

ECONOMETRIA

Tema 1: Datos Económicos y Modelización Econométrica

César Alonso

Universidad Carlos III de Madrid



¿Qué es la ECONOMETRÍA?

- Es una disciplina basada en el desarrollo de **modelos probabilísticos** y de métodos de **inferencia estadística**, para el estudio de **relaciones económicas**, la **contrastación** de teorías económicas o la **evaluación e implementación de políticas** económicas o empresariales.
- Para ello, tiene en cuenta la particular **naturaleza de los datos económicos**,
- La econometría combina elementos de:
 - Teoría económica
 - Matemáticas
 - Estadística



- La econometría nos puede ayudar a responder cuestiones tales como:
 - Efectos de un programa de formación en la productividad o en el salario de los trabajadores.
 - Rendimientos de diversas estrategias de inversión.
 - Efectos de una campaña publicitaria.
 - Efecto del tamaño de la clase en el rendimiento escolar.
 - Impacto de los seguros médicos en el uso de servicios médicos.



- Uno de los objetivos principales de la Econometría es realizar **análisis causal**, es decir:
 - Analizar cualitativa y cuantitativamente cómo ciertos factores afectan a una variable asociada a un fenómeno económico de interés.
- El análisis causal permite:
 - Determinar los efectos de ciertas políticas (caracterizados por cambios en determinados factores que afectan al fenómeno de interés).
 - Caracterizar y cuantificar la relación de comportamiento entre variables económicas, de acuerdo con lo que sugiere la teoría económica.
 - Simular los efectos de políticas alternativas.



- **Ejemplo 1:** Efecto causal de la educación en el salario.
 - Es el incremento salarial que conseguiría un individuo de la población objeto de estudio si, manteniéndose constantes sus demás características, tuviera un nivel mayor de educación (por ejemplo, un año adicional, tener o no un título universitario, etc.).
 - Esta cuestión es de interés:
 - práctico: inferir las consecuencias de (i) aumentar el gasto público o las subvenciones a la educación; (ii) incrementar el número de años de escolarización obligatoria; etc.
 - teórico: la decisión de escolarización puede derivarse de modelos económicos basados en la teoría del capital humano.
 - *Los estudios empíricos encuentran que el efecto causal de tener un título universitario suponen un salario que es en promedio un 40% mayor que el de los que no tienen un título universitario.*



- La econometría ha evolucionado como una disciplina independiente de la estadística, al centrarse en los problemas inherentes a la recopilación y al análisis de datos económicos, que típicamente son **datos no experimentales** (también llamados datos observacionales).
- Los **datos experimentales** se recopilan en entornos controlados, generalmente en un laboratorio, y habitualmente en las **ciencias naturales**.
- En un experimento (apropiadamente diseñado) se controlan los valores de aquellos factores susceptibles de afectar al fenómeno que se quiere estudiar.



Datos experimentales y datos observacionales (ii)

- En un experimento ideal se puede:
 - alterar de forma controlada el valor de un determinado factor que puede afectar al fenómeno objeto de estudio, manteniendo constantes todo lo demás.
 - atribuir (y cuantificar) sin ambigüedad las variaciones en la variable objeto de estudio a las variaciones en dicho factor.
 - En tal caso, el efecto obtenido es el **efecto ceteris paribus** o **efecto causal** del factor cuyo valor hemos alterado.
- Con datos no experimentales (u observacionales):
 - tanto los valores de la variable objeto de estudio como de los factores susceptibles de afectarla están fuera de nuestro control
 - no podemos interpretar un co-movimiento o correlación entre variables como un efecto causal.



- En el caso infrecuente de que dispongamos de datos que reproduzcan el experimento apropiado, sí podemos inferir causalidad.
- Al plantearnos un problema empírico, resulta útil reflexionar sobre cuál sería el experimento apropiado y en qué o por qué nuestros datos se alejan de dicho experimento.
- Los métodos econométricos pueden permitir estimar el efecto “ceteris paribus” e inferir causalidad entre variables, simulando una situación similar a la del experimento apropiado.



