INTRODUCCIÓN A LA ECONOMETRÍA

3er curso LE y LADE
Tema 4

Dpto. de Econometría y Estadística (EA3)

UPV—EHU

4.1 Variables Ficticias. De nición y uso en el MRLG.

Variables Ficticias: De nición

- Var explicativa cualitativa $\rightsquigarrow$ sub-muestras $T_1, T_2, \ldots$ según categoría o características
- ejemplos:
  - vars puramente cualitativas:
    - difs individuales: sexo, raza, estado civil, etc.
    - difs temporales: estación, guerra/paz, etc.
    - difs espaciales: países, regiones, urbano/rural, etc.
  - vars cuantitativas por tramos: renta, edad, etc.
- Acordar: no podemos utilizar vars cualitativas. . . entonces substituir por **vars cticias** . . .
- Def. de Variable Ficticia:

$$D_{jt} = \begin{cases} 1, & \text{si } t \in \text{categoria } j \text{ ;} \\ 0, & \text{si no.} \end{cases}$$

$$\Rightarrow D_{jt} = \mathcal{I}(t \in T_j)$$

1 VC con 2 categorías

Consumo $=$ $f($[cte], renta, sexo$)$

$\downarrow$ Y_t $\quad \downarrow$ $[1]$ $\quad \downarrow$ R_t $\quad$ sexo: $Masc$ / Fem

$S_{1t} = \mathcal{I}(t \in M) \quad S_{2t} = \mathcal{I}(t \in F)$

Muestra:

t	Y	cte	R	S
1	Y_1	1	R_1	M
2	Y_2	1	R_2	F
3	Y_3	1	R_3	F
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
T	Y_T	1	R_T	M
		¿ X ?		

$\Rightarrow$

t	Y	cte	R	S_1	S_2
1	Y_1	1	R_1	1	0
2	Y_2	1	R_2	0	1
3	Y_3	1	R_3	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
T	Y_T	1	R_T	1	0
			X		

En principio: substituir VC por tantas VFs como categorías haya.

Trampa de Var Ficticias: 1 var cualitativa

- Modelo: $Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \gamma_2 S_{2t} + u_t$
- Problema (Trampa VF):
  X es una matriz $(T \times 4)$, pero

  $S_1 + S_2 = [1]$ (c.l. exacta) $\Rightarrow \mathrm{rk}(X) = 3 < 4$ (*es decir* MC perfecta)
- $\Rightarrow \det(X'X) = 0$
  $\Rightarrow$ ¡$(X'X)^{-1}$ no existe!! y
  ¡no se puede calcular $\widehat{\beta}$!!
- Solución General: eliminar UNA de las cols. que causa el problema:[1] o S_1 o S_2 .
- (POSIBLE Solución: eliminar intercepto . . . pero. . .

Solución: VF sin categoría

SOLUCIÓN MÁS HABITUAL: eliminar categoría: *p.ej.* F (S_2):

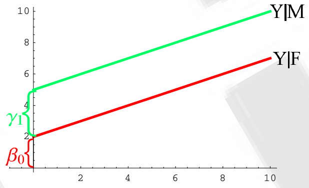
- Modelo a estimar:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \cancel{\gamma_2} S_{2t} + u_t \\ &= \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + u_t \end{aligned}$$

- Modelos Submuestras:

$$E(Y_t | S = F) = \beta_0 + \beta_1 R_t$$
$$E(Y_t | S = M) = \beta_0 + \beta_1 R_t + \gamma_1 \quad \Rightarrow$$

sin categoría F

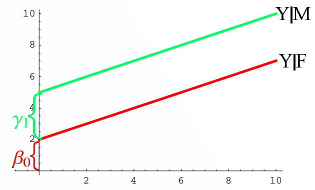


- Interpretación Coeficientes:

$$E(Y_t | R_t = 0, S = F) = \beta_0$$
$$E(Y_t | S = M) - E(Y_t | S = F) = \gamma_1$$

Interpretación Coeficientes

sin categoría F



$$E(Y_t | S = M) - E(Y_t | S = F) = \gamma_1$$
$$E(Y_t | R_t = 0, S = F) = \beta_0$$

- es decir,

β_0 = consumo esperado Mujeres (base) si $R_t = 0$.

