza di $\widehat{\beta}_1$ con $\widehat{\beta}_2$. Sappiamo che

$$\widehat{\beta}_1 - \beta_1 = (\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\boldsymbol{\epsilon}$$
$$\widehat{\beta}_2 - \beta_2 = (\mathbf{x}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{M}_{\mathbf{X}_1}\boldsymbol{\epsilon}$$

$$\begin{aligned}
Cov(\widehat{\beta}_1, \widehat{\beta}_2) &= E\left[(\widehat{\beta}_1 - \beta_1)(\widehat{\beta}_2 - \beta_2)\right] \\
&= E\left[(\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2(\mathbf{x}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2)^{-1}\right] \\
&= \sigma^2(\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2(\mathbf{x}_2'\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2)^{-1}
\end{aligned}$$

ma

$$\begin{aligned}
\mathbf{x}_1'\mathbf{M}_{\mathbf{X}_2}\mathbf{M}_{\mathbf{X}_1}\mathbf{x}_2 &= \mathbf{x}_1'(\mathbf{I}_N - \mathbf{x}_2(\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2')(\mathbf{I}_N - \mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1')\mathbf{x}_2 \\
&= \mathbf{x}_1'\mathbf{x}_2(\mathbf{x}_2'\mathbf{x}_2)^{-1}\mathbf{x}_2'\mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{x}_2 - \mathbf{x}_1'\mathbf{x}_2 \\
&= \left(\frac{(\mathbf{x}_1'\mathbf{x}_2)^2}{\mathbf{x}_1'\mathbf{x}_1\mathbf{x}_2'\mathbf{x}_2}\right)\mathbf{x}_1'\mathbf{x}_2 - \mathbf{x}_1'\mathbf{x}_2 \\
&= \widehat{\rho}^2\mathbf{x}_1'\mathbf{x}_2 - \mathbf{x}_1'\mathbf{x}_2 = (\widehat{\rho}^2 - 1)\mathbf{x}_1'\mathbf{x}_2
\end{aligned}$$

$$Cov(\widehat{\beta}_1, \widehat{\beta}_2) = \frac{\sigma^2(\widehat{\rho}^2 - 1)\mathbf{x}_1'\mathbf{x}_2}{\sigma^4\mathbf{x}_1'\mathbf{x}_1\mathbf{x}_2'\mathbf{x}_2} = \frac{(\widehat{\rho}^2 - 1)\mathbf{x}_1'\mathbf{x}_2}{\sigma^2\mathbf{x}_1'\mathbf{x}_1\mathbf{x}_2'\mathbf{x}_2}$$

da $Cov(\widehat{\beta}_1, \widehat{\beta}_2)$, $Var(\widehat{\beta}_1)$ e $Var(\widehat{\beta}_2)$ otteniamo il coefficiente di correlazione $\widehat{\rho}$:

$$\frac{(\widehat{\rho}^2 - 1)\mathbf{x}_1'\mathbf{x}_2}{(1 - \widehat{\rho}^2)(\mathbf{x}_1'\mathbf{x}_1)^{1/2}(\mathbf{x}_2'\mathbf{x}_2)^{1/2}} = -\widehat{\rho}.$$

(b) L'assunzione che i termini di errore siano normalmente distribuiti può essere abbandonata senza cambiare il risultato al punto precedente poichè non usiamo mai questa ipotesi. Tuttavia l'assunzione $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2\mathbf{I}_N$ è essenziale per ottenere $Cov(\widehat{\beta}_1, \widehat{\beta}_2)$, $Var(\widehat{\beta}_1)$ e $Var(\widehat{\beta}_2)$.

3. Nel file *advertising.csv* trovate tre variabili S, le vendite di un certo prodotto, P, il prezzo e A le spese in pubblicità, per ciascuna avete un totale di $N = 75$ osservazioni. Stimate il seguente modello:

$$S_t = \beta_1 + \beta_2 P_t + \beta_3 A_t + \epsilon_t \quad t = 1, \ldots, N$$

assumendo che ϵ_t sia $i.i.d.N(0, \sigma^2)$.

Soluzione

(a) La regressione è significativa (significa testare per $\beta_2 = \beta_3 = 0$)?
Statistica $F(2, 72) = 29,2479$ (p-value $< 0,00001$).

