TEMA 1. Introducción al análisis empírico de variables económicas.

Profesor: Pedro Albarrán Pérez

Universidad de Alicante. Curso 2010/2011.

Contenido

La importancia de saber analizar datos

- En el mundo laboral, muchos profesionales de la economía, la empresa y las finanzas deben **trabajar con datos**:
  - para contrastar explicaciones teóricas sobre el comportamiento de la economía o sobre el comportamiento de activos financieros
  - para anticipar los efectos de cambios en las políticas públicas (tipo de interés, prestaciones por desempleo) o en la gestión de la empresa (exportar a nuevos mercados, inversiones)
  - para predecir la evolución de la producción, los mercados financieros, los tipos de cambio, los precios de los bienes
  - para la analizar el estado y desempeño de una empresa
- En este curso, estudiaremos las herramientas que se utilizan en la práctica
  - motivando los conceptos generales con ejemplos particulares y
  - desarrollando las habilidades informáticas necesarias.

Tipos de Datos

1. **Series Temporales:** los datos de nuestra variable están ordenados temporalmente
   - Ejemplos: el PIB, tipos de interés, ventas de una empresa, precios de activos financieros, etc.
   - La *frecuencia* de los datos puede ser anual, trimestral, mensual, semanal, diaria, cada hora, cada minuto, etc.
2. **Datos de sección cruzada:** los datos se refieren a unidades individuales (en un momento del tiempo)
   - las unidades puede ser personas, empresas, países, activos en una cartera de inversión
   - el orden de las unidades no importa
3. **Datos de Panel:** información sobre unidades individuales a lo largo del tiempo
   - combinan dimensión temporal y de corte transversal
   - Ejemplo: países de la UE15 entre 1992-2010, rentabilidad de los activos en una cartera durante los últimos 10 años, etc.

- **Información cuantitativa:** los datos establecen un *valor numérico* para cada unidad y/o periodo de tiempo
  - Ej.: el precio de un activo son 10 euros
  - Ej.: las ventas en agosto han sido 20.000 euros
- **Información cualitativa:** los datos informan de una cualidad o una elección
  - Ejemplos:
    - Hombre/Mujer
    - Sí/No
    - Bueno/Regular/Malo
    - Coche/Autobús/Tren
  - A veces, los datos se convierten en valores numéricos:
    - Variables binarias (dummies): Sí=1, No=0
    - Variables categóricas: Bueno=3, Regular=2, Malo=1
  - Aunque no siempre, algunas veces la información cualitativa puede interpretarse ordinalmente (Bueno/Regular/Malo)
  - NUNCA se pueden interpretar cardinalmente como la información cuantitativa

Transformaciones de los Datos

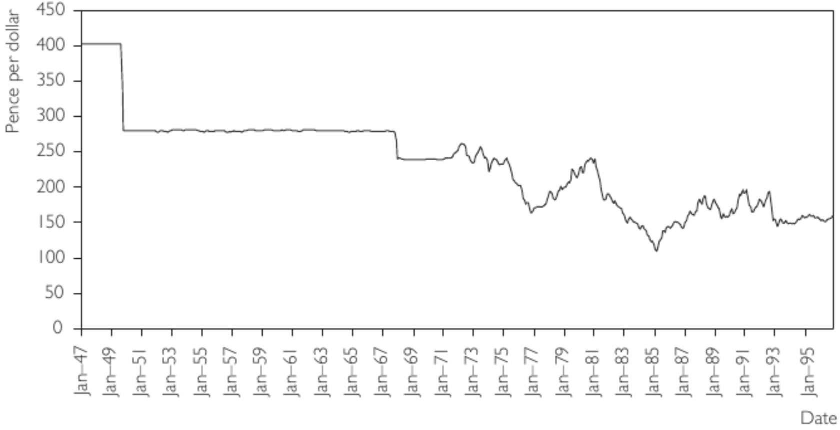
- En el curso, supondremos que los datos de interés (la variable Y) está directamente disponible
- En la práctica, el profesional obtiene los datos (“en bruto”) de una fuente de información y los transforma para realizar su análisis empírico
- Frecuentemente se tienen que calcular **ratios**:
  - X = beneficios de la empresa, W = número de acciones $\Longrightarrow Y = \frac{X}{W}$ = beneficio por acción
- Muchas datos informan sobre el **nivel** de una variable: ej., el precio de una acción o de un bien, ventas, PIB, etc.
- Pero podemos estar interesados en cómo cambia esa variable en el tiempo, es decir, **la tasa de crecimiento**
- Otras transformaciones puede ser necesarias según el contexto: deflactar variables, calcular rentabilidades, trabajar con números índice, etc.