γ_1 = dif. consumo esperado de Hombres (frente a base = Mujeres).

$\beta_1 = \Delta$ consumo si $\Delta R_t = 1$ (*c.p.*).

Recuérdese: Este caso sólo significa diferentes interceptos para cada categoría.
Nota: Eliminar una categoría $\rightsquigarrow$ transformarla en base de referencia.

Contrastes Habituales con 1 VC

Hipótesis: variable cualitativa (Sexo) no significante (no afecta el Consumo) *es decir* $Masc$ y Fem mismo Consumo:

- Modelo Sin Restringir

$$Y_t = \beta_0 + \beta_1 R_t + \gamma S_{1t} + u_t$$

- Hipótesis: $H_0 : \gamma = 0$ frente a $H_a : \gamma \neq 0$
- Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t + u_t$$

- Véase el habitual Estadístico t (o Estadístico F basado en SCR).

1 VC con 2 cats + 1 VC con 3 cats

$$\begin{array}{ccccc} \text{Consumo} = & f([\text{cte}], & \text{renta}, & \text{sexo}, & \text{territorio CAV}) \\ \downarrow & \downarrow & \downarrow & \swarrow\ \searrow & \swarrow \downarrow \searrow \\ Y_t & [1] & R_t & Masc \quad Fem & A \quad B \quad G \end{array}$$

$S_{1t} = \mathcal{I}(t \in M)$
$S_{2t} = \mathcal{I}(t \in F)$

$T_{1t} = \mathcal{I}(t \in A)$
$T_{2t} = \mathcal{I}(t \in B)$
$T_{3t} = \mathcal{I}(t \in G)$

Muestra:

t	Y	cte	R	S	T
1	Y_1	1	R_1	M	B
2	Y_2	1	R_2	F	G
3	Y_3	1	R_3	F	B
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
T	Y_T	1	R_T	M	A
		¿ X ?			

$\Rightarrow$

t	Y	cte	R	S_1	S_2	T_1	T_2	T_3
1	Y_1	1	R_1	1	0	0	1	0
2	Y_2	1	R_2	0	1	0	0	1
3	Y_3	1	R_3	0	1	0	1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$			$\vdots$	
T	Y_T	1	R_T	1	0	1	0	0
				X				

Recuérdese: En principio, sustituir var cualitativa por tantas vars Ficticias como categorías haya.

Trampa de Var Ficticias: 2 vars cualitativas

- Modelo: $Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \gamma_2 S_{2t} + \delta_1 T_{1t} + \delta_2 T_{2t} + \delta_3 T_{3t} + u_t$
- Problema (Trampa VF):
  X es una $(T \times 7)$ matriz, pero

$$S_1 + S_2 = T_1 + T_2 + T_3 = [1]$$

(2 c.l. exactas) $\Rightarrow \mathrm{rk}(X) = 5 < 7$ (*es decir* MC perfecta)

- $\Rightarrow \det(X'X) = 0$
  $\Rightarrow$ ¡$(X'X)^{-1}$ no existe!! y
  ¡no se puede calcular $\widehat{\beta}$!!

- Solución General: eliminar UNA de las col's que causa el problema: [1] o (S_1 o S_2) o (T_1 o T_2 o T_3).

Solución: VF sin combinación de categorías

SOLUCIÓN MÁS HABITUAL:
eliminar última categoría de cada VF: S_2 y T_3:

- Modelo a estimar:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \cancel{\gamma_2 S_{2t}} + \delta_1 T_{1t} + \delta_2 T_{2t} + \cancel{\delta_3 T_{3t}} + u_t$$
$$= \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

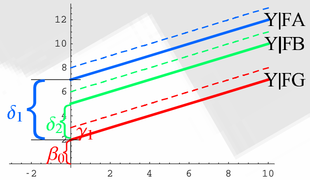
- <6>Modelos Submuestras:

