Econometría I

Grado en Economía UAM

Apuntes

1. NATURALEZA DE LA ECONOMETRÍA Y DE LOS DATOS ECONÓMICOS

RECORDATORIO DEL MODELO DE REGRESIÓN

¿QUÉ ES LA ECONOMETRÍA?

La **Econometría** es la rama de la Economía que se ocupa del análisis cuantitativo de los fenómenos económicos reales, siguiendo un enfoque probabilístico. Trata de encontrar, cuantificar y contrastar relaciones económicas entre variables utilizando la modelización matemática y las técnicas basadas en los métodos inferenciales de la estadística. El Econometría se nutre de tres disciplinas: Teoría Económica, Matemáticas y Estadística.

Diferencias modelo económico y econométrico

El **modelo econométrico** exige una especificación estadística más precisa de las variables que lo componen y exige una forma funcional definida, mientras que el económico es más teórico.
Los modelos econométricos se establecen, comúnmente, como relaciones aleatorias o no deterministas (estocásticas) entre variables, suponiéndose la existencia de elementos de azar, frente a las relaciones exactas o fijas que proponen los modelos económicos (determinista).

Modelo económico: $Consumo = f(Renta)$ **Model econométrico**: $Consumo_i = \beta_o + \beta_1 Renta_i + u_i$

CASUALIDAD Y NOCIÓN CETERIS PARIBUS

Si encontramos una relación entre dos variables no implica que una sea causa de la otra, significan que varían conjuntamente, que correlacionan. Para valorar la **causalidad** se debería hacer un experimento y recoger **datos experimentales** y no observacionales, modificando la VI y observando si se producen cambios en la VD, y controlando todas las demás variables de confusión. Únicamente se puede establecer causalidad bajo el supuesto **Ceteris Paribus** (manteniendo todos los demás factores constantes).

Ejemplo:

ETAPAS DEL ANÁLISIS ECONÓMICO EMPÍRICO

- Especificación:

- Estimación:

- Validación:

- Utilización o predicción:

ESTRUCTURA DE LOS DATOS ECONÓMICOS

Corte transversal o sección cruzada (x_i): Las observaciones corresponden a diferentes unidades económicas (individuos, empresas países ...) por un único período de tiempo.

Series temporales o longitudinales (x_t): Las observaciones corresponden a diferentes periodos de tiempo (años, meses, días ...) por un única unidad económica.

Datos de panel (x_{it}): Las observaciones corresponden a diferentes unidades económicas y para cada una de las diferentes unidades tenemos información para varios períodos de tiempo.

2. EL MODELO CLÁSICO DE REGRESIÓN

PRESENTACIÓN DEL MODELO

$Notación\ algebraica\quad y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$

$Notación\ matricial\quad Y = X\beta + u$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} (nx1) \quad X = \begin{bmatrix} 1 & x_{11} & \ldots & x_{k1} \\ 1 & x_{12} & & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & & x_{kn} \end{bmatrix} (nxK) \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} (Kx1) \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} (nx1)$$

Variable Y: endógena, explicada, dependiente, respuesta, regresando o causada.

Variables X: exógenas, explicativas, independientes, de control, regresoras o causantes.

Coeficientes: es la medida del efecto de x en y. En el modelo poblacional son fijos y desconocidos y el objetivo será estimarlos a partir de los datos de una muestra.

Variable u, residual, error o perturbación: $u_i = y_i - \hat{y}_i$

Ejemplo Eviews:

Dependent Variable: SALARIOM
Method: Least Squares
Sample (adjusted): 1 300
Included observations: 300

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	**767.2631**	180.4930	4.250929	0.0000
EDUC	**72.84627**	14.05438	5.183172	0.0000
EXPER	**20.06167**	2.207412	9.088321	0.0000
MALE	**93.45670**	13.54719	6.898605	0.0000

R-squared	**0.951607**	Mean dependent var	17702.50
Adjusted R-squared	0.949015	S.D. dependent var	1153.264
S.E. of regression	113.2656	F-statistic	367.0678
Sum squared resid	3797413.	Prob(F-statistic)	0.000000

Modelo simple: $y_i = \beta_0 + \beta_1 x_{1i} + u_i$

Modelo múltiple: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$

HIPÓTESIS DEL MODELO

H1: Linealidad

H2: Exogeneidad estricta

No debe haber relación lineal entre la perturbación aleatoria y las exógenas. $E(u/x) = 0$

H3: Ausencia de multicolinealidad

No es posible expresar una columna de la matriz X como una combinación lineal del resto de columnas.

