Econometria II
Part 1: Repàs del Model Lineal General

Riste Gjorgjiev
pareto.uab.es/~rgjorgjiev/econ_cat

Lorenzo Burlon
idea.uab.es/lburlon/teaching_econometria2.html

Universitat Autònoma de Barcelona

L'objectiu de l'econometria és investigar la relació lineal* entre

- una variable econòmica Y_i (variable dependent)
- i k variables explicatives $X_1, \ldots, X_k$

Per tant tenim una mostra de grandor T

A més, en el model hi ha un element aleatori u perquè l'equació lineal és una aproximació de la veritable per les raons:

- pot ser que haguem omès variables rellevants
- pot ser que hi hagi errors de mesura
- ...

Notació

Es poden utilitzar uns tipus de notació

- $Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$
- $Y_i = X_i' \beta + u_i$ o
- $Y = X\beta + u$, on

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_T \end{pmatrix}_{T\times 1} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_T \end{pmatrix}_{k\times 1} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_T \end{pmatrix}_{T\times 1}$$

$$X = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ 1 & X_{22} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & X_{21} & \ldots & X_{k1} \end{pmatrix}_{T\times k}$$

Supòsits del Model Lineal General

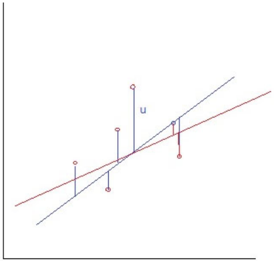
1. Les variables independents són fixes, E(X)=X
2. Les variables independents no estan correlacionades, $(X'X)^{-1}$ existeix
3. El valor mitjà de les pertorbacions és zero, $E(u_i) = 0$, $\forall i$
4. La variància de les pertorbacions és constant, $var(u_i) = \sigma^2$
5. No hi ha correlació entre les pertorbacions, $cov(u_i, u_s) = 0$
6. Les pertorbacions segueixen la distribució normal, $u_i \sim N(0, \sigma^2)$

Els incompliments dels supòsits 2, 4 i 5 es diuen: *Col·linealitat, homoscedasticitat i correlació*

Minims Quadrats Ordinaris (MQO)

L'estimador MQO dels parametres β és el que minimitza la suma de les distàncies u^2

En el cas $k = 2$



Després un exemple, veurem unes propietats del MQO

MQO amb Gretl

El fitxer MI.xls conté les dades sobre la mortalitat infantil (MI), PNB per càpita (PNBC) i la taxa d'analfabetisme de les dones (TAD).
Volem investigar l'influència de PNBC i TAD a la MI, utilitzant l'equació

$$MI = \beta_1 + \beta_2 PNBC + \beta_3 TAD + u_i$$

Aleshores, fem els següents passos:

- obrir Gretl
- introduir l'arxiu (archivo$-$ $>$abrir datos$-$ $>$importar$-$ $>$excel)
- escollir fer un MQO (Modelo$-$ $>$MQO)

Els resultats són

$$\widehat{\mathrm{MI}} = \underset{(11.593)}{263.642} - \underset{(0.0020033)}{0.00564659}\,\mathrm{PNBC} - \underset{(0.20995)}{2.23159}\,\mathrm{TAD}$$
$$T = 64 \quad \bar{R}^2 = 0.6981 \quad F(2,61) = 73.833 \quad \hat{\sigma} = 41.748 \tag{1}$$

Modelo 1: MCO, usando las observaciones 1–64
Variable dependiente: MI

	Coeficiente	Desv. Tpica	Estadstico t	Valor p
const	263.642	11.5932	22.7411	0.0000
PNBC	−0.00564659	0.00200326	−2.8187	0.0065
TAD	−2.23159	0.209947	−10.6293	0.0000

Media de la vble. dep.	141.5000	D.T. de la vble. dep.	75.97807
Suma de cuad. residuos	106315.6	D.T. de la regresin	41.74780
R^2	0.707665	R^2 corregido	0.698081
$F(2, 61)$	73.83254	Valor p (de F)	5.12e–17
Log-verosimilitud	−328.1012	Criterio de Akaike	662.2023
Criterio de Schwarz	668.6790	Hannan–Quinn	664.7538