- **Ejemplo 2:** Efecto de los fertilizantes sobre el rendimiento de los cultivos.
 - La cantidad de fertilizante empleado es sólo uno de los factores que afectan al rendimiento. (otros: humedad, insolación, calidad del terreno, etc.)
 - Pregunta de interés: si elegimos una parcela cualquiera de 1 Ha. e incrementamos el fertilizante en una determinada cantidad, ¿en cuánto aumentaría el rendimiento?
 - Podríamos realizar un experimento, con los siguientes pasos:
 1. Elegir varias parcelas de terreno de igual superficie (p.ej., 1 Ha.)
 2. Aplicar diferentes cantidades de fertilizante a cada parcela.
 3. Medir el rendimiento (en kg. de soja) de cada parcela.



El diseño ideal requeriría que las parcelas fueran idénticas en todo excepto en la cantidad de fertilizante. PERO puede haber diferencias en calidad del terreno, humedad, insolación, etc., lo que hace imposible ese diseño.

- Un diseño factible que permite identificar las diferencias en el rendimiento de las parcelas como efecto del fertilizante consistiría en:
 - **asignar** a cada parcela la cantidad de fertilizante **de forma completamente aleatoria**, ignorando por completo las demás características de la parcela.



- **Ejemplo 1 (cont.):** Efecto causal de la educación en el salario
 - Podríamos pensar en la posibilidad de realizar un experimento similar al de los fertilizantes para el caso de la educación.
 - La pregunta es ahora: si a una persona elegida al azar en la población le incrementamos en un año su nivel de educación, ¿en cuánto aumentaría su salario?
 - De nuevo, nos interesa el efecto *ceteris paribus*, que supone que todos los demás factores susceptibles de afectar al salario se mantienen invariantes para la persona elegida, a la que se aumenta en un año su nivel de educación.
 - El experimento ideal consistiría en que un planificador social:
 1. Elige varias personas entre la población.
 2. Les asigna distintos años de educación de forma aleatoria.
 3. Mide sus salarios.



- Si los **niveles de educación** de que disfruta cada persona **se pudieran establecer de forma independiente** de otras características que afectan a la productividad (como experiencia laboral o capacidad innata), podremos atribuir las diferencias en los salarios a diferencias en los niveles de educación.
- Pero este ejercicio no es factible en la realidad.



- Los datos experimentales son mucho más difíciles (o imposibles) de obtener en las **ciencias sociales**. Aunque es posible elaborar ciertos experimentos sociales, los experimentos controlados que se necesitan pueden:
 - no ser factibles,
 - ser muy costosos,
 - ser éticamente rechazables.
- Por el contrario, los datos no experimentales se generan en un entorno no controlado.
- Por tanto, las realizaciones (valores observados) tanto de la variable objeto de estudio como de los factores susceptibles de afectarla no están sujetos a control.



- Emplea datos para cuantificar una relación o contrastar una teoría.
- Aspecto previo esencial: plantear de manera cuidadosa el problema que deseamos estudiar. Por ejemplo:
 - contrastar uno o varios aspectos de una determinada teoría
 - contrastar o evaluar los efectos de una determinada política pública, de una determinada estrategia empresarial, etc.
- El diseño de un trabajo empírico se estructura en los siguientes elementos:
 - Modelo económico
 - Modelo econométrico
 - Datos



- La **teoría económica** propone modelos que explican el comportamiento de una o varias variables, Y_1, \dots, Y_m , en función de otra u otras variables, X_1, \dots, X_k , que se determinan fuera del modelo.
- El modelo planteado puede ser más o menos formal.
 - Modelo formal: ecuaciones matemáticas que describen relaciones entre variables.
(Basados en maximización de la utilidad -consumidores- o beneficio -empresas-, sujeto a diversas restricciones
Ejemplo:
 - demanda de bienes - en términos de precio del bien y de sus sustitutos y complementos, de la renta y de características de los consumidores;
 - demanda de factores productivos (trabajo, capital, consumos intermedios) - en términos de los precios de los factores.
 - Modelo informal: su descripción se inspira en la teoría y en aspectos más intuitivos.