Debe haber siempre más observaciones que parámetros.

H4: Perturbaciones esféricas

La perturbación aleatoria debe mantener su varianza constante (**homocesdasticidad**) y que no mantenga ningún tipo de tendencia (**no autocorrelación**)

- Homocedasticidad $E(u_i^2/X) = \sigma^2$

- No autocorrelación $E\left(u_i u_j/X\right) = 0$

$$Esfericidad \quad \rightarrow \quad E\,(uu'/X) = \sigma^2 \cdot I == \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Expresión matricial de los supuestos:

Linealidad: $Y = X\beta + u$.

X *no aleatoria:* X es una matriz deterministica.

Esperanza nula: $E(u) = 0_{n\times 1}$

Homoscedasticidad y no correlacion serial: $V(u) = \sigma^2 I_n$.

No multicolinealidad: $\rho(X) = K$.

ESTIMACIÓN MCO

Modelo poblacional $y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u_i$

Modelo muestral $y_i = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \cdots + \widehat{\beta_k} x_k + \widehat{u_i}$

Un buen estimador es insesgado, eficiente y consistente/robusto.

Mínimos Cuadrados Ordinarios (MCO): los coeficientes que minimizan la suma de los residuos al cuadrado.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{(nx1)} \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & & x_{kn} \end{bmatrix}_{(nxK)} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(Kx1)} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{(nx1)}$$

$$\hat{\beta}_{MCO} = (X'X)^{-1}(X'Y)$$

PROPIEDADES ALGEBRÁICAS DE LOS ESTIMADORES

P1.- $X'\hat{u} = 0$ **Regresores y residuos son ortogonales**

P2.- $\hat{y}'\hat{u} = 0$ **siendo** $\hat{y} = X\hat{\beta}_{MCO}$

P3.- $y'y = \hat{y}'\hat{y} + \hat{u}'\hat{u}$

P4.- $\hat{u}'\hat{u} = y'y - \hat{y}'\hat{y} = y'y - \hat{\beta}'X'X\hat{\beta} = y'y - \hat{\beta}'X'y$

P5.- Si el modelo tiene <u>término constante</u>:

$$\sum_{i=1}^{n} \hat{u}_i = 0$$

Demostraciones…..

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

PROPIEDADES ESTADÍSTICAS DE LOS ESTIMADORES

Linealidad: Los estimadores MCO son una combinación lineal de las perturbaciones aleatorias

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + U) = \beta + (X'X)^{-1}X'U$$

Insesgadez: la esperanza del estimador es su parámetro $E(\widehat{\beta_i}) = \beta_i$

$$E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'U) = \beta + (X'X)^{-1}X' \cdot E(U) = \beta + (X'X)^{-1}X' \cdot 0 = \beta$$

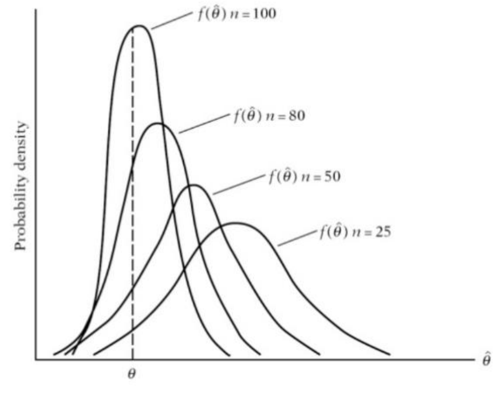
Optimalidad o eficiencia: la varianza del estimador es la mínima posible: $var(\widehat{\beta_{MCO}}) < var(\widehat{\beta_i})$**.** También se puede definir como que la varianza del estimador coincida con la cota Cramer-Raó (la mínima posible)

