

Dummy Variables

Dummy Variables are explanatory variables that measure qualitative characteristics. They take values of 1 or 0

$$D_i = \begin{cases} 1 & \text{if Event happens or possesses a characteristic.} \\ 0 & \text{otherwise} \end{cases}$$

Used for gender, race, geographic regions etc.

They allow us to construct models in which some or all of the model parameters change for some obs in sample.

Can allow intercept to change, slopes, or both

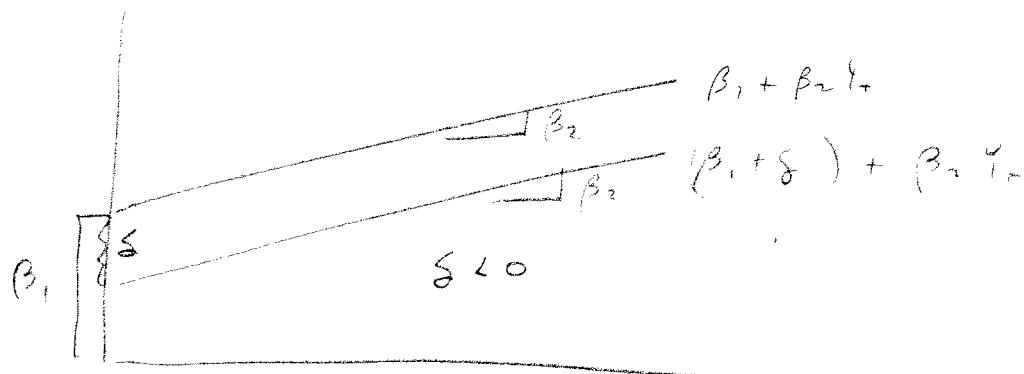
$$C_t = \beta_1 + \beta_2 T_t + \epsilon_t \quad t = 1929 - 1970$$

Suppose we define a dummy for war years

$$D_t = \begin{cases} 1 & \text{if } t = 1941 - 1945 \\ 0 & \text{otherwise} \end{cases}$$

$$C_t = \beta_1 + \gamma D_t + \beta_2 T_t + \epsilon_t$$

$$E(C_t) = \begin{cases} (\beta_1 + \gamma) + \beta_2 T_t & t = 1941 - 1945 \\ \beta_1 + \beta_2 T_t & t = 1929 - 1940 \\ & 1941 - 1970 \end{cases}$$



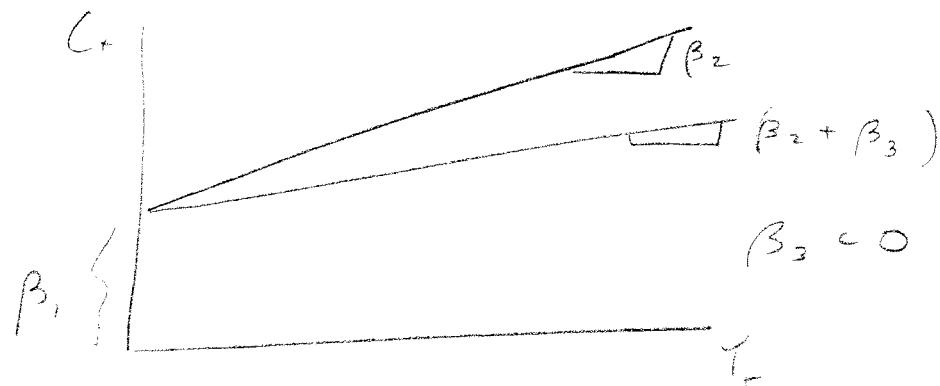
γ results in a parallel shift of the consumption downward during the WWII.

Slope Dummy

We could also allow the slope to be different during other

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 D_t Y_t + e_t$$

$$E(C_t) = \begin{cases} \beta_1 + (\beta_2 + \beta_3) Y_t & t = 1941 - 1946 \\ \beta_1 + \beta_2 Y_t & 1929 - 1940 \\ & 1947 - 1970 \end{cases}$$



Or, you would do both.