	$S = M$	$S = F$	$M - F$
$T = A$	$\beta_0 + \beta_1 R_t + \gamma_1 + \delta_1$	$\beta_0 + \beta_1 R_t + \delta_1$	γ_1
$T = B$	$\beta_0 + \beta_1 R_t + \gamma_1 + \delta_2$	$\beta_0 + \beta_1 R_t + \delta_2$	γ_1
$T = G$	$\beta_0 + \beta_1 R_t + \gamma_1$	$\beta_0 + \beta_1 R_t$	γ_1
$A - G$	δ_1	δ_1	
$B - G$	δ_2	δ_2	
$A - B$	$\delta_1 - \delta_2$	$\delta_1 - \delta_2$	

Coe ciente Interpretación

sin categorías F ni G

$$\mathrm{E}(Y_t | S = M) - \mathrm{E}(Y_t | S = F) = \gamma_1$$
$$\mathrm{E}(Y_t | T = A) - \mathrm{E}(Y_t | T = G) = \delta_1$$
$$\mathrm{E}(Y_t | T = B) - \mathrm{E}(Y_t | T = G) = \delta_2$$
$$\mathrm{E}(Y_t | R_t = 0, S = F, T = G) = \beta_0$$



- <5>es decir,

β_0 = consumo esperado Mujeres G (base) if $R_t = 0$.

γ_1 = dif. consumo esperado Hombres frente a Mujeres .

δ_1 = dif. consumo esperado A frente a G.

δ_2 = dif. consumo esperado B frente a G.

$\beta_1 = \Delta$ consumo si $\Delta R_t = 1$ (*c.p.*).

Recuérdese: Este caso sólo significa diferentes interceptos para cada categoría.
Recuérdese: Eliminar una (combinación de) categoría(s)
$\rightsquigarrow$ transformarla en base de referencia.

Contrastes habituales con 2 VCs

Hipótesis: Variable Sexo no afecta Consumo
(pero territorio de residencia puede que sí)

- Modelo sin Restringir:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t}$$
$$+ \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

(γ_1 = dif. esperada Consumo $Masc$ frente a Fem)

- Hipótesis: $H_0 : \gamma_1 = 0$ frente a $H_a : \gamma_1 \neq 0$
- Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t$$
$$+ \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

- Usar el habitual t Estadístico (o F Estadístico basado en SCR)

Otros Contrastes habituales con 2 VCs

- Modelo sin Restringir (sin S_2 ni T_3):

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

  - Recuérdese: γ_1 es dif. Consumo esperado $Masc$ frente a Fem (base) δ_1 y δ_2 son dif. Consumo esp. de A y B frente a G (base)
- Hipótesis: Mismo Consumo para todos (independientemente del Sexo y Territorio):
  - $H_0 : \gamma_1 = \delta_1 = \delta_2 = 0$
  - Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t + u_t$$

- Hipótesis: Territorio de Residencia no afecta Consumo (pero $Masc$ frente a Fem puede que si):
  - $H_0 : \delta_1 = \delta_2 = 0$
  - Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + u_t$$

Otros Contrastes habituales con 2 VCs

- Modelo sin Restringir (sin S_2 ni T_3):

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + \delta_2 T_{2t} + u_t$$

  - Recuérdese: δ_1 y δ_2 son dif. Consumo esperados de A y B frente a G (base)
- Hipótesis: Residentes del mismo sexo en A y B tienen el mismo nivel de consumo (pero G puede ser diferente):
  - $H_0 : \delta_1 = \delta_2$ frente a $H_a : \delta_1 \neq \delta_2$
  - Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta \underbrace{(T_{1t} + T_{2t})}_{1-T_{3t}} + u_t$$