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1} = \begin{pmatrix} Var(\widehat{\beta_1}) & Cov(\widehat{\beta_2\beta_1}) & \dots & Cov(\widehat{\beta_k\beta_1}) \\ Cov(\widehat{\beta_1\beta_2}) & Var(\widehat{\beta_2}) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ Cov(\widehat{\beta_1\beta_k}) & \dots & \dots & Var(\widehat{\beta_k}) \end{pmatrix}$$

Teorema de Gauss-Markov: "Bajo los supuestos clásicos, los estimadores MCO son Estimadores Lineales, Insesgados y Óptimos (ELIO)"

Además de las propiedades anteriores, el estimador MCO es **consistente,** es decir, que cuando contamos con todos los datos y no con una muestra, el cálculo de MCO da como resultado los parámetros reales.

$$p\lim_{n\to\infty}(\widehat{\beta_i}) = \beta_i \quad p\lim_{n\to\infty}(var\,\widehat{\beta_i}) = 0$$



BONDAD DE AJUSTE

Siempre que el modelo tenga término independiente se puede afirmar que: $STC = SEC + SRC$

STC: Suma de cuadrados totales → $\sum(y_i - \bar{y})^2$

SEC: Suma de cuadrados explicado por el modelo → $\sum(\hat{y}_i - \bar{y})^2$

SRC: Suma de cuadrados no explicado o de residuos cuadrados → $\sum(y_i - \hat{y}_i)^2 = \sum \hat{u}_i^{\,2}$

El **coeficiente de determinación R^2**:

$$R^2 = \frac{SEC}{STC} = 1 - \frac{SRC}{STC} = \quad R^2 = r_{xy}{}^2(modelo\ simple)$$

El **coeficiente de determinación R^2 ajustado o corregido**:

$$\bar{R}^2 = 1 - \left[\frac{n-1}{n-K} \cdot (1 - R^2)\right] \qquad \bar{R}^2 \leq R^2$$

ERROR ESTÁNDAR DE LA REGRESIÓN

$$\widehat{u_i} = e_i = y_i - \hat{y}_i \qquad \hat{\sigma}^2 = Var(\widehat{u_i}) = \frac{SRC}{n-K} = \frac{\sum \hat{u}_i^{\,2}}{n-K} = \frac{\hat{u}'\hat{u}}{n-K} \qquad \boldsymbol{\hat{\sigma}} = \sqrt{\hat{\sigma}^2}$$

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{pmatrix} Var(\widehat{\beta_1}) & Cov(\widehat{\beta_2}\widehat{\beta_1}) & \dots & Cov(\widehat{\beta_k}\widehat{\beta_1}) \\ Cov(\widehat{\beta_1}\widehat{\beta_2}) & Var(\widehat{\beta_2}) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ Cov(\widehat{\beta_1}\widehat{\beta_k}) & \dots & \dots & Var(\widehat{\beta_k}) \end{pmatrix}$$

Ejemplo Eviews:

Dependent Variable: SALARIOM
Method: Least Squares
Sample (adjusted): 1 300
Included observations: 300

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	767.2631	**180.4930**	4.250929	0.0000
EDUC	72.84627	**14.05438**	5.183172	0.0000
EXPER	20.06167	**2.207412**	9.088321	0.0000
MALE	93.45670	**13.54719**	6.898605	0.0000

R-squared	**0.951607**	Mean dependent var	17702.50
Adjusted R-squared	**0.949015**	S.D. dependent var	1153.264
S.E. of regression	**113.2656**	F-statistic	367.0678
Sum squared resid	3797413.	Prob(F-statistic)	0.000000

FUNCIONES DE REGRESIÓN NO LINEALES

En muchas ocasiones nos encontramos en Economía con relaciones no lineales. En este caso muchas veces podemos transformar el modelo con logaritmos y conseguir un modelo lineal:

$$y_i = L_i^{\beta_0} * K_i^{\beta_1} * u_i \;\rightarrow\; \ln(y_i) = \beta_0 \ln(L_i) + \beta_1 \ln(K_i) + \ln(u_i)$$

Pero la interpretación de los coeficientes cambia:

1. $\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$ el coef. $\widehat{\beta_1}$ nos indica el cambio medio esperado en y por cada unidad de x.