$D_t Y_t$ is called an interaction variable.

| | $D_t Y_t$ | $D_t Y_t$ |
|---|-----------|-----------|
| 1 | 29 | 29 |
| 0 | 37 | 0 |
| 1 | 21 | 61 |
| 0 | 19 | 0 |

Multiple Categories

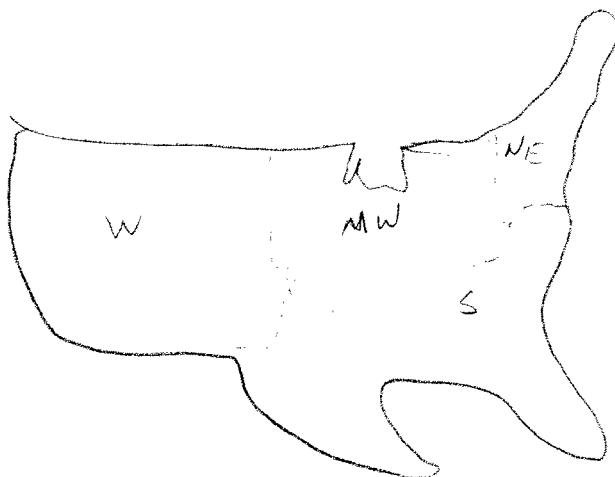
4 Regions requires 3 dummy

Suppose we have regional demands

$$D_{NE} = \begin{cases} 1 & \text{northeast} \\ 0 & \text{other} \end{cases}$$

$$D_S = \begin{cases} 1 & \text{south} \\ 0 & \text{other} \end{cases}$$

$$D_W = \begin{cases} 1 & \text{west} \\ 0 & \text{other} \end{cases}$$



$$C_F = \beta_1 + \beta_2 D_{NE} + \beta_3 D_S + \beta_4 D_W + \beta_5 S + \beta_6 W$$

$$/ (\beta_1 + \beta_2) + \beta_6 T_c \quad NE$$

$$E(C_F) = / (\beta_1 + \beta_3) + \beta_6 T_c \quad S$$

$$\beta_1 + \beta_4 + \beta_6 T_c \quad W$$

$$\beta_1 + \beta_5 T_c \quad MW.$$

Systematically varying Parameter Models

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + \epsilon_t$$

suppose you thought the coefficient β_2 ($\partial Y / \partial X_2$) depended on whether some event happens

$$D_t = \begin{cases} 1 & \text{if event occurs} \\ 0 & \text{otherwise.} \end{cases}$$

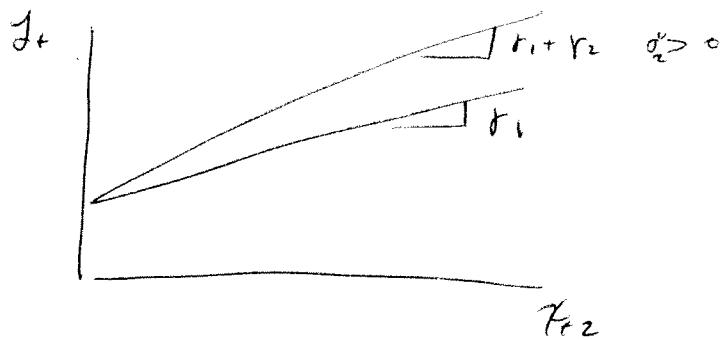
$$\therefore \beta_{t2} = \gamma_1 + \gamma_2 D_t$$

$$\text{so } \beta_{t2} = \begin{cases} \gamma_1 + \gamma_2 & \text{if event occurs} \\ \gamma_1 & \text{otherwise.} \end{cases}$$

Subst in to model

$$Y_t = \beta_1 + (\gamma_1 + \gamma_2 D_t) X_{t2} + \beta_3 X_{t3} + \epsilon_t$$

$$= \beta_1 + \gamma_1 X_{t2} + \gamma_2 D_t X_{t2} + \beta_3 X_{t3} + \epsilon_t$$



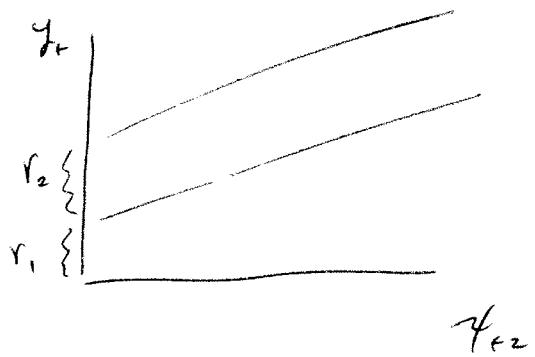
r_2 measures difference in slope between curves by event D_t .