- El modelo económico, de manera más o menos formal, permite expresar Y a través de una ecuación del tipo:

$$Y - f(X_1, \dots, X_k) = 0$$

- O, de forma equivalente,

$$Y = f(X_1, \dots, X_k)$$

- Habitualmente, la función $f(\cdot)$ no queda caracterizada por la teoría. Dicha función depende del modelo de decisión subyacente, que rara vez se conoce.



- En ambos casos, podemos clasificar las variables en endógenas y exógenas.
 - los valores de las variables endógenas se determinan dentro del modelo;
 - los valores de las variables exógenas están determinados fuera del modelo.
- En el caso más sencillo, podemos considerar una variable endógena, Y , y un conjunto de K variables exógenas X_1, \dots, X_K .



Ejemplos: modelo económico

- Función de oferta y demanda de un bien agrícola:

$$\text{Demanda : } Q = f_1(P, R) \qquad \text{Oferta : } Q = f_2(P, LI)$$

- $Y_1 = Q, Y_2 = P, X_1 = R, X_2 = LI.$
 - $Q =$ cantidades, $P =$ precios, $R =$ renta y $LI =$ lluvias.
- Modelo de capital humano:

$$\text{Ec. salarial : } W = f_1(S, EX) \qquad \text{Ec. educación : } S = f_2(S_P, S_M)$$

- $Y_1 = W, Y_2 = S, X_1 = EX, X_2 = S_P, X_3 = S_M$
 - $W =$ salario, $S =$ años de estudio, $EX =$ años de experiencia laboral, $S_P =$ nivel educativo del padre, $S_M =$ nivel educativo de la madre.
- Curva de Engel:

$$G = f(GT)$$

- $Y_1 = G =$ gasto en un bien, $X_1 = GT =$ gasto total.



- Permite **cuantificar y contrastar las relaciones entre variables** postuladas por los modelos económicos a partir de los datos (la evidencia empírica).
- Recoge la **naturaleza estocástica** que gobierna las relaciones entre variables.
- Establece cómo tratar el problema de las variables que, aunque postuladas por el modelo económico, no pueden observarse.
- Parametriza $f(\cdot)$, la **forma funcional**, estableciendo una relación entre Y y X_1, \dots, X_K por medio de **parámetros constantes**.



Modelo econométrico: Naturaleza estocástica

- Dada la naturaleza aleatoria de las variables económicas (tanto exógenas como endógenas),
 - es de esperar que cada uno de los datos no cumpla exactamente cada una de las ecuaciones que especifica un modelo económico.
 - Sin embargo, siempre se puede encontrar un conjunto de funciones que satisfaga:

$$E [Y - f (X_1, \dots, X_k)] = 0$$

- Esta naturaleza aleatoria de las relaciones económicas se puede expresar a través de un **error inobservable** de la siguiente manera:

$$Y - f (X_1, \dots, X_k) = \varepsilon$$

o, de forma equivalente

$$Y = f (X_1, \dots, X_k) + \varepsilon$$



- Para cuantificar las relaciones entre variables económicas se propone una **forma funcional específica**:

$$Y = f(X_1, \dots, X_k; \mathbf{f}) + \varepsilon$$

donde $\mathbf{f} = (\beta_1, \dots, \beta_k)$ es un conjunto de parámetros desconocidos.

- La naturaleza del modelo y la interpretación de los parámetros dependen de las propiedades del término de error (que veremos luego).
- El carácter experimental o no experimental de dichos datos determina en buena medida las propiedades del término de error.
 - Con datos experimentales, el término de error es meramente aleatorio (producto del azar);
 - Con datos observacionales, el término de error puede contener, además, factores relevantes no observados.