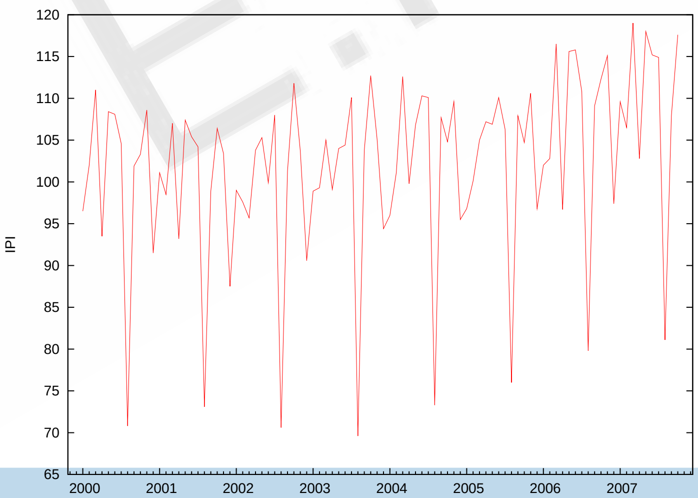
- Hipótesis: Residentes del mismo sexo en B y G tienen el mismo nivel de consumo (pero A puede ser diferente):
  - $H_0 : \delta_2 = 0$ frente a $H_a : \delta_2 \neq 0$
  - Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \delta_1 T_{1t} + u_t$$

4.2 Efecto estacional

Efecto estacional

- Efecto estacional:
-
- Var estacional $\rightsquigarrow$ submuestras $T_1, T_2, \ldots$ según estaciones/meses



Variables Ficticias Estacionales: Definición

- Def. Variable Ficticia Estacional:

$$D_{jt} = \begin{cases} 1, & \text{en } t \in \text{estación } j = 1,2,3,4,\dots; \\ 0, & \text{sino.} \end{cases}$$

- *p.ej.* para datos trimestrales:

date (t)	IPI_t	X_t	D_{1t}	D_{2t}	D_{3t}	D_{4t}
1975.1	·	·	1	0	0	0
1975.2	·	·	0	1	0	0
1975.3	·	·	0	0	1	0
1975.4	·	·	0	0	0	1
1976.1	·	·	1	0	0	0
1976.2	·	·	0	1	0	0
1976.3	·	·	0	0	1	0
1976.4	·	·	0	0	0	1
1977.1	·	·	1	0	0	0
⋮	⋮	⋮	…			
2000.1	·	·	1	0	0	0
2000.2	·	·	0	1	0	0
2000.3	·	·	0	0	1	0
2000.4	·	·	0	0	0	1
2001.1	·	·	1	0	0	0
…	…	…	…			

Variables Ficticias Estacionales: Definición (2)

- Modelo a estimar:

$$IPI_t = \beta_0 + \beta_1 X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \cancel{\gamma_4} D_{4t} + u_t$$
$$= \beta_0 + \beta_1 X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + u_t$$

- ¿Interpretación de los parámetros γ?
- <1>¿Qué ocurre si los datos son observaciones mensuales (como de hecho en el ejemplo IPI)?

fecha (t)	IPI_t	X_t	D_{1t}	D_{2t}	D_{3t}	D_{4t}
1975.ene	·	·	1	0	0	0
1975.feb	·	·	1	0	0	0
1975.mar	·	·	1	0	0	0
1975.abr	·	·	0	1	0	0
1975.may	·	·	0	1	0	0
1975.jun	·	·	0	1	0	0
1975.jul	·	·	0	0	1	0
1975.ago	·	·	0	0	1	0
1975.sep	·	·	0	0	1	0
1975.oct	·	·	0	0	0	1
1975.nov	·	·	0	0	0	1
1975.dic	·	·	0	0	0	1
1976.ene	·	·	1	0	0	0
1976.feb	·	·	1	0	0	0
1976.mar	·	·	1	0	0	0
…	…	…	…			

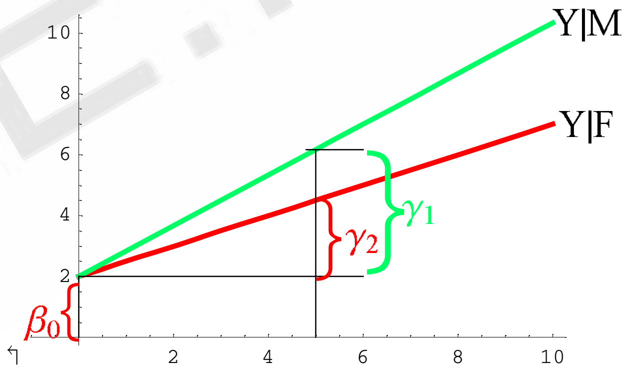
ó

$D_{1t}\dots$											$\dots D_{12t}$
1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
…											