¿Cómo analizar la información?

- Para cualquier análisis empírico, resulta crucial ofrecer un resumen relevante de los datos
    - los datos originales (cientos de observaciones) no son informativos por sí mismos
    - de hecho, la econometría simplemente desarrolla métodos por los cuales la información de los datos se *resumen de forma informativa*
- También se pueden usar tablas y gráficos para presentar la información de forma muy útil.

Gráficos de series temporales

La forma más simple de representar la evolución de una serie temporal es un gráfico de líneas de los datos en función del tiempo

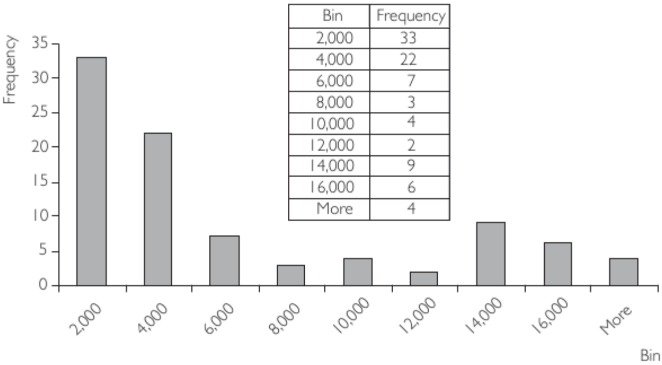


Histograma

Con datos de sección cruzada, un gráfico de líneas NO es informativo porque el orden no importa.

Podemos representar la **distribución** de distintas categorías de valores

1. *Elegir* los intervalos de valores
2. Calcular las frecuencias muestrales en cada intervalo
3. Realizar un gráfico de barras (de las frecuencias en cada intervalo)



Bin	Frequency
2,000	33
4,000	22
6,000	7
8,000	3
10,000	4
12,000	2
14,000	9
16,000	6
More	4

Este gráfico puede ayudar con varias cuestiones de interés:

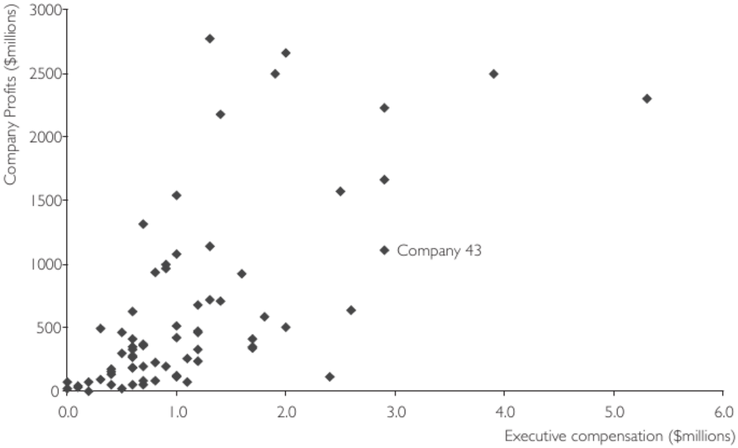
- ¿Cómo es la distribución de renta per cápita entre países?
- ¿Cómo de importante es la desigualdad?

NOTA: el histograma depende de la elección de la amplitud y el número de intervalos

- muchos programas realizan automáticamente histogramas eligiendo automáticamente este aspecto
- distintas elecciones pueden llevarnos a conclusiones diferentes (puede que incorrectas)

Gráfico de Dispersión

- Muchas veces, estaremos interesados en analizar *relaciones* entre dos o más variables
- Aunque no es fácil mediante gráficos, se pueden explorar mediante gráficos de dispersión



En este ejemplo, parece existir una relación positiva entre el salario de los ejecutivos y los beneficios que obtienen las empresas

- Relación positiva: valores altos de una variable suelen asociarse con valores altos de la otra variable
- Relación negativa: valores altos de una variable suelen aparecer junto con valores pequeños de la otra variable

En muchas ocasiones encontraremos valores "atípicos": observaciones que no se ajustan al patrón general.