Uns resultats de la estimació per MQO

- $\hat{\beta} = (X'X)^{-1}X'Y$
- $E(\hat{\beta}) = \beta$ - sense biaix (quins supòsits necessitem)
- residus - $\hat{u}_i = Y_i - \hat{Y}_i$
- $var(\hat{\beta}) = \sigma^2(X'X)^{-1}$ però $v\hat{a}r(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$
- $\hat{\beta} \sim N(\beta, var(\hat{\beta}))$
- $\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{T-k}$
- Error estàndard $ee(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2(X'X)^{-1}_{jj}}$
- coeficient de determinació $R^2 = 1 - \frac{SQR}{SQT} = \frac{SQE}{SQT}$
- prediccions - $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki}$
- interpretació dels coeficients estimats

Contrast d'Hipòtesis

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

- contrast d'una restricció lineal de dues cues

$$H_0 : \beta_i = \beta_{i0}$$
$$H_1 : \beta_i \neq \beta_{i0}$$

Per fer aquest contrast fem servir l'estadístic

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{\hat{\sigma}_{\hat{\beta}_i}} \sim t_{T-k}$$

Prenem la decisió de rebutjar o no a nivell α utilitzant les valors critiques de la distribució $t_{T-k,\alpha/2}$

Si $|t| > t_{T-k,\alpha/2}$ - rebutjem

Si no tenim el supòsit de perturbacions normals? $t \approx N(0,1)$

Contrast d'Hipòtesis

- contrast d'una restricció lineal d'una cua

$$
\begin{array}{lcl}
H_0 : \beta_i = \beta_{i0} & \qquad & H_0 : \beta_i = \beta_{i0} \\
H_1 : \beta_i > \beta_{i0} & \qquad & H_1 : \beta_i < \beta_{i0}
\end{array}
$$

- Utilitzem el mateix estadístic t. Quina és la diferencia?

En el primer cas rebutjem si $t > t_{T-k,\alpha}$ i en el segon si $t < -t_{T-k,\alpha}$
També en els casos de restricció lineal d'una cua, si no sabem la distribució de u, aproximem la t amb la normal
En Gretl, on veiem els resultats de l'estadístic t?

Tornem a l'exemple d'abans. Calculem l'estadistic t per la variable β_2. Hi ha dues maneres

- mirar-la directe del cuadre amb tots els resultats, $t = -2.8187$
- calcular-la utilitzant la primera equació $t = \frac{\hat{\beta}_2 - 0}{ee(\hat{\beta}_2)} = \frac{-0.00564659}{0.0020033} = -2.818644$

El valor crític de la distribució t, amb 5% de significació és

- 2.0 per contrast de dues cues
- 1.671 per contrast d'una cua

Llavors, rebutjem l'hipotesi

$$H_0 : \beta_2 = 0, \text{ v.s. } H_1 : \beta_2 \neq 0$$

Contrast de significació global del model

Havent fet la estimació del model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

és possible fer el següent contrast

$$\beta_1 = \beta_k = \ldots = \beta_k = 0$$

Què vol dir aquest contrast? Quin estadístic utilitzem?

$$F = \frac{SQE/k-1}{SQR/T-k} = \frac{R^2/(k-1)}{(1-R^2)/(T-k)}$$

La distribució de $F \sim F_\alpha(k-1, T-k)$

En el mateix exemple d'abans,

$$F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} = \frac{0.707665/(3-1)}{(1-0.707665)/(64-3)} = 73.8325$$

Però $F_{5\%}(2, 60) = 3.15$

- Conclusió: rebutjem l'hipotesi H_0

Inferència en el model restringit

Utilitzant el model amb k regressors,

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

també podem fer inferència sobre més d'una restricció,

$$R\beta = r$$

Quines són les dimensions de R i r?
L'estadístic què utilitzem per fer aquest d'inferència amb k restriccions és

$$F = \frac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/q}{SQR/(T-k)} \sim F_{q,T-k}$$

Tornem una altra vegada al mateix exemple per fer el següent contrast:

$$H_0 : \beta_3 - \beta_2 = -2$$

- Què hem de fer per utilitzar la formula d'abans?
- I amb Gretl?

L'estadístic F $= 1.15214$ i el valor crític $F_{5\%}(1, 61) = 252$

La conclusió sobre l'H_0?