2. $\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} \ln(x_i)$ el coef. $\widehat{\beta_1}$ nos indica que si x ↑ 1%, esperamos que y aumente en $\frac{\widehat{\beta_1}}{100}$ unidades

3. $\widehat{\ln(y_i)} = \widehat{\beta_0} + \widehat{\beta_1} x_i$ el coef. $\widehat{\beta_1}$ nos indica que si x ↑ 1u, esperamos que y aumente en $100 \cdot \hat{\beta}_1$%

4. $\widehat{\ln(y_i)} = \widehat{\beta_0} + \widehat{\beta_1} \ln(x_i)$ el coef. $\widehat{\beta_1}$ nos indica que si x ↑ 1%, esperamos que y aumente en $\hat{\beta}_1$%

Ejemplos:

MODELO CUADRÁTICO

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + u_i$$

El efecto marginal es la derivada respecto a la x: $\beta_1 + 2\beta_2 x_1$

Ejemplo:

EXÓGENAS CATEGÓRICAS

Si queremos incluir una variable categórica dicotómica en el modelo lo podemos hacer, siempre que la convirtamos en una variable "dummy" o fictícia, donde los dos valores sea 0 y 1. De esa manera, el coeficiente significará el aumento esperado de y por ser de la categoría 1 respecto a la 0.

Ejemplos:

Ejemplo Eviews:

Dependent Variable: L_NOTA_FINAL
Method: Least Squares
Sample (adjusted): 1 22
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	**0.564198**	0.563828	1.0010	0.3310
NOTA_PARCIAL	**0.115344**	0.371480	3.1050	0.0064
L_ASISTENCIA	**-0.891568**	0.866862	-1.0280	0.3181
EDAD	**0.013787**	0.019807	0.6961	0.4958
ACADEMIA	**0.356599**	0.144600	2.4660	0.0246

R-squared	0.658501	Mean dependent var	1.824567
Adjusted R-squared	0.578149	S.D. dependent var	0.420078
S.E. of regression	0.272841	F-statistic	8.195145
Sum squared resid	1.265520	Prob(F-statistic)	0.000713

MODELO DE DESVIACIONES RESPECTO A LA MEDIA

$$y_i - \bar{y} = \widehat{\beta_0} + \widehat{\beta_1}(x_{1i} - \overline{x_1}) + \widehat{\beta_2}(x_{2i} - \overline{x_2}) + \cdots + \widehat{\beta_k}(x_{ki} - \overline{x_k}) + \widehat{u_i}$$

$$Q = I_n - \frac{1}{n} 1_n 1_n^t \quad \text{simétrica e idempotente}$$

$$\widetilde{Y} \equiv QY = \begin{pmatrix} y_1 - \overline{y} \\ y_2 - \overline{y} \\ \vdots \\ y_n - \overline{y} \end{pmatrix}$$

$$\widetilde{X} \equiv QX = \begin{pmatrix} 0 & x_{11} - \overline{x}_1 & \dots & x_{1k} - \overline{x}_k \\ 0 & x_{21} - \overline{x}_1 & \dots & x_{2k} - \overline{x}_k \\ \vdots & \vdots & \vdots & \vdots \\ 0 & x_{n1} - \overline{x}_1 & \dots & x_{nk} - \overline{x}_k \end{pmatrix} = \left[0; \widetilde{X}_1\right]$$

Este dosier está hecho para seguir la clase de prueba.

Si te apuntas al curso te enviaremos por correo el dosier entero con todos los temas que faltan, ejercicios y exámenes de años anteriores

Más información en:

www.estadistix.com

Y si tienes cualquier consulta,
escríbenos un whatsapp al 644310902