- Oferta y demanda de un bien agrícola:
 - Demanda: $Q = \beta_{01} + \beta_{11}R + \beta_{21}P + \varepsilon_1$
 - Oferta: $Q = \beta_{02} + \beta_{12}P + \beta_{22}LI + \varepsilon_2$
 - Variables endógenas: Q (cantidades), P (precios).
 - Variables exógenas: R (renta), LI (lluvia).
- Modelo de capital humano:
 - Ecuación salarial: $W = \beta_0 + \beta_1S + \beta_2EX + \beta_3EX^2 + \varepsilon_1$
 - Ecuación de educación: $S = \alpha_0 + \alpha_1S_P + \alpha_2S_M + \varepsilon_2$
 - Variables endógenas: W (salario), S (años de estudio).
 - Variables exógenas: EX (experiencia), S_P (educ. padre), S_M (educ. madre).



- A partir de unos datos, empleando un modelo econométrico estaremos interesados en cuantificar y responder a preguntas sobre los parámetros del modelo.
- Ejemplos:
 - Función de oferta y demanda de un bien agrícola:
 - ¿Es β_{22} muy grande? ¿Es distinto de cero?
 - Modelo de capital humano:
 - ¿Es $\beta_3 = 0$? ¿Qué signo tiene? ¿Es $\alpha_1 = \alpha_2$?



- Dependiendo del problema a analizar, del interés de investigador y de la naturaleza de los datos se formulan diferentes tipos de modelos econométricos.
- Algunos tipos de modelos econométricos:
 - Univariantes versus multivariantes
 - Uniecuacionales versus multiecuacionales



Modelo econométrico

Modelos univariantes vs. multivariantes

- Los modelos univariantes emplean exclusivamente la información que contienen los datos disponibles de una única variable.
- Los modelos multivariantes emplean la información que contienen los datos disponibles de varias variables.

Ejemplo: Deseamos realizar una predicción para la evolución futura de la tasa de inflación de un país π .

- Si disponemos sólo de la historia pasada de dicha tasa puedo formular un modelo univariante para π , de manera que $\pi = f(\text{pasado de } \pi)$.

- Si disponemos de datos sobre otras variables que inciden sobre dicha tasa (tipos de interés, variación de la oferta monetaria, crecimiento económico, etc.) puedo formular un modelo multivariante para π , es decir, $\pi = f(\text{pasado de } \pi, \text{ otras variables})$.



Modelo econométrico

Modelos uniecuacionales vs. multiecuacionales

- Los modelos uniecuacionales formulan la existencia de una relación entre una variable (endógena) y otras que inciden sobre ésta (exógenas).
- Los modelos multiecuacionales formulan la existencia de una relación simultánea entre diversas variables

Ejemplo:

Modelos uniecuacionales:

- $\pi = f(\text{pasado de } \pi, \text{ otras variables})$.

- $G = \alpha + \beta GT + \varepsilon$.

Modelo multiecuacional:

- función de oferta $Q = \beta_{02} + \beta_{12}P + \beta_{22}LI + \varepsilon_2$ y demanda

$Q = \beta_{01} + \beta_{11}R + \beta_{21}P + \varepsilon_1$ de un bien agrícola

- Q = cantidades, P = precios, R = renta y LI = llluvias



- Las variables de un modelo económico representan aspectos del comportamiento de los agentes económicos en el ámbito individual o agregado.
- Los economistas observamos, directamente o de forma aproximada, el comportamiento y características de los agentes. Esta observación constituye la evidencia empírica, los datos.
- Tipos de datos:
 - Sección cruzada
 - Series temporales
 - Datos de Panel



- Proviene de encuestas sobre familias, individuos, empresas, etc., en un momento del tiempo. La ordenación no importa.
- Ejemplos: Encuesta de Presupuestos Familiares, Encuesta de Población Activa.
- En general, cabe esperar que las observaciones sean independientes entre sí.
- Son los datos en los que nos centraremos en este curso



