Introduction to Econometrics
Chapter 5

Ezequiel Uriel Jiménez
University of Valencia

Valencia, September, 2013

5 Multiple regression analysis with qualitative information

5.1 Introduction of qualitative information in econometric models

5.2 A single dummy independent variable

5.3 Multiple categories for an attribute

5.4 Several attributes

5.5 Interactions involving dummy variables

5.6 Testing structural changes

Exercises

5.1 Introduction of qualitative information in econometric models

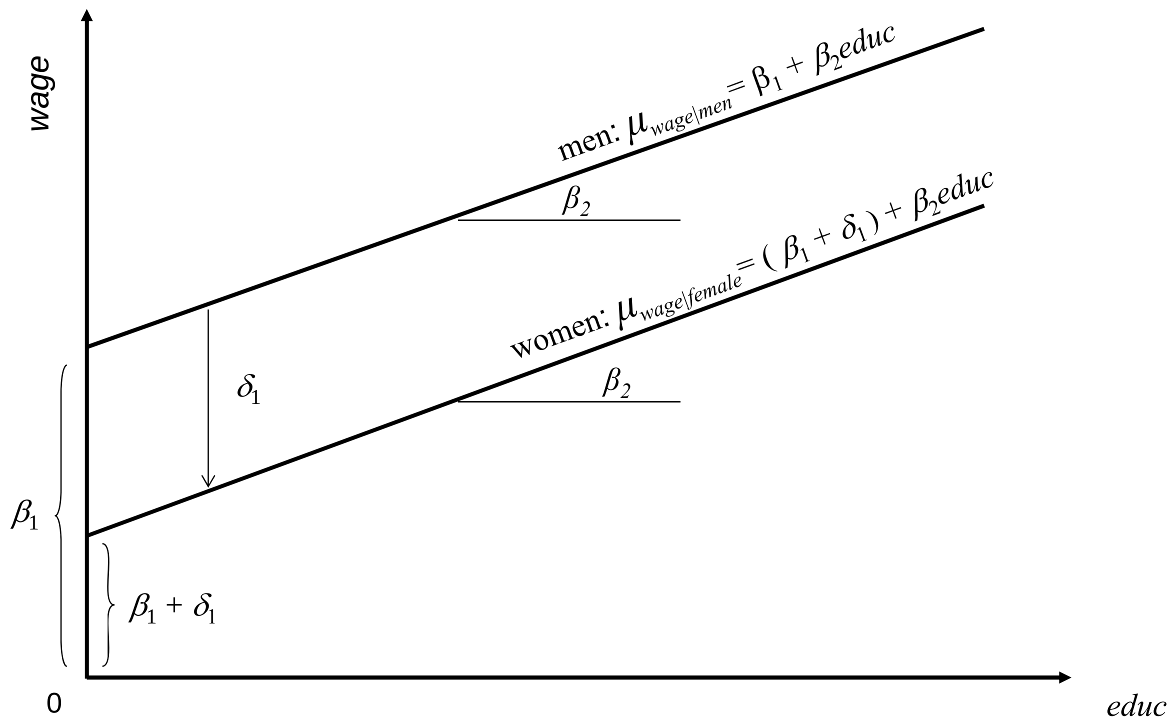

FIGURE 5.1. Same slope, different intercept.

5.2 A single dummy independent variable

EXAMPLE 5.1 Is there wage discrimination against women in Spain? (file wage02sp)

$$\ln(wage) = \beta_1 + \delta_1 female + \beta_2 educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.026)}{1.731} - \underset{(0.022)}{0.307}\ female + \underset{(0.0025)}{0.0548}\, educ$$

$$RSS = 393 \quad R^2 = 0.243 \quad n = 2000$$

$$H_0: \delta_1 = 0$$
$$H_1: \delta_1 < 0$$

$$t = \frac{-0.3070}{0.0216} = -14.26$$

Percentage difference in hourly wage between men and women

$$= 100 \times (e^{0.307} - 1) = 35.9\%$$

5.2 A single dummy independent variable

EXAMPLE 5.2 Analysis of the relation between market capitalization and book value: the role of ibex35 (file bolmad11)