Estadísticos descriptivos

Una distribución se puede resumir en una *tabla de frecuencias* (equivalente al histograma)

- útil cuando la variable tiene pocos valores diferentes
- PERO puede contener demasiada información para ser interpretable

Estadísticos descriptivos: números que resumen las propiedades de la distribución de una variable económica

- ofrecen precisión numérica frente al impacto visual de los gráficos

(1) **Medidas de localización:** intuitivamente, indican el centro de la distribución, individuo promedio o típico
  - Media (o promedio): $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$

(2) **Medidas de dispersión:** muestran la variabilidad/dispersión de la distribución en torno al valor central, desigualdad
  - desviación estándar: $s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}$
  - varianza$=s^2$

- Otros estadísticos: mediana, moda, percentiles (cuartiles, deciles, etc.); rango intercuartílico, etc.

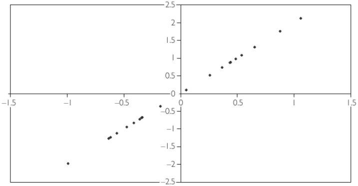
Correlación

Correlación: mide numéricamente la relación entre dos variables

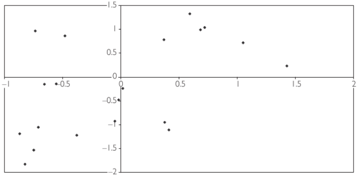
$$r = r_{YX} = r_{XY} = \frac{\sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sqrt{\sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2} \sqrt{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2}}$$

$-1 \leqslant r \leqslant 1$

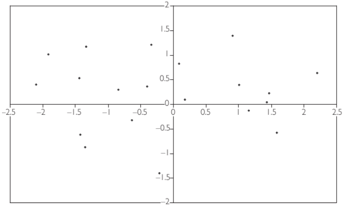
- $r > 0$ indica relación positiva entre X e Y
  - la relación es más fuerte cuanto mayor sea r
  - $r = 1$ indica una correlación positiva perfecta
- $r < 0$ indica relación negativa entre X e Y
  - la relación es más fuerte cuanto menor sea r
  - $r = -1$ indica una correlación negativa perfecta
- $r = 0$ indica ausencia de relación entre X e Y
- $r_{XX} = 1$



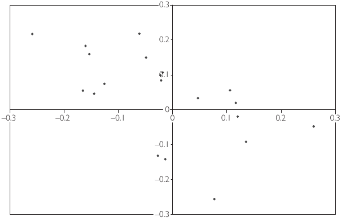
$r = 1$



$r = 0{,}51$



$r = 0$



$r = -0{,}58$

Correlación entre variables variables

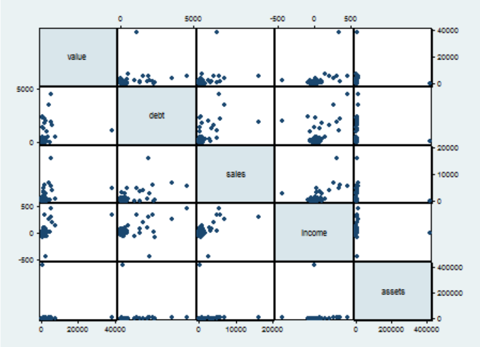
- La correlación es una propiedad que relaciona DOS variables
- Pero frecuentemente querremos trabajar con varias variables
  - los modelos de *regresión* son los más adecuados en este caso
- Se pueden calcular las correlaciones de CADA PAR de variables, que se presentan en una matriz de correlaciones
  - si tenemos muchas variables pueden ser muchas correlaciones
- Pero no conocemos si la relación entre dos variables es directa o indirecta (por medio de una tercera)