The usual Dummmy variable model is created by forming an auxiliary model for the intercept

$$\beta_{1t} = r_1 + r_2 D_t$$

$$\text{If } D_t = 1 \quad \beta_1 = r_1 + r_2$$

$$\text{If } D_t = 0 \quad \beta_1 = r_1$$



It is possible to make intercept or slope functions of continuous variable

$$y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + \epsilon_t$$

$$\beta_2 = r_1 + r_2 X_{t4}$$

$$y_t = \beta_1 + r_1 X_{t2} + r_2 X_{t2} X_{t4} + \beta_3 X_{t3} + \epsilon_t$$

$$\frac{\partial y}{\partial X_{t2}}$$

$$\frac{\partial y_t}{\partial X_{t2}} = r_1 + r_2 X_{t4}$$

$$\frac{\partial y}{\partial X_{t4}}$$

$$\frac{\partial y_t}{\partial X_{t4}} = r_2 X_{t2}$$

The term $\gamma_{12} \gamma_{24}$ is called an "interaction effect" and its interpretation is clear given the aux model

$$\beta_{12} = \gamma_1 + \gamma_2 \gamma_4$$

To test for the existence of interaction

$$H_0: \gamma_2 = 0$$

$$H_A: \gamma_2 \neq 0$$

Example:

$$\ln(\text{wage}_t) = \beta_1 + \beta_2 S_t + \beta_3 (X_{PR_t}) + \beta_4 (X_{PR_t^2}) + \epsilon_t$$

The effect of schooling on (ln(wage)) depends on age & gender.

$$D_{gt} = \begin{cases} 1 & \text{Female} \\ 0 & \text{male} \end{cases}$$

$$\beta_{12} = \gamma_1 + \gamma_2 AGE_t + \gamma_3 D_{gt}$$

$$\beta_1 + \gamma_1 S_t + \gamma_2 AGE_t \cdot S_t + \gamma_3 S_t \cdot D_t + \beta_3 (X_{PR_t}) + \beta_4 (X_{PR_t^2})$$

Does AGE affect returns to schooling?

$$H_0: \gamma_2 = 0$$

$$H_A: \gamma_2 \neq 0$$

Does gender affect return to schooling?

$$H_0: \gamma_3 = 0$$

$$\text{Price} = \beta_1 + \beta_2 \text{SQFT} + \beta_3 \text{lot} + \beta_4 \text{AGE} \\ + s_1 \text{Pool} + s_2 \text{Hood} + c.$$

Also

$$\beta_2 = r_1 + r_2 \text{Hood.}$$

$$\beta_1 + r_1 \text{SQFT} + r_2 (\text{SQFT} - \text{Hood}) + \beta_3 \text{lot} + \beta_4 \text{AGE} \\ + s_1 \text{Pool} + s_2 \text{Hood.}$$

$$\frac{\partial E(\text{Price})}{\partial \text{Hood}} = r_1 + r_2 \text{Hood} = \begin{cases} \beta_2 r_1 + r_2 \text{Hood} \\ r_1 \text{Hood} \end{cases}$$

per house large size.

