INTRODUCTORY ECONOMETRICS

Lesson 4

Dr Javier Fernández

etpfemaj@ehu.es

Dpt. of Econometrics & Statistics

UPV—EHU

4.1 Dummy Variables.
Definition and use in the GLRM.

Dummy Variables: Definition

- Qualitative explanatory var $\rightsquigarrow$ subsamples $T_1, T_2, \ldots$
  according to category or characteristics
- examples:
  - pure qualitative vars:
    - individual diffs: sex, race, civil state, etc.
    - time diffs: season, war/peace, etc.
    - spatial diffs: countries, A.C.'s, urban/rural, etc.
  - quantitative vars by sections: income, age, etc.
- Recall: we cannot use qualitative vars...
  then substitute by **dummy vars**...
- Def. of Dummy Variable:

$$D_{jt} = \begin{cases} 1, & \text{if } t \in \text{category } j\,; \\ 0, & \text{otherwise.} \end{cases}$$

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

1 QV with 2 categories

Consumption = f([ctnt], income, sex)

$$\begin{array}{cccc} \downarrow & \downarrow & \downarrow & \swarrow \searrow \\ Y_t & [1] & R_t & M \quad F \end{array}$$

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

Sample:

t	Y	ctnt	R	S
1	Y_1	1	R_1	M
2	Y_2	1	R_2	F
3	Y_3	1	R_3	F
⋮	⋮	⋮	⋮	⋮
T	Y_T	1	R_T	M
		X ?		

$\Rightarrow$

t	Y	ctnt	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
⋮	⋮	⋮	⋮	⋮	
T	Y_T	1	R_T	1	0
			X		

In principle: substitute QV by

as many DVs as categories we have.

Dummy Var Trap: 1 qualitative var

- Model: $Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \gamma_2 S_{2t} + u_t$
- Problem (Dummy Variable trap):
  X is a $(T \times 4)$ matrix, but

  $S_1 + S_2 = [1]$ (exact l.c.) $\Rightarrow \mathrm{rk}(X) = 3 < 4$ (*i.e.* perfect MC)
- $\Rightarrow \det(X'X) = 0$
  $\Rightarrow (X'X)^{-1}$ doesn't exist!! and
  $\widehat{\beta}$ cannot be calculated!!
- General Solution: eliminate ONE of the col's causing the problem: [1] or S_1 or S_2 .
- (POSSIBLE Solution: eliminate intercept... but...

Solution: DV without a category

MOST USUAL SOLUTION: eliminate category: *e.g.* F (S_2):

- Model to estimate:

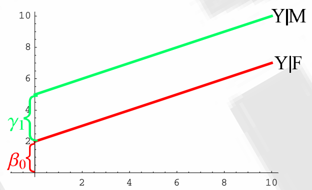
$$Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \cancel{\gamma_2 S_{2t}} + u_t$$
$$= \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + u_t$$

- Subsample Models:

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

$\Rightarrow$

without category F

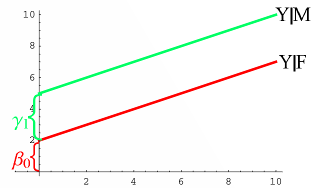


- Coefficient interpretation:

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

Coefficient Interpretation

without category F



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

- that is,

β_0 = expected consumption Women (base) if $R_t = 0$.

γ_1 = diff expected consumption of Men (vs. base = Women).

$\beta_1 = \Delta$ consumption if $\Delta R_t = 1$ (*c.p.*).

Recall: This case just means different intercepts for each category.
Note: Eliminating a category $\rightsquigarrow$ transforms it into reference base.

Usual Tests with 1 QV

Hypothesis: qualitative variable (Sex) not significant (it doesn't affect Consumption) *i.e.* M and F same Consumption:

- Unrestricted Model

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

- Hypothesis: $H_0 : \gamma = 0$ vs. $H_a : \gamma \neq 0$
- Restricted Model:

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

- Use usual t Statistic (or F Statistic based on RSS)

1 QV with 2 cats + 1 QV with 3 cats

Consumption $= f(\text{[ctnt]}, \text{income}, \text{sex}, \text{territory CAV})$

Y_t, $[1]$, R_t, $M \;\; F$, $A \;\; B \;\; G$

$$S_{1t} = \mathscr{I}(t \in M) \qquad T_{1t} = \mathscr{I}(t \in A)$$
$$S_{2t} = \mathscr{I}(t \in F) \qquad T_{2t} = \mathscr{I}(t \in B)$$
$$T_{3t} = \mathscr{I}(t \in G)$$

Sample:

t	Y	ctnt	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
⋮	⋮	⋮	⋮	⋮	⋮
T	Y_T	1	R_T	M	A

X ?

$\Rightarrow$

t	Y	ctnt	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
⋮	⋮	⋮	⋮	⋮			⋮	
T	Y_T	1	R_T	1	0	1	0	0

X

Recall: In principle, substitute qualitative var by as many Dummy vars as categories we have.