$$\ln(marketcap) = \beta_1 + \delta_1 ibex35 + \beta_2 \ln(bookvalue) + u$$

$$\widehat{\ln(marketcap)} = \underset{(0.243)}{1.784} + \underset{(0.179)}{0.690}\, ibex35 + \underset{(0.037)}{0.675} \ln(bookvalue)$$

$$RSS = 35.672 \quad R^2 = 0.893 \quad n = 92$$

$$H_0 : \delta_1 = 0$$
$$H_1 : \delta_1 > 0$$

$$t = \frac{0.690}{0.179} = 3.85$$

Percentage difference= $100 \times (e^{0.690} - 1) = 99.4\%$

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

$$t = \frac{0.675}{0.037} = 18$$

5.2 A single dummy independent variable

EXAMPLE 5.3 Do people living in urban areas spend more on fish than people living in rural areas? (file demand)

$$\ln(fish) = \beta_1 + \delta_1 urban + \beta_2 \ln(inc) + u$$

$$\widehat{\ln(fish)} = \underset{(0.511)}{-6.375} + \underset{(0.055)}{0.140}\, urban + \underset{(0.070)}{1.313} \ln(inc)$$

$$RSS = 1.131 \quad R^2 = 0.904 \quad n = 40$$

$$H_0 : \delta_1 = 0$$

$$H_1 : \delta_1 > 0$$

$$t = \frac{0.140}{0.055} = 2.55$$

5.3 Multiple categories for an attribute

Dummy variable trap

Example

$$\ln(wage) = \beta_1 + \theta_0 small + \theta_1 medium + \theta_2 large + \beta_2 educ + u$$

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & educ_1 \\ 1 & 1 & 0 & 0 & educ_2 \\ 1 & 0 & 1 & 0 & educ_3 \\ 1 & 0 & 1 & 0 & educ_4 \\ 1 & 0 & 0 & 1 & educ_5 \\ 1 & 0 & 0 & 1 & educ_6 \end{bmatrix}$$

Solutions:

$$\ln(wage) = \beta_1 + \theta_1 medium + \theta_2 large + \beta_2 educ + u$$

$$\ln(wage) = \theta_0 small + \theta_1 medium + \theta_2 large + \beta_2 educ + u$$

5.3 Multiple categories for an attribute

EXAMPLE 5.4 Does firm size influence wage determination? (file wage02sp)

$$\ln(wage) = \beta_1 + \theta_1 medium + \theta_2 large + \beta_2 educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.027)}{1.566} + \underset{(0.025)}{0.281}\, medium + \underset{(0.024)}{0.162}\, large + \underset{(0.0025)}{0.0480}\, educ$$

$$RSS = 406 \quad R^2 = 0.218 \quad n = 2000$$

$$H_0 : \theta_1 = \theta_2 = 0$$
$$H_1 : H_0 \text{ is not true}$$

$$\ln(wage) = \beta_1 + \beta_2 educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.026)}{1.657} + \underset{(0.0026)}{0.0525}\, educ$$

$$RSS = 433 \quad R^2 = 0.166 \quad n = 2000$$

$$F = \frac{\left[RSS_R - RSS_{UR}\right]/q}{RSS_{UR}/(n-k)} = \frac{\left[433-406\right]/2}{406/(2000-4)} = 66.4$$

5.3 Multiple categories for an attribute

EXAMPLE 5.5 In the case of Lydia E. Pinkham, are the time dummy variables introduced significant individually or jointly? (file pinkham)

$$sales_t = \beta_1 + \beta_2 advexp_t + \beta_3\, sales_{t-1} + \beta_4 d1_t + \beta_5 d2_t + \beta_6 d3_t + u_t$$

$$\widehat{sales_t} = \underset{(96.3)}{254.6} + \underset{(0.136)}{0.5345}\, advexp_t + \underset{(0.0814)}{0.6073}\, sales_{t-1} - \underset{(89)}{133.35}\, d1_t + \underset{(67)}{216.84}\, d2_t - \underset{(67)}{202.50}\, d3t$$

