UNIVERSIDAD CARLOS III DE MADRID

CURSO DE ECONOMETRÍA-I

2003/2004

Profesor: Francesc Marmol

TEMA I: DATOS ECONÓMICOS Y MODELIZACIÓN ECONOMÉTRICA

INGREDIENTES DEL TRABAJO EMPÍRICO EN ECONOMÍA

- Modelo económico
- Modelo econométrico
- Datos

MODELO ECONÓMICO

Un modelo económico formal consta de ecuaciones matemáticas que tratan de explicar o describir el comportamiento de ciertas variables de interés en términos de otro conjunto de variables (relaciones de causalidad).

Distinguiremos entre: Variables endógenas, cada una de ellas con su correspondiente ecuación que trata de explicar su comportamiento, y variables exógenas, cuyo comportamiento no es el objeto principal del estudio, y por tanto se considera explicado fuera del modelo.

Los modelos pueden ser uniecuacionales, si tienen solamente una ecuación, y multiecuacionales cuando tiene más de una ecuación.

Ejemplo 1: Curva de Engel

$$G = f(R)$$

siendo G = gasto en un bien, y R =renta.

Es un modelo uniecuacional, con una variable endógena (G) y una exógena (R).

Ejemplo 2: efectos de la formación (job training) en la productividad de los trabajadores.

$$sala = f(educ, exper, form),$$

donde $sala$ = salario por hora, $educ$ = años de educación, $exper$ =años de experiencia laboral, y $form$ = semanas dedicadas a formación.

Hay otros factores que influyen sobre el salario percibido, pero se espera que con estas tres variables se capte la esencia del problema.

En este ejemplo estamos ante un modelo uniecuacional, con una variable endógena ($sala$) y tres exógenas ($educ, form, exper$).

MODELO ECONOMÉTRICO

Un modelo econométrico se construye para cuantificar y contrastar las relaciones postuladas entre las variables económicas a partir de la evidencia empírica (los datos).

Características de un modelo econométrico:

- Reconoce el carácter aleatorio que gobierna las relaciones entre las variables.
- Postula formas funcionales para relacionar las diferentes variables a estudiar. Estas formas funcionales van a depender de parámetros, cantidades fijas, pero desconocidas, que describen las direcciones e intensidades de las relaciones entre las variables del modelo.

El modelo econométrico presupone que:

variable dependiente = parte explicada + parte no explicada

debiendo especificarse:

- las variables explicativas del modelo (parte explicada)
- su relación (estadística) con la parte no explicada o término de error.

DATOS ECONÓMICOS

Por definición, un análisis empírico necesita datos.

En este sentido, cabe recordar que en Economía, la mayoría de las series económicas están compuestas por datos no experimentales, esto es, datos que no se obtienen mediante experimentos controlados acerca de individuos, empresas o sectores de la economía, sino que se toman como dados.

El económetra depende generalmente de cifras que no pueden ser controladas directamente. Cifras sobre consumo, ingreso, inversión, ahorro, precios, etc., vienen recogidas por agencias oficiales y privadas, y son por tanto información no experimental. A veces se denominan datos observacionales para destacar el hecho de que el investigador es un recopilador pasivo de ellos. Además, es probable que tales cifras contengan errores de medición.

La información económica suele presentarse en diversas formas:

- datos de corte transveral o de sección cruzada (cross-section) o longitudinales,
- datos temporales,
- series temporales de secciones cruzadas,
- paneles de datos.

DATOS DE CORTE TRANSVERSAL O DE SECCIÓN CRUZADA (CROSS-SECTIONAL DATA)

Consiste en una muestra de individuos, hogares, empresas, ciudades, regiones, países u otro tipo de unidades, tomada en un momento determinado del tiempo

En este tipo de datos, el ordenamiento no importa para el análisis econométrico.

En el EJEMPLO 2 sobre determinación del salario, podríamos tener la siguiente información muestral de sección cruzada (en el ejemplo, referida a una muestra de 526 trabajadores para el año 2003:

$obs.$	$sala$	$educ$	$exper$	$sexo$	$ecivil$
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
525	11.56	16	5	0	1
526	3.50	14	5	1	0

donde $sala$ = salario en dólares por hora, $educ$ = años de escolaridad, $exper$ = años de experiencia laboral, $sexo$ = que vale 1 si la persona es mujer, 0 si es hombre, y $ecivil$ = que vale 1 es la persona está casada, 0 si no lo está.

NOTA: $sexo$ y $ecivil$ son variables denominadas binarias o dicotómicas, al tomar solamente dos valores. Son de gran importancia en el análisis econométrico al permitir el estudio según categorías (sexo y estado civil en este ejemplo).

A diferencia de las otras variables, no vienen dadas, sino que son creadas por el investigador con vistas a ampliar el estudio. Es por eso que también son conocidas como variables artificiales.

DATOS TEMPORALES

Son observaciones de una o más variables, hechas en el tiempo (días, semanas, meses, años,....)