Ejemplo: relación entre salario de ejecutivos, beneficio de la empresa, cambio en las ventas y cambio en la deuda

	compensation	profits	changesales	changedebt
compensation	1			
profits	0.6591	1		
changesales	-0.0524	-0.1269	1	
changedebt	0.3016	0.0886	0.2489	1

Ejemplo: relación entre valor de mercado de una empresa, su deudas, sus ventas, sus ingresos netos y el valor contable de sus activos

	value	debt	sales	income	assets
value	1				
debt	0.322	1			
sales	0.410	0.638	1		
income	0.486	0.511	0.624	1	
assets	0.014	0.041	0.025	0.019	1



¿Por qué están dos variables correlacionadas?

- Cuando dos variables están correlacionadas, se puede pensar en términos de influencia o causalidad
- Las correlaciones pueden sugerir pero no establecer por sí misma la existencia de **relación causal**.

Ejemplo: Cigüeñas y nacimientos en una muestra de 17 países europeos

	area	**storks**	humans	birth_rate
area	1.0000			
storks	0.5793	1.0000		
	(0.0148)			
humans	0.8122	0.3542	1.0000	
	(0.0001)	(0.1630)		
birth_rate	0.9225	**0.6203**	0.8512	1.0000
	(0.0000)	**(0.0079)**	(0.0000)	

Causalidad Directa vs. Causalidad Indirecta

- *Causalidad directa o inmediata*: una variable influye real e independientemente sobre otra
  - un cambio en una variable provocará necesariamente un cambio en la otra sin que interfieran cambios en otras variables
  - El ejemplo anterior sobre la relación entre ventas y valor de mercado puede interpretarse como causalidad directa
- *Causalidad indirecta* (intermediada o aproximada): la relación aparente entre dos variables se debe a una tercera que sí causa a ambas
  - una mayor área permite observar más cigüeñas y que haya más población humana y, por tanto, nacimientos
  - sólo una mayor deuda orientada a aumentar las ventas aumenta el valor de mercado
- Una correlación (debida a causalidad indirecta) siempre puede usarse para predecir.
- El sentido común o los razonamientos y teorías económicas nos debe ayudar a establecer si una correlación puede interpretarse como relación causal.

Modelo de Regresión Lineal

- Supongamos una relación lineal

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t$$

- En notación vectorial

$$Y_t = \boldsymbol{X}_t' \beta + u_t$$

  - $\boldsymbol{X}_t = (1, X_{2t}, .., X_{kt})'$ es un vector $(k \times 1)$ de variables aleatorias y u_t es una variable aleatoria
  - $\beta = (\beta_1, \beta_2, .., \beta_k)'$ un vector $(k \times 1)$ de parámetros (poblacionales, verdaderos)

- Para datos de corte transversal, suponemos que $\left\{(Y_t, \boldsymbol{X}_t')'\right\}_{t=1}^{T}$ **i.i.d.**
  - por tanto, $\{u_t\}_{t=1}^{T}$ también es i.i.d.

- Para datos de series temporales, suponemos que $\left\{(Y_t, \boldsymbol{X}_t')'\right\}_{t=1}^{T}$es un **proceso estocástico estacionario**
  - por tanto, $\{u_t\}_{t=1}^{T}$ también es un proceso estocástico estacionario

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t$$

Supuestos básicos:

(a) $E\left(u_t | \boldsymbol{X_t}\right) = 0, \quad \Longleftrightarrow E\left(Y_t | \boldsymbol{X_t}\right) = \boldsymbol{X_t'}\beta$

- Esto implica: $E\left(u_t \boldsymbol{X_t}\right) = 0$

(b) $Var\left(u_t | \boldsymbol{X_t}\right) = \sigma^2$ (homocedasticidad)

(b)' NO ES NECESARIO: $Var\left(u_t | \boldsymbol{X_t}\right) = E\left(u_t^2 \middle| \boldsymbol{X_t}\right) = g\left(\boldsymbol{X_t}\right) = \sigma^2 \omega\left(\boldsymbol{X_t}\right)$

(c) $\Sigma = E(X_t X_t')$ es definida positiva

- (no hay multicolinearidad perfecta)