Ejemplo: Muestra de 1000 individuos españoles en el año 2002

Individuo (i)	Edad $_i$	Renta $_i$	Sexo $_i$	Ecivil $_i$
1	68	12000	1	1
2	43	24324	0	1
3	23	17345	0	0
\vdots	\vdots	\vdots	\vdots	\vdots
999	63	54987	1	1
1000	32	67677	1	0



- Observación de una o varias variables a lo largo del tiempo.
- El orden cronológico y la frecuencia (anual, trimestral, mensual, diaria) de los datos son importantes.
- En general, cabe esperar que exista dependencia entre observaciones (en particular, las series económicas suelen presentar dependencia temporal, de manera que el pasado nos da una idea de lo que podemos esperar en el futuro cercano).

Ejemplos: IPC, PIB, Ventas anuales de una empresa, etc.



Ejemplo: Datos anuales de tasas de inflación (tasa anual de variación del IPC), tasa de desempleo y crecimiento (tasa anual de variación del PIB) para un país determinado

Año (t)	π_t	u_t	dy_t
1975	3.8	5.8	3.6
1976	5.4	6.4	2.8
1977	5.3	8.9	2.9
\vdots	\vdots	\vdots	\vdots
2001	1.3	6.3	1.5
2002	1.1	6.9	1.2



- Consisten en una serie temporal por cada miembro de una sección cruzada. Para cada miembro de la sección cruzada el orden cronológico es importante.
- El número de series temporales suele ser muy pequeño con relación al número de unidades de sección cruzada.
- Los datos de panel son diferentes de las series temporales de secciones cruzadas (cohortes), que consisten en diferentes secciones cruzadas en distintos momentos del tiempo en las que, en general, en cada período tendremos individuos diferentes.



Datos

Datos de panel o datos longitudinales (cont.)

Ejemplo: Muestra de 525 empresas españolas, años 1995-2000

Empresa (i)	Año (t)	Beneficios $_{it}$	Empleados $_{it}$	Cotiza $_{it}$
1	1995	200	150	0
⋮	⋮	⋮	⋮	⋮
1	2000	4566	356	1
2	1995	2624	345	1
⋮	⋮	⋮	⋮	⋮
2	2000	7500	345	1
⋮	⋮	⋮	⋮	⋮
525	1995	456890	1456	1
⋮	⋮	⋮	⋮	⋮
525	2000	134893	1321	0



Ejemplo de análisis empírico: Modelo económico

- Becker (1968). "Crime and punishment: an economic approach", *Journal of Political Economy* 76, pp. 169-217.
- Decisión de cometer delitos –¿ decisión de asignación de recursos considerando beneficios / costes de cometer delitos frente a los de actividades legales.

$$Y = f(X_1, X_2, X_3, \dots, X_7)$$

- Y = horas de actividades delictivas
- X_1 = salario-hora de la actividad delictiva
- X_2 = salario-hora del empleo legal
- X_3 = otras rentas
- X_4 = probabilidad de ser detenido
- X_5 = probabilidad de ser condenado
- X_6 = duración esperada de la condena
- X_7 = edad



Ejemplo de análisis empírico: Modelo econométrico

- Impone una forma funcional (en general, la lineal).
- Algunas variables no son observadas y debemos elegir variables aproximativas.
- Además, puede haber otros muchos factores que afecten a la actividad criminal.

$$\text{crimen} = \beta_0 + \beta_1 \text{salario} + \beta_2 \text{otrarenta} + \beta_3 \text{frecde} + \beta_4 \text{frecon} + \beta_5 \text{durmed} + \beta_6 \text{edad} + \varepsilon,$$

- *crimen* = medida de actividad criminal (Y)
- *salario* = salario hora empleo legal (X_2)
- *otrarenta* = otros ingresos (X_3)
- *frecde* = frecuencia de detenciones en el pasado (X_4)
- *frecon* = frecuencia de condenas (X_5)
- *durmed* = duración media de las condenas (X_6)
- *edad* = edad (X_7)
- ε = error (incluye errores de medida y factores no observados por actividad delictiva, carácter del individuo, entorno familiar)

