Dummy Variables

$$C_i = \begin{cases} 1, \text{ if bad} \\ 2, \text{ if medium} \\ 3, \text{ if good} \end{cases}$$

$$y_i = \beta_0 + \beta_1 \cdot C_i + \beta_2 \cdot p_i + \varepsilon_i$$

$C_i = 1$ $\qquad \beta_0 + \beta_1$

$C_i = 2$ $\qquad \beta_0 + 2\beta_1$

$C_i = 3$ $\qquad \beta_0 + 3\beta_1$

$$y_i = \beta_0 + \beta_1 \cdot d_b + \beta_2 \cdot d_g + \beta_3 \cdot p_i + \varepsilon_i \qquad (+\cancel{\beta_4} \cdot d_m)$$

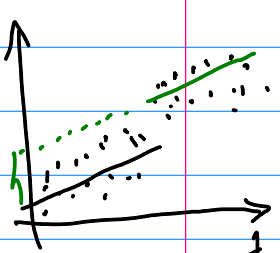
$$d_b = \begin{cases} 0, & \text{else} \\ 1, & C_i = 1 \end{cases}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & p_1 \\ \vdots & \vdots & \vdots & \vdots & \\ & 1 & 0 & \vdots & \vdots \\ & 0 \;+ & 1 \;+ & \vdots & \vdots \\ & \vdots & \vdots & & \\ & \vdots & 1 & 0 & \\ & \vdots & 0 & 1 & \\ & & \vdots & \vdots & \\ 1 & 0 & 0 & 1 & p_n \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1} X'y$$

→ dummy variable trap

Question 1: $\cdot\; Y_i = \beta_0 + \beta_1 \cdot P_i + \varepsilon_i$



a)

$$D_i = \begin{cases} 0, & \text{before crisis} \\ 1, & \text{after crisis} \end{cases}$$

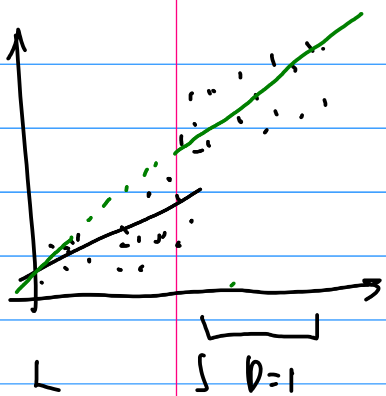
$$Y_i = \beta_0 + \beta_1 \cdot P_i + \beta_2 \cdot D_i + \varepsilon_i$$

(Ref. model) before crisis: $Y_i = \beta_0 + \beta_1 P_i + \varepsilon_i$

after crisis: $Y_i = (\beta_0 + \beta_2) + \beta_1 \cdot P_i + \varepsilon_i$

b) change in slope

$$Y_i = \beta_0 + \beta_1 \cdot P_i + \beta_2 \cdot P_i \cdot D_i + \varepsilon_i$$



Ref. mod. B.C. : $Y_i = \beta_0 + \beta_1 \cdot P_i + \varepsilon_i$

A.C. : $Y_i = \beta_0 + (\beta_1 + \beta_2) P_i + \varepsilon_i$

c) $$Y_i = \beta_0 + \beta_1 \cdot P_i + \beta_2 \cdot D_i + \beta_3 \cdot P_i D_i + \varepsilon_i$$

Ref. mod. B.C. : $Y_i = \beta_0 + \beta_1 \cdot P_i + \varepsilon_i$

A.C. : $Y_i = \beta_0 + \beta_2 + (\beta_1 + \beta_3) P_i + \varepsilon_i$

F - test : $\beta_2 = \beta_3 = 0$

Chow test : (for structural break for T.S.)