(d) Una forma concreta de la dependencia entre observaciones del término de error $\{u_t\}_{t=1}^{T}$

(d.1) Ausencia de correlación: $Cov\left(u_t, u_s | X_t, X_s\right) = 0, \quad t \neq s$

(d.2) $u_t \sim AR(p), \quad u_t \sim MA(q), \ldots$

Estimador MCO

Utilizando los supuestos (a)

$$E\left[\mathbf{X}_t u_t\right] = E\left[\mathbf{X}_t\left(Y_t - \mathbf{X}_t'\beta\right)\right] = E\left(\mathbf{X}_t Y_t\right) - E\left(\mathbf{X}_t\mathbf{X}_t'\right)\beta = 0$$

y (c) para despejar β

- Se puede identificar el parámetro β en *la población*

$$\beta = \left[E\left(\mathbf{X}_t\mathbf{X}_t'\right)\right]^{-1} E\left(\mathbf{X}_t Y_t\right)$$

- El estimador MCO se obtiene en *nuestra muestra* como un estimador del método de los momentos:

$$\widehat{\beta} = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t Y_t\right)$$

En notación matricial,

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{u}$$

donde

- $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}$ y $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$ son vectores $(T \times 1)$
- $\mathbb{X} = \begin{pmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_T' \end{pmatrix}$ es una matriz $(T \times k)$

El estimador de MCO es

$$\widehat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\boldsymbol{Y}$$

Se puede definir el vector $(T \times 1)$ de residuos

$$\boldsymbol{e} = \boldsymbol{Y} - \mathbb{X}\widehat{\beta} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{pmatrix}$$

donde $e_t = Y_t - X_t'\widehat{\beta}, \quad t = 1, 2, .., T$

Propiedades del estimador MCO

Bajo (algunos de) los supuestos anteriores el estimado de MCO tiene "buenas" propiedades

- Insesgado: $E\left(\widehat{\beta}\middle|\mathbb{X}\right) = E\left(\widehat{\beta}\right) = \beta$
- Varianza

$$Var\left(\widehat{\beta}\middle|\mathbb{X}\right) = \left(\mathbb{X}'\mathbb{X}\right)^{-1}\mathbb{X}'\Omega\mathbb{X}\left(\mathbb{X}'\mathbb{X}\right)^{-1}$$

donde

$$\Omega = Var\left(\boldsymbol{u}\right) = Var\left[\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}\right]$$

  - Bajo homocedasticidad y ausencia de correlación, $\Omega = \sigma^2 I_I$, por lo que $Var\left(\widehat{\beta}\middle|\mathbb{X}\right) = \sigma^2\left(\mathbb{X}'\mathbb{X}\right)^{-1}$
  - Los programas informáticos también calculan *errores estándar robustos* (y otros estadísticos de contraste) usando la formula general

- Es consistente

$$\widehat{\beta}_{MCO} \xrightarrow{p} \beta$$

- Es asintóticamente normal

$$\widehat{\beta}_{MCO} \simeq N\left(\beta, Var\left(\widehat{\beta}\right)\right)$$

Esta propiedad se puede explotar para hacer inferencia sobre el vecto de parámetros β:

- Hipótesis nula y alternativa

$$\begin{array}{lll} H_0 : \beta_j = \beta_j^0 & H_0 : \beta_j = \beta_j^0 & H_0 : \beta_j = \beta_j^0 \\ H_1 : \beta_j \neq \beta_j^0 & H_1 : \beta_j > \beta_j^0 & H_1 : \beta_j < \beta_j^0 \end{array}$$

- y el estadístico de contraste

$$t = \frac{\widehat{\beta}_j - \beta_j^0}{SE\left(\widehat{\beta}_j\right)} \underset{H_0}{\simeq} N(0,1)$$