Test ~~Neig has no effect 1. Neigh has no effect for 2000 sqft house.~~
~~Price for Movie in Hood 1 = Hood 0~~

$$\frac{\partial E(\text{Price})}{\partial \text{Hood}} = r_2 \text{SQFT} + s_2$$

$$H_0: r_2 \text{SQFT} + s_2 = 0$$

$$H_A: r_2 \text{SQFT} + s_2 \neq 0$$

$$H_0: r_2 = s_2 = 0$$

Certainly true if $r_2 = s_2 = 0$

$$H_A: r_2 \neq 0 \text{ or } s_2 \neq 0$$

what if $s_2 < 0$ $r_2 > 0$?

$$H_0: r_2 \frac{2000}{\text{SQFT}} + s_2 = 0$$

would evaluate so at $\overline{\text{SQFT}}$.

$$H_A: r_2 \frac{2000}{\text{SQFT}} + s_2 \neq 0$$

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$$

$$\text{Lwage}_i = \beta_1 + \beta_2 \text{edule} + \beta_3 \text{exper} + u_i$$

(1) Starting wage for female is lower than that of male.
Returns to ed and exper same.

$$\beta_1 = r_1 + r_2 D_F$$

$$D_F = \begin{cases} 1 & \text{if Female} \\ 0 & \text{otherwise} \end{cases}$$

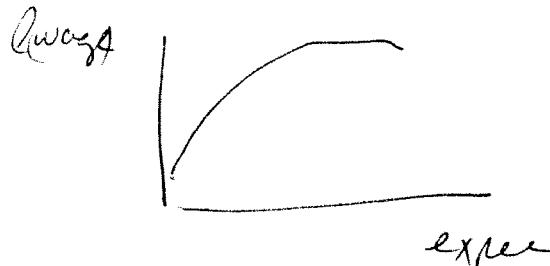
$$= r_1 + r_2 D_F + \beta_2 \text{edule} + \beta_3 \text{exper} + u_i$$

\Rightarrow std. dummy var model.

(2) Return to another year of experience
Varies according to your existing level of experience

$$\beta_3 = s_1 + s_2 \text{exper.}$$

$$= \beta_1 + \beta_2 \text{edule} + s_1 \text{exper} + s_2 \text{exper}^2 + u_i$$



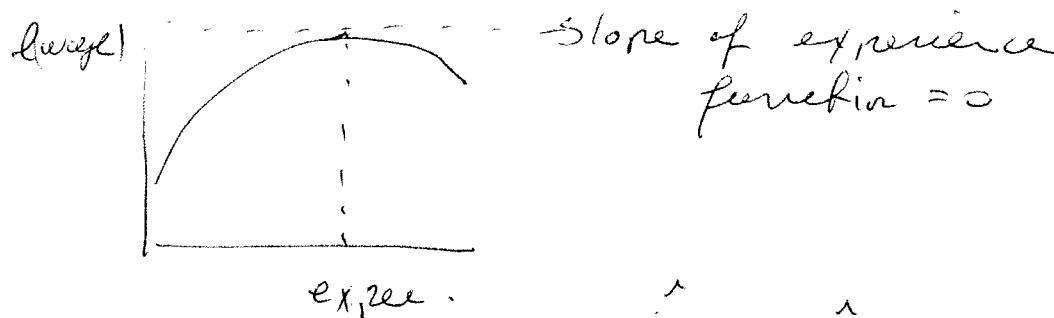
$$s_1 > 0, s_2 < 0$$

AT what point does function peak.

When Does another year of expre \Rightarrow no change

$$\frac{\partial E(y)}{\partial \text{exper}} = S_1 + 2S_2 \text{exper} = 0$$

$$+ (-)$$



$$S_1 = -2 S_2 \text{Exper}$$

$$\text{Exper} = -\frac{S_1}{2S_2} > 0$$

(3) β_2 Return to schooling depends on gender and a persons age.

$$\beta_{i2} = \theta_1 + \theta_2 \text{AGE}_i + \theta_3 D_F$$

$$\beta_1 + \beta_{2,1} (\theta_1 + \theta_2 \text{AGE}_i + \theta_3 D_F) \text{Exper}_i + \beta_3 \text{Exper}_i + ll.$$

$$+ \theta_1 E_{deee}_i + \theta_2 AGE_i E_{deee}_i + \theta_3 D_F E_{deee}_i + \beta_3 Exper_i + ll$$

To Test the hypothesis that this is untrue

$$H_0: \theta_2 = \theta_3 = 0$$

$$H_A \text{ not } H_0$$

use F-test.