A diferencia de los datos de sección cruzada, en los datos temporales el orden de los datos es muy importante.

Como ejemplo de datos temporales tenemos la siguiente base de datos:

$obs.$	$año$	$salamin$	cob	$desem$	PIB
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
47	1996	3.35	58.1	18.9	4281.6
48	1997	3.35	58.2	16.8	4496.7

siendo $salamin$ = salario mínimo medio anual (dólares por hora), cob = porcentaje de trabajadores cobrando el salarios mínimo, $desem$ = índice de desempleo, PIB = producto interior bruto (millones de dólares).

TODO JUNTO:

Supongamos el caso más simple donde tenemos una variable endógena (Y) y una variable exógena (X).

De acuerdo con el modelo económico, se supone que están relacionadas (con lo que se establece la causalidad):

$$Y = f(X)$$

A veces la teoría económica nos dice algo sobre la relación entre ambas variables:

$$\frac{df(X)}{dX}, \frac{d^2 f(X)}{dX^2} \substack{\geq \\ \leq} 0$$

Pasemos ahora al modelo econométrico. Se supone en general que las variables son aleatorias.

Como variables aleatorias que son, en general tendremos que

$$Y \neq f(X),$$

por lo que se introduce un nuevo término aleatorio, ε,denominado término de error o perturbación, de tal forma que

$$Y = f(X) + \varepsilon.$$

Se entiende que este término de error contiene todos aquellos efectos no capturados por X y que pueden ayudar a explicar el comportamiento de la variable Y.

Con esto tenemos:

- $f(X)$: parte explicada por el modelo,
- ε : parte no explicada por el modelo (error).

A continuación, hay que darle una forma funcional a la parte explicada. La forma más común es la lineal (que da lugar al denominado modelo lineal):

$$f(X) = \alpha + \beta X.$$

En este sentido, se demuestra en el siguiente tema que la “mejor relación $f(.)$” viene dada por la función de esperanza condicional (FEC):

$$f(X) = E(Y|X),$$

de tal forma que el modelo (denominado modelo de regresión simple) queda:

$$Y = E(Y|X) + \varepsilon$$

El siguiente paso es especificar la relación estadística de X, ε, o bien de X, Y.

En general,

$$E(Y|X) \neq \alpha + \beta X.$$

Caso particular: Esperanza condicional lineal:

$$E(Y|X) = a + bX.$$

En este caso,

$$a = \alpha, b = \beta.$$

El siguiente paso es especificar las propiedades de la muestra de datos. Se supone que disponemos de n parejas de datos:

$$(Y_1, X_1), (Y_2, X_2), ..., (Y_n, X_n),$$

generadas cada una de ellas a partir del modelo econométrico (modelo de regresión):

$$Y_i = E(Y_i|X_i) + \varepsilon_i.$$

No confundir con la muestra de datos disponible

$$(y_1, x_1), (y_2, x_2), ..., (y_n, x_n),$$

que no es más que una de las muchas posibles realizaciones que se podrían haber obtenido.

En el caso de tener más de una variable exógena (o regresor), tendremos:

$$Y = E(Y|X_1, X_2, ..., X_k) + \varepsilon$$

denominado modelo de regresión múltiple. Cuando la FEC es lineal, tenemos el modelo de regresión lineal múltiple.

En este curso trabajaremos con modelos lineales,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon,$$

o bien, al trabajar con datos muestrales,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i.$$

Con esto, se hacen explícitos los parámetros:

$$\beta_0, \beta_1, \beta_2, ..., \beta_k.$$

Ejemplo 2: Ecuación de salarios

En este caso tendremos el siguiente modelo lineal:

$$sala = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 form + \varepsilon,$$

donde en este caso el término ε contendría factores tales como la capacidad innata, la calidad de la instrucción, los antecedentes familiares, etc.

$\beta_0, \beta_1, \beta_2, \beta_3$ serían parámetros de este modelo.

Ejemplo 4: ¿De qué depende la decisión de delinquir?

(G. Becker, "Crime and punishment: An economic approach", *Journal of Political Economy*, 1968)

Es un modelo para describir la conducta delictiva desde el punto de vista de un típico problema de asignación de recursos, donde tenemos unos beneficios y unos costes a tener en cuenta. El modelo económico que se deriva es el siguiente:

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7),$$

donde (la lista de variables no es exhaustiva):

Y = horas dedicadas a las actividades delictivas,

X_1 = "salario" por una hora de actividades delictivas,

X_2 = salario por hora en un empleo legal,

X_3 = otros ingresos, aparte de la delincuencia y el empleo,

X_4 = probabilidad de ser capturado,

X_5 = probabilidad de ser declarado culpable, si se es capturado,

X_6 =sentencia esperada, si se es hallado culpable,

X_7 = edad.