- Regla de rechazo: rechazar H_0 (con un nivel de significación α)

$$\begin{array}{llllll} (a) & H_0 : \mathbb{R}\boldsymbol{\beta} = r & (b) & H_0 : \mathbb{R}\boldsymbol{\beta} = r & (c) & H_0 : \mathbb{R}\boldsymbol{\beta} = r \\ & H_1 : \mathbb{R}\boldsymbol{\beta} \neq r & & H_1 : \mathbb{R}\boldsymbol{\beta} > r & & H_1 : \mathbb{R}\boldsymbol{\beta} < r \\ & & & & & \\ & \text{si } |t| > C_{\alpha/2} & & \text{si } t > C_\alpha & & \text{si } t < -C_\alpha \end{array}$$

  - donde $\Pr(N(0,1) > C_\alpha) = \alpha$ y $\Pr\left(N(0,1) > C_{\alpha/2}\right) = \alpha/2$

Efectos Marginales

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t$$

- ¿Cómo se interpretan los coeficientes β del modelo de regresión?
- Si el modelo tiene constante

$$\beta_1 = E\left[Y | X_{2t} = 0, \ldots, X_{kt} = 0\right]$$

- El coeficiente de una variable continua (digamos, X_2) es un efecto marginal

$$\beta_2 = \frac{\delta}{\delta X_{2t}} E\left[Y | X_{2t}, X_{3t} = x_3, \ldots, X_{kt} = x_k\right]$$

- El coeficiente de una variable discreta (digamos, X_k) también es un efecto marginal (discreto)

$$\begin{aligned} \beta_k &= E\left[Y | X_{2t} = x_2, , \ldots, X_{k-1,t} = x_{k-1}, X_{kt} = x_k + 1\right] \\ &- E\left[Y | X_{2t} = x_2, , \ldots, X_{k-1,t} = x_{k-1}, X_{kt} = x_k\right] \end{aligned}$$

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t$$

$$\beta_j = \frac{\delta}{\delta X_{jt}} E\left[Y | X_{2t} = x_2, \ldots, X_{j-1,t} = x_{j-1}, X_{j,t}, X_{j+1,t} = x_{j+1}, \ldots, X_{kt} = x_k \right]$$

El coeficiente β_j informa del

- cambio esperado en la variable dependiente Y
- ante un cambio de unidad o infinitesimal de la variable explicativa X_j
- manteniendo el resto en valores dados/constantes (*ceteris paribus*)

$$\beta_j \approx \left. \frac{\Delta E\left[Y | X_{2t}, \ldots, X_{kt} \right]}{\Delta X_{jt}} \right|_{X_{2t}=x_2, \ldots, X_{j-1,t}=x_{j-1}, X_{j+1,t}=x_{j+1}, \ldots, X_{kt}=x_k}$$

Por tanto, el coeficiente de un modelo de regresión (correctamente especificado) ofrece el **efecto causal directo** de X_{jt} sobre Y_t

Modelos no lineales

- El modelo de regresión *lineal* puede estimarse por MCO siempre que la especificación sea lineal en los parámetros
    - (no necesariamente lineal en las variables del modelo)
- Es decir, se pueden incluir como variable dependiente y como variables explicativas **transformaciones no lineales** de las variables originales

$$\ln Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_3 + u_t$$

$$Y_t = \beta_1 + \beta_2 X_{2t}^3 + \beta_3 \frac{1}{X_3} + u_t$$

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{2t}^2 + \beta_4 \frac{1}{X_{2t}} + \beta_5 X_{3t} + u_t$$

- El estimador de MCO de estos modelos tiene las "buenas" propiedades que ya hemos discutido
- PERO debemos tener cuidado al interpretar los coeficientes y los efectos marginales

Modelos no lineales: efectos marginales

- Seguimos interesados en los **efectos marginales**: ofrecen el efecto causal
- PERO el efecto marginal de interés de la variable X_j NO coincide necesariamente con un único coeficiente β_j

En modelos con variables transformadas en logaritmos, los efectos marginales se interpretan como (semi-)elasticidades

1. $\ln Y_t = \beta_1 + \beta_2 \ln X + u_t$

$$\beta_2 \approx \frac{\Delta \ln Y}{\Delta \ln X} \approx \frac{\Delta Y/Y}{\Delta X/X} = \frac{\%\Delta Y}{\%\Delta X} = elasticidad$$