Dummy Var Trap: 2 qualitative vars

- Model: $Y_t = \beta_0 + \beta_1 R_t + \gamma_1 S_{1t} + \gamma_2 S_{2t} + \delta_1 T_{1t} + \delta_2 T_{2t} + \delta_3 T_{3t} + u_t$
- Problem (DV trap):
  X is a $(T \times 7)$ matrix, but

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

(2 exact l.c.) $\Rightarrow \mathrm{rk}(X) = 5 < 7$ (*i.e.* perfect MC)

- $\Rightarrow \det(X'X) = 0$
  $\Rightarrow (X'X)^{-1}$ doesn't exist!! and
  $\widehat{\beta}$ cannot be calculated!!
- General Solution: eliminate ONE of the col's causing the problem: [1] or (S_1 or S_2) or (T_1 or T_2 or T_3).

Solution: DV without combination of categories

MOST USUAL SOLUTION:
eliminate last category of each DV: S_2 and T_3:

- Model to estimate:

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

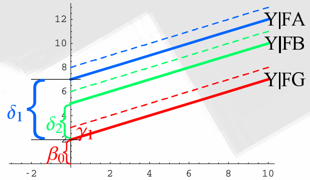
- Subsample Models:

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

Coefficient Interpretation

without categories F nor G

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



- that is,

$\beta_0 =$ expected consumption Women G (base) if $R_t = 0$.

$\gamma_1 =$ diff. expected consumption Men vs. Women .

$\delta_1 =$ diff. expected consumption A vs. G.

$\delta_2 =$ diff. expected consumption B vs. G.

$\beta_1 = \Delta$ consumption if $\Delta R_t = 1$ (*c.p.*).

Recall: This case just means different intercepts for each category. Recall: Eliminating a (combination of) category(ies)
$\rightsquigarrow$ transforms it into reference base.

Usual Tests with 2 QVs

Hypothesis: Variable Sex doesn't affect Consumption
(but place of residence might do)

- Unrestricted Model:

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

($\gamma_1 =$ diff. exp. C of M vs. F)

- Hypothesis: $H_0 : \gamma_1 = 0$ vs. $H_a : \gamma_1 \neq 0$
- Restricted Model:

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

- Use usual t Statistic (or F Statistic based on RSS)

Other usual Tests with 2 QVs

- Unrestricted Model (without S_2 nor T_3):

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

  - Recall: γ_1 is diff. expected C of M vs. F (base)
    δ_1 and δ_2 are diff. exp. C of A and B vs. G (base)
- Hypothesis: Same Consumption overall (independently of Sex and Residence):
  - $H_0 : \gamma_1 = \delta_1 = \delta_2 = 0$
  - Restricted Model:

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

- Hypothesis: Place of Residence doesn't affect Consumption (but M vs. F might do):
  - $H_0 : \delta_1 = \delta_2 = 0$
  - Restricted Model:

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

Other usual Tests with 2 QVs

- Unrestricted Model (without S_2 nor T_3):

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

  - Recall: δ_1 and δ_2 are diff. expected C of A and B vs. G (base)
- Hypothesis: Residents of same sex in A and B have same consumption level (but G might be different):
  - $H_0 : \delta_1 = \delta_2$ vs. $H_a : \delta_1 \neq \delta_2$
  - Restricted Model:

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

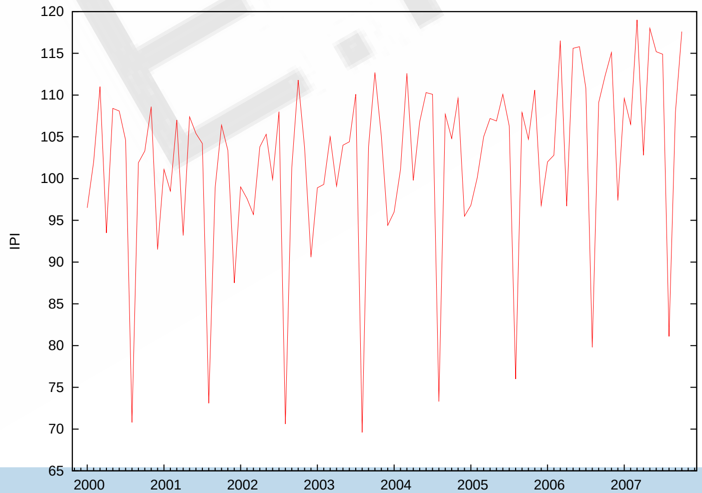
- Hypothesis: Residents of same sex in B and G have same consumption level (but A might be different):
  - $H_0 : \delta_2 = 0$ vs. $H_a : \delta_2 \neq 0$
  - Restricted Model:

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

4.2 Seasonal effects

Seasonal effect

- Seasonal effect:
-
- Seasonal var $\rightsquigarrow$ subsamples $T_1, T_2, \ldots$ according to seasons/months



Seasonal Dummy Variables: Definition

- Def. of Seasonal Dummy Variable:

$$D_{jt} = \begin{cases} 1, & \text{if } t \in \text{season } j = 1,2,3,4,\ldots; \\ 0, & \text{otherwise.} \end{cases}$$