4.3 Interacción entre VFs y Vars cuantitativas

Interacción entre VFs y Vars cuantitativas

En vez de diferentes *interceptos*, necesitamos diferentes pendientes para cada categoría:



es decir, diferente respuesta "Y" para el mismo "X"

Trampa Var Ficticias: interacción

- <1>Matriz X:

cte	R	$R\times S_1$	$R\times S_2$
1	R_1	$R_1\times 1$	$R_1\times 0$
1	R_2	$R_2\times 0$	$R_2\times 1$
1	R_3	$R_3\times 0$	$R_3\times 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	R_T	$R_T\times 1$	$R_T\times 0$

$\Rightarrow$

cte	R	RS_1	RS_2
1	R_1	R_1	0
1	R_2	0	R_2
1	R_3	0	R_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	R_T	R_T	0

- Modelo:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 R_t S_{1t} + \gamma_2 R_t S_{2t} + u_t$$

- Problema (trampa VF): X es matriz $T\times 4$, pero

$$RS_1 + RS_2 = R \Rightarrow \mathrm{rk}(X) = 3 < 4 \quad \text{(MC exacta)}$$

- Solución General: eliminar UNA de las col's que causan el problema: R ó RS_1 ó RS_2 .

Solución: Interacción sin categoría

- SOLUCIÓN MÁS HABITUAL:

eliminar última categoría de la VF: F (RS_2):

- Modelo a estimar:

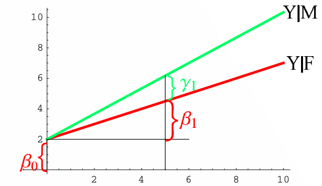
$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 R_t S_{1t} + u_t$$

no $R\,S_2$

- <2>Modelos Submuestras:

$$\mathsf{E}(Y_t|S=F) = \beta_0 + \beta_1 R_t$$
$$\mathsf{E}(Y_t|S=M) = \beta_0 + \underbrace{(\beta_1+\gamma_1)}_{\beta_1^*}R_t$$

$\Rightarrow$



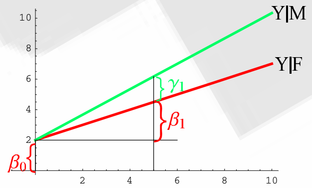
- Interpretación de los coeficientes:

$$\mathsf{E}(Y_t|R_t=0) = \beta_0$$
$$\frac{\partial \mathsf{E}(Y_t|S=F)}{\partial R_t} = \beta_1$$
$$\frac{\partial \mathsf{E}(Y_t|S=M)}{\partial R_t} = \beta_1+\gamma_1$$

Interpretación de los coe cientes

no $R\,S_2$

$$\mathsf{E}(Y_t|R_t=0) = \beta_0$$
$$\frac{\partial \mathsf{E}(Y_t|S=F)}{\partial R_t} = \beta_1$$
$$\frac{\partial \mathsf{E}(Y_t|S=M)}{\partial R_t} = \beta_1+\gamma_1$$



- <3>es decir,

$\beta_0=$ consumo esperado si $R_t = 0$.

$\beta_1=$ Δ consumo Mujeres si $\Delta R_t = 1$ (*c.p.*).

$\gamma_1=$ dif Δ consumo para Hombres (frente a base $=$ Mujer).

Recuérdese: Este caso significa diferentes pendientes para cada categoría.
Recuérdese una vez más: Eliminar una (combinación de) categoría(s)
$\rightsquigarrow$ transformarla en base de referencia.

Contrastes Habituales con Interacción

Hipótesis: M y F Consumo igual
o la variable Sexo no afecta Consumo:

- Modelo sin Restringir:

$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 R_t S_{1t} + u_t$$

- Hipótesis: $H_0 : \gamma_1 = 0$ frente a $H_a : \gamma_1 \neq 0$
- Modelo Restringido:

$$Y_t = \beta_0 + \beta_1 R_t + u_t$$

- Usar el habitual Estadístico t (o Estadístico F basado en SCR)

Fin

FIN