Pooled reg. $y_i = \beta_1 + \beta_2 x_{1,i} + \ldots + \beta_k \cdot x_{k-1,i} + \varepsilon_i$ $\quad n = n_A + n_B$

Subsample A $y_i = \beta_1^A + \beta_2^A x_{1,i} + \ldots + \beta_k^A \cdot x_{k-1,i} + \varepsilon_i$ $\quad n_A \, (D_i = 0)$

Subsample B $y_i = \beta_1^B + \beta_2^B x_{1,i} + \ldots + \beta_k^B \cdot x_{k-1,i} + \varepsilon_i$ $\quad n_B \, (D_i = 1)$

H_0: $\beta_1^A = \beta_1^B, \ldots, \beta_k^A = \beta_k^B$

H_a: at least one rest. is not held

$$F = \frac{(RSS_P - RSS_A - RSS_B)/k}{(RSS_A + RSS_B)/(n - 2 \cdot k)} \sim F(k, n-2k)$$

$$F = \frac{\text{improvement in fit / extra d.o.f.}}{\text{remaining RSS / remaining d.o.f.}}$$

Problem 2. (ICEF exam) A student decided to investigate the market of private mathematics teachers in Moscow, with particular interest to those who can teach in English. He took a random sample of 30 profiles of teachers who provide private teaching in math (taken from population of 300 profiles registered in certain internet site) and run some regressions trying to find factors influencing the prices of teaching ($PRICE_i$ - price of a standard two-hour lesson in thousands of roubles, $DIST_i$ - distance in the number of metro stations from the center of Moscow to the teacher's place, $HOME_i$ - dummy variable indicating visit of the tutor to the client, ENG_i - dummy variable indicating ability to teach the subject in English):

R:

$$\widehat{PRICE}_i = \underset{(0.49)}{6.59} - \underset{(0.06)}{0.16}DIST_i \qquad R^2 = 0.185 \qquad (1)$$

$$\widehat{PRICE}_i = \underset{(0.40)}{\overset{\beta_0}{4.51}} + \underset{(0.58)}{\overset{\beta_1}{2.54}}HOME_i \qquad R^2 = 0.40 \qquad (2)$$

β_0 home = 0

$\beta_0 + \beta_1$ home = 1

$$\widehat{PRICE}_i = \underset{(0.64)}{5.13} - \underset{(0.06)}{0.08}DIST_i + \underset{(0.95)}{1.95}HOME_i + \underset{(0.07)}{0.07}DIST*HOME_i \qquad R^2 = 0.437$$

(3)

$$\widehat{PRICE}_i = \underset{(0.61)}{4.52} - \underset{(0.06)}{0.08}DIST_i + \underset{(0.75)}{2.18}HOME_i + \underset{(0.76)}{1.58}ENG_i - \underset{(1.09)}{0.39}HOME*ENG_i \qquad R^2 = 0.553$$

(4)

a) $4,51 + 2,54 = 7,05$

b) home = 1 $\qquad \hat{P}_i = 7,08 - 0,01 D_i + \hat{e}_i$

c) home = 0 eng = 0

d) is model (2) sign.:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0,437/3}{(1-0,437)/(30-4)} = 6,73$$

$$F_{crit} = 4,6$$

is factor "dist" sig.:

F - test $\quad H_0: \beta_1 = \beta_3 = 0$

$$F = \frac{(R^2_{ur} - R^2_R)/q}{(1-R^2_{UR})/(n-k)} = \frac{0,037/2}{(1-0,437)/26} = 5,8$$

$$F_{crit} = 5,6$$

$$H_0: \beta_2 = \beta_3 = 0$$

$$F = \dots$$

d) for mod. (3):

$$F = \frac{(RSS_p - RSS_A - RSS_B)/k}{(RSS_A + RSS_B)/(n - 2 \cdot k)} \sim F(k, n - 2k)$$

for mod. (4):

R

$$F = \frac{(RSS_p - RSS_{he} - RSS_{He} - RSS_{hE} - RSS_{HE})/2 \cdot 3}{(RSS_{he} + RSS_{He} + RSS_{hE} + RSS_{HE})/(30 - 4 \cdot 2)}$$

Alternative: (equivalent to Chow test)

$$P_i = \beta_0 + \beta_1 \cdot D_i + \beta_2 H_i + \beta_3 \cdot E_i +$$

$$\beta_4 \cdot H_i \cdot E_i + \beta_5 \cdot D_i \cdot H_i + \beta_6 D_i \cdot E_i +$$

$$\beta_7 \cdot D_i \cdot H_i \cdot E_i + w_i$$

F_{test}: $\beta_2 = \dots = \beta_7 = 0$

$$F = \frac{(R^2_{UR} - R^2_R)/6}{(1 - R^2_{UR})/(30 - 8)}$$