- *e.g.* for quarterly data:

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
…	…	…	…			

Seasonal Dummy Variables: Definition (2)

- Model to estimate:

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

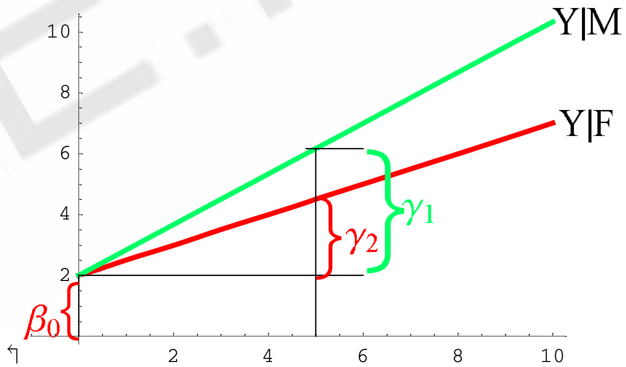
- interpretation of γ parameters?
- What if data are monthly observations (as in the IPI example actually)?

date (t)	IPI_t	X_t	D_{1t}	D_{2t}	D_{3t}	D_{4t}		D_{1t} …											… D_{12t}
1975.jan	·	·	1	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0
1975.feb	·	·	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0	0
1975.mar	·	·	1	0	0	0		0	0	1	0	0	0	0	0	0	0	0	0
1975.apr	·	·	0	1	0	0		0	0	0	1	0	0	0	0	0	0	0	0
1975.may	·	·	0	1	0	0		0	0	0	0	1	0	0	0	0	0	0	0
1975.jun	·	·	0	1	0	0		0	0	0	0	0	1	0	0	0	0	0	0
1975.jul	·	·	0	0	1	0		0	0	0	0	0	0	1	0	0	0	0	0
1975.ago	·	·	0	0	1	0	or	0	0	0	0	0	0	0	1	0	0	0	0
1975.sep	·	·	0	0	1	0		0	0	0	0	0	0	0	0	1	0	0	0
1975.oct	·	·	0	0	0	1		0	0	0	0	0	0	0	0	0	1	0	0
1975.nov	·	·	0	0	0	1		0	0	0	0	0	0	0	0	0	0	1	0
1975.dec	·	·	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	1
1976.jan	·	·	1	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0
1976.feb	·	·	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0	0
1976.mar	·	·	1	0	0	0		0	0	1	0	0	0	0	0	0	0	0	0
…	…	…	…					…											

4.3 Interaction between DVs and quantitative Vars

Interaction between DVs and quantitative Vars

Instead of different *intercepts*, we require different slopes for each category:



that is, different response "Y" for same "X"

Dummy Var Trap: interaction

- Matrix X:

$$\begin{array}{|c|c|c|c|}\hline \text{ctnt} & R & R\times S_1 & R\times S_2 \\ \hline 1 & R_1 & R_1\times 1 & R_1\times 0 \\ 1 & R_2 & R_2\times 0 & R_2\times 1 \\ 1 & R_3 & R_3\times 0 & R_3\times 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & R_T & R_T\times 1 & R_T\times 0 \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|c|c|}\hline \text{ctnt} & R & RS_1 & RS_2 \\ \hline 1 & R_1 & R_1 & 0 \\ 1 & R_2 & 0 & R_2 \\ 1 & R_3 & 0 & R_3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & R_T & R_T & 0 \\ \hline \end{array}$$

- Model:

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

- Problem (DV trap): X is $T \times 4$, but

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

- General Solution: eliminate ONE of the col's causing the problem: R or RS_1 or RS_2 .

Solution: Interaction without a category

- MOST USUAL SOLUTION:

eliminate last category of the DV: F (RS_2):

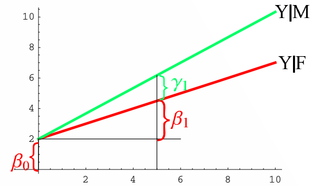
- Model to estimate:

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

- Subsample Models:

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

no RS_2



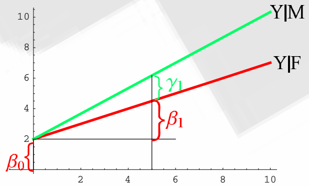
- Coefficient interpretation:

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

Coefficient Interpretation

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

no RS_2



- that is,

β_0= expected consumption if $R_t = 0$.

β_1= Δ consumption Women if $\Delta R_t = 1$ (*c.p.*).

γ_1= diff Δ consumption for Men (vs. base = Female).

Recall: This case means different slopes for each category. Recall: again eliminating a category $\rightsquigarrow$

transforms it into reference base.

Usual Tests with Interaction

Hypothesis: M and F equal Consumption
or variable Sex doesn't affect Consumption:

- Unrestricted Model:

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

- Hypothesis: $H_0 : \gamma_1 = 0$ vs. $H_a : \gamma_1 \neq 0$
- Restricted Model:

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

- Use usual t Statistic (or F Statistic based on RSS)

THE END