2. $\ln Y_t = \beta_1 + \beta_2 X + u_t$

$$\beta_2 \approx \frac{\Delta \ln Y}{\Delta X} \approx \frac{\Delta Y/Y}{\Delta X} = \frac{\%\Delta Y}{\Delta X}$$

3. $Y_t = \beta_1 + \beta_2 \ln X + u_t$

$$\beta_2 \approx \frac{\Delta Y}{\Delta \ln X} \approx \frac{\Delta Y}{\Delta X/X} = \frac{\Delta Y}{\%\Delta X}$$

En el modelo,

$$Y_t = \beta_1 + \beta_2 \frac{1}{X_{2t}} + \beta_3 X_{3t} + u_t$$

¿cuál es el efecto marginal (causal) de X_2 sobre Y?

$$\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, X_{3t} \right] = \beta_2 \frac{1}{X_{2t}^2}$$

Se pueden extraer varias conclusiones de este ejemplo (válidas en general)

1. el efecto marginal NO coincide con un coeficiente del modelo (en este caso, β_2)
2. el efecto marginal NO es constante
   - depende, en este caso, de X_{2t}
   - individuos con distinto valor de X_{2t} tendrán distinto cambio esperado en Y_t ante un mismo cambio (infinitesimal) en X_{2t}

Frecuentemente, se utilizan modelos lineales con polinomios e interacciones de las variables explicativas

1. $$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{2t}^2 + \beta_4 X_{2t}^3 + \beta_5 X_{3t} + u_t$$
   $$\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, X_{3t} \right] = \beta_2 + 2\beta_3 X_{2t} + 3\beta_4 X_{2t}^2$$
   - Depende de varios parámetros
   - Depende del valor de X_{2t}
2. $$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{2t} X_{3t} + \beta_4 X_{3t} + u_t$$
   $$\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, X_{3t} \right] = \beta_2 + \beta_3 X_{3t}$$
   - Depende de varios parámetros
   - Depende del valor de otra variable: X_{3t}
3. $$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{2t}^2 + \beta_4 X_{2t} X_{3t} + \beta_5 X_{2t}^2 X_{3t} + \beta_6 X_{3t} + u_t$$
   $$\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, X_{3t} \right] = \beta_2 + 2\beta_3 X_{2t} + \beta_4 X_{3t} + 2\beta_5 X_{2t} X_{3t}$$
   - Depende de varios parámetros
   - Depende del valor de X_{2t} y de X_{3t}
4. Existen muchas otras combinaciones posibles

Sesgo de variable omitidas

- Una especificación incorrecta del modelo implica que los efectos marginales NO son interpretables como efectos causales
  - el efecto marginal de X_j es manteniendo constante el resto de *variables incluidas* en el modelo
  - por tanto, no se tienen en cuenta las variables (relevantes) omitidas

Ejemplo:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

- Sólo nos interesa el efecto de X_1 (se estima un modelo reducido):

$$Y_t = \delta_0 + \delta_1 X_{1t} + \epsilon_t$$

- Si $\delta_1 = \beta_1$, este modelo reducido sí proporciona el verdadero efecto causal de interés de X_1 sobre Y
- Se puede mostrar fácilmente que

$$\delta_1 = \beta_1 + \beta_2 \frac{Cov\left(X_1, X_2\right)}{Var\left(X_1\right)}$$

- En general, omitir una variable sesga el coeficiente: no coincide con el efecto causal

$$\delta_1 = \beta_1 + \beta_2 \frac{Cov\left(X_1, X_2\right)}{Var\left(X_1\right)}$$

- Sin embargo, omitir la variable X_2 no sesga el efecto causal de X_1 si
    1. realmente X_2 no afecta a Y: no es relevante porque $\beta_2 = 0$
    2. se tiene que $Cov\left(X_1, X_2\right) = 0$.

- El argumento (no la formula concreta) se puede generalizar al caso de múltiples variables
    - si las variables omitidas (relevantes) no están correlacionadas con las variables explicativas incluidas en la regresión, los coeficientes estimados para éstas no estarán sesgados.