Tenemos:

$$\begin{aligned} delinc &= \beta_0 + \beta_1 sala_m + \beta_2 oting + \beta_3 frearr + \beta_4 frecon \\ &\quad + \beta_5 senprom + \beta_6 edad + \varepsilon \end{aligned}$$

donde $delinc$ es alguna medida de la frecuencia de las actividades delictivas, $sala_m$ es el salario base que puede ganarse en un empleo legal, $oting$ son los ingresos por otras fuentes (bienes, herencias, etc.), $frearr$ es la frecuencia de arrestos por infracciones anteriores (para aproximarse a la probabilidad de captura), $frecon$ es la frecuencia de condenas y $senprom$ es la duración promedio de las sentencias, despúes de la condena.

Las constantes $\beta_0, \beta_1, \beta_2, ..., \beta_6$ son parámetros de este modelo.

Dado el modelo de regresión lineal múltiple,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon,$$

uno de los pasos más importantes es saber qué se está haciendo, es decir, saber interpretar los parámetros $\beta_0, \beta_1, \beta_2, ..., \beta_k$.

El siguiente paso es estimar los parámetros, en el modelo lineal

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i.$$

Veremos que, bajo ciertos supuestos sobre las $X's$ y los $\varepsilon's$, es posible obtener “buenos”(en sentido estadístico) estimadores de $\beta_0, \beta_1, \beta_2, ..., \beta_k$:

$$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k.$$

Para una muestra de datos disponible,

$$
\begin{aligned}
&(y_1, x_{11}, x_{21}, ..., x_{k1}), \\
&(y_2, x_{12}, x_{22}, ..., x_{k2}), \\
&\vdots \\
&(y_n, x_{1n}, x_{2n}, ..., x_{kn}),
\end{aligned}
$$

vamos a obtener estimaciones (números)

$$\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_k,$$

con los que,

- interpretar el modelo (en lugar de β_i se utiliza $\widehat{\beta}_i$)

- predecir el comportamiento de Y para diferentes valores de las $X's$:

$$y^0 = \widehat{\beta}_0 + \widehat{\beta}_1 x_1^0 + \widehat{\beta}_2 x_2^0 + \cdots + \widehat{\beta}_k x_k^0,$$

- análisis de política económica: Por ejemplo, dado $\hat{\beta}_0, \hat{\beta}_1 \hat{\beta}_2, ... \hat{\beta}_k,\ x_2^0, ..., x_k^0,$

¿qué valor de x_1 hace falta para que $\hat{y}$ valga y^0?

Solución: Despejando,

$$x_1^0 = \frac{1}{\hat{\beta}_1}\left(y^0 - \hat{\beta}_0 - \hat{\beta}_2 x_2^0 - \cdots - \hat{\beta}_k x_k^0\right).$$

EJEMPLO: FUNCIÓN KEYNESIANA DE CONSUMO

Modelo:

$$Y = \beta_1 + \beta_2 X, \qquad 0 < \beta_2 < 1$$

- donde Y = gasto de consumo, X = ingreso, y donde β_1 y β_2 son, respectivamente, los coeficientes de la constante y de la pendiente. El coeficiente β_2 mide la PMC.

Tenemos datos sobre el consumo y la renta de USA para los años 1980-1991:

$familias$	$consumo$	$ingreso$
1	2477.1	3776.3
2	2476.9	3843.1
3	2503.7	3760.3
4	2619.4	3906.6
5	2746.1	4148.5
6	2865.8	4279.8
7	2969.1	4404.5
8	3052.2	4539.9
9	3.162.4	4718.6
10	3223.3	4838.0
11	3260.4	4877.5
12	3240.8	4821.0

La estimación del modelo de regresión lineal simple

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon$$

nos permite obtener las estimaciones:

$$\begin{aligned} \hat{\beta}_2 &= 0{,}722, \\ \hat{\beta}_1 &= -242{,}9 \end{aligned}$$

Y ahora podemos preguntarnos, por ejemplo, si se espera un PIB real de USA para 1994 de 6000 (mil millones de dólares), ¿cuál es el gasto de consumo esperado para 1994?. De acuerdo con nuestro modelo, el gasto de consumo esperado para 1994 será de:

$$\begin{aligned} \hat{Y}_{1994} &= -242{,}9 + 0{,}722 X_{1994} \\ &= -242{,}9 + 0{,}722(6000) = 4089{,}1 \end{aligned}$$

miles de millones de dólares.

Asimismo, si las autoridades han concluído que para mantener la tasa de desempleo de 1994 en el 6.5 % hace falta que el gasto en consumo de 1994 sea de 4000 (mil millones de dólares de 1987). La pregunta es: ¿Qué nivel de ingreso garantizará ese nivel de consumo en 1994?. La respuesta es fácil:

$$\begin{aligned} \hat{Y}_{1994} &= -242{,}9 + 0{,}722 X_{1994} \\ &\Rightarrow 4000 = -242{,}9 + 0{,}722 X_{1994} \\ &\Rightarrow X_{1994} = (4000 + 242{,}9)/0{,}722 = 5877 \end{aligned}$$

(aproximadamente) miles de millones de dólares.