(b) Testate l'ipotesi nulla che il prezzo non abbia alcun effetto sulle vendite. Commentate il risultato del test.
$H_0 : \beta_2 = 0, H_1 \beta_2 \neq 0$, $t - ratio = -7,215$ p-value $= 4,42E - 010$, rifiutiamo H_0.

(c) Includete tra i regressori A_t^2 che permette di investigare l'esistenza di rendimenti decrescenti nelle spese pubblicitarie. Stimate il nuovo modello e commentate i risultati.

Il nuovo modello è:

$$S_t = \beta_1 + \beta_2 P_t + \beta_3 A_t + \beta_4 A_t^2 + \epsilon_t \quad t = 1, \ldots, N$$

	Coefficiente	Errore Std.	rapporto t	p-value
const	$109,719$	6,799	$16,137$	0,000
P	$-7,640$	1,046	$-7,304$	0,000
A	$12,151$	3,556	$3,417$	0,001
A^2	$-2,768$	0,941	$-2,943$	0,004

Media della variabile dipendente	77,3747
D.S. della variabile dipendente	6,48854
Somma dei quadrati dei residui	1532,08
Errore standard della regressione ($\hat{\sigma}$)	4,64528
R^2	0,508235
$\bar{R}^2$ corretto	0,487456
$F(3, 71)$	24,4593

Adesso la variazione delle vendite a seguito di un incremento delle spese pubblictarie dipende dal livello di queste. Questo è dovuto alla presenza di A_t^2,l'effetto di quest'ultimo è negativo (rendimenti decrescenti).

(d) Come misurate l'effetto marginale della pubblicità sulle vendite attese?

$$\frac{\partial E[S_t|P_t, A_t]}{\partial A_t} = \beta_3 + \beta_4 A_t$$

la stima di $\frac{\partial E[S_t|P_t, A_t]}{\partial A_t}$ è $12,151 - 2,768 A_t$.

(e) Testate l'ipotesi nulla che la pubblicità non abbia effetti sulle vendite.
$H_0 = \beta_3 = \beta_4 = 0$. Statistica test: $F(2, 71) = 8,441$ con p-value $= 0,0005$.

(f) Trovate il livello ottimo di pubblicità tale che l'ultimo euro speso in pubblicità risulti in un euro di vendite addizionali, sotto l'ipotesi che il costo marginale di produrre e vendere un'altra unità di prodotto sia zero.
Il livello ottimo di pubblicità è la soluzione del seguente problema:

$$\beta_3 + 2\beta_4 A_t = 1$$

da cui si ottiene

$$A_t = \frac{1 - \beta_3}{2\beta_4}$$

(g) Testate l'ipotesi che il livello ottimo sia 1900 euro (A è misurato in migliaia di euro).
Per testare l'ipotesi che il livello ottimo odipubblictà sia 1900 definiamo l'ipotesi nulla e alternativa:

$$\begin{aligned} H_0 &: \beta_3 + 2\beta_4 1,9 = 1 \\ H_1 &: \beta_3 + 2\beta_4 1,9 \neq 1 \end{aligned}$$

La statistica test: $F(1, 71) = 0,936$, con p-value $= 0,337$.

4. Scrivi una funzione in GRETL che permetta di determinare le osservazioni influenti in un modello di regressione lineare classico. In particolare, tale funzione dovrà restituire i seguenti output:

- I residui, $\hat{\epsilon}_i$, per ciascuna osservazione;
- Gli elementi sulla diagonale principale della matrice di proiezione $\mathbf{P}_X$;
- Il vettore delle differenze tra le stime OLS di $\boldsymbol{\beta}$ ottenute utlizzando l'intero campione e quelle ottenute omettendo, di volta in volta, ciascuna osservazione;

Utilizzando i dati in ESERCIZIO4.gdt e la funzione appena creata, individua l'osservazione più influente nella seguente relazione lineare

$$producgr_t = \beta_1 + \beta_2 invgdp_t + \epsilon_t$$

dove $producgr$ è la crescita del PIL per lavoratore, mentre $invgdp$ è il rapporto tra il livello degli investimenti in capitale fisso e il PIL.