Predicción: R^2

- Puesto que el modelo de regresión establece que

$$E\left[\, Y_t \mid X_{2t}, \ldots, X_{kt}\right] = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt}$$

- Se puede predecir el valor esperado de la variable dependiente (en función de las variables explicativas):

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + \cdots + \widehat{\beta}_k X_{kt}$$

- El valor observado de Y_t se descompone en una parte predicha por el modelo y un residuo

$$Y_t = \widehat{Y}_t + e_t$$

- Se puede demostrar que la "varianza" de los datos observados también se puede descomponer en dos partes:

$$\begin{aligned} \widehat{Var}\left(Y_t\right) &= \widehat{Var}\left(\widehat{Y}_t\right) + \widehat{Var}\left(e_t\right) \\ \sum_{t=1}^{T}\left(Y_t - \overline{Y}\right)^2 &= \sum_{t=1}^{T}\left(\widehat{Y}_t - \overline{Y}\right)^2 + \sum_{t=1}^{T} e_t^2 \\ SCT &= SCE + SCR \end{aligned}$$

$$SCT = SCE + SCR$$

- Se define el R^2 como

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

  - mide la proporción de la variabilidad de Y_t explicada por el modelo: $0 \leqslant R^2 \leqslant 1$
  - cuanto más próxima a 1, mayor es la capacidad predictiva del modelo (menores residuos=parte no explicada)
- Si usamos el modelo de regresión para predecir, NO estamos interesados en saber si los coeficientes tienen interpretación causal o no
  - en el ejemplo anterior, podemos predecir Y_t sólo con X_{1t} igual de bien que con X_{1t} y X_{2t}
  - nos da igual si Y_t es aumenta con X_{1t} directamente o indirectamente a través de X_{2t}
- Un mayor valor de R^2 indica mejor capacidad predictiva, pero NO nos informa sobre la interpretación causal de los coeficientes del modelo.

Regresores Binarios

- El modelo de regresión admite datos cuantitativos o cualitativos
- Una variable binaria (o "dummy") es una forma de transformar variables cualitativos en cuantitativos

$$D_t = \begin{cases} 1, & \text{si el individuo es mujer} \\ 0, & \text{si el individuo es hombre} \end{cases}$$

- La interpretación de los efectos marginales de un regresor binario son un caso particular de lo que hemos visto para variables discretas:

$$Y_t = \beta_1 + \beta_2 X_{2t} + .. + \beta_2 X_{k-1t} + \beta_k D_{kt} + u_t$$

$$\begin{aligned} \beta_k &= E\left[Y \mid X_{2t} = x_2, , \ldots, X_{k-1,t} = x_{k-1}, D_{kt} = 1\right] \\ &- E\left[Y \mid X_{2t} = x_2, , \ldots, X_{k-1,t} = x_{k-1}, D_{kt} = 0\right] \end{aligned}$$

- Por tanto, β_k es el aumento esperado en el valor de Y_t por pertenecer al grupo $D_{kt} = 1$ frente al grupo $D_{kt} = 0$, *ceteris paribus*
  - el valor esperado de Y_t para el grupo con $D_{kt} = 1$ menos su valor esperado para el grupo con $D_{kt} = 0$

El uso de variables binarios permite que cada grupo tenga un valor esperado diferente (dado un mismo valor del resto de variables)

- En un caso simple,

$$Y_t = \beta_1 + \beta_2 D_t + u_t$$

- Se tiene que
  - $E\left[Y_t | D_t = 1\right] = \beta_1 + \beta_2$
  - $E\left[Y_t | D_t = 0\right] = \beta_1$

También se puede interactuar una variable binaria con otra variable del modelo; por ejemplo,

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{2t} D_t + \beta_4 X_{3t} + u_t$$

- En este caso, el efecto marginal de X_{2t} depende de D_t

$$\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, D_t, X_{3t}\right] = \beta_2 + \beta_3 D_t$$

- Esto implica que el efecto marginal de X_{2t} es diferente para el grupo $D_{kt} = 1$ y para el grupo $D_{kt} = 0$
  - $\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, D_t = 1, X_{3t}\right] = \beta_2 + \beta_3$
  - $\frac{\delta}{\delta X_{2t}} E\left[Y_t | X_{2t}, D_t = 0, X_{3t}\right] = \beta_2$