$$R^2 = 0.929 \quad n = 53$$

$$\begin{cases} H_0: \theta_i = 0 \\ H_1: \theta_i \neq 0 \end{cases} \qquad i = 1,2,3$$

$$t_{\hat{\theta}_1} = \frac{-133.35}{89} = -1.50 \qquad t_{\hat{\theta}_2} = \frac{216.84}{67} = 3.22 \qquad t_{\hat{\theta}_3} = \frac{-202.50}{67} = -3.02$$

$$\begin{cases} H_0: \theta_1 = \theta_2 = \theta_3 = 0 \\ H_1: H_0 \text{ is not true} \end{cases}$$

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(n-k)} = \frac{(0.9290 - 0.8770)/3}{(1 - 0.9290)/(53-6)} = 11.47$$

5.4 Several attributes

EXAMPLE 5.6 The influence of gender and length of the workday on wage determination (file wage06sp)

$$\ln(wage) = \beta_1 + \delta_1 female + \phi_1 partime + \beta_2 educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.026)}{2.006} - \underset{(0.021)}{0.233}\, female - \underset{(0.027)}{0.087}\, partime + \underset{(0.0023)}{0.0531} educ$$

$$RSS = 365 \quad R^2 = 0.235 \quad n = 2000$$

EXAMPLE 5.7 Trying to explain the absence from work in the company Buenosaires (file absent)

$$absent = \beta_1 + \delta_1 bluecoll + \phi_1 male + \beta_2 age + \beta_3 tenure + \beta_4 wage + u$$

$$\widehat{absent} = \underset{(1.640)}{12.444} + \underset{(0.669)}{0.968} bluecoll + \underset{(0.712)}{2.049} male - \underset{(0.047)}{0.037} age - \underset{(0.065)}{0.151} tenure - \underset{(0.007)}{0.044} wage$$

$$RSS = 161.95 \quad R^2 = 0.760 \quad n = 48$$

$$H_0: \delta_1 = 0 \qquad H_1: \delta_1 \neq 0$$

$$H_0: \delta_1 = 0 \qquad H_1: \delta_1 > 0$$

$$t = \frac{0.968}{0.669} = 1.45$$

$$H_0: \varphi_1 = 0 \qquad H_1: \varphi_1 \neq 0$$

$$t = \frac{2.049}{0.712} = 2.88$$

EXAMPLE 5.8 Size of firm and gender in determining wage (file wage02sp)

$$\ln(wage) = \beta_1 + \delta_1 female + \theta_1 medium + \theta_2 large + \beta_2 educ + u$$

$$H_0 : \delta_1 = \theta_1 = \theta_2 = 0$$
$$H_1 : H_0 \text{ is not true}$$

$$\widehat{\ln(wage)} = \underset{(0.026)}{1.639} - \underset{(0.021)}{0.327}\, female + \underset{(0.023)}{0.308}\, medium + \underset{(0.023)}{0.168}\, large + \underset{(0.0024)}{0.0499}\, educ$$
$$RSS = 361 \quad R^2 = 0.305 \quad n = 2000$$

$$F = \frac{\left[RSS_R - RSS_{UR}\right]/q}{RSS_{UR}/(n-k)} = \frac{\left[433-361\right]/3}{361/(2000-5)} = 133$$

EXAMPLE 5.9 Is the interaction between females and part-time work significant? (file wage06sp)

$$\ln(wage) = \beta_1 + \delta_1 female + \phi_1 partime + \varphi_1 female \times partime + \beta_2 educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.026)}{2.007} - \underset{(0.022)}{0.259}\, female - \underset{(0.047)}{0.198}\, partime + \underset{(0.058)}{0.167}\, female \times partime + \underset{(0.0024)}{0.0538}\, educ$$

$$RSS = 363 \quad R^2 = 0.238 \quad n = 2000$$

$$H_0 : \varphi_1 = 0$$
$$H_1 : \varphi_1 \neq 0$$

$$t = \frac{0.167}{0.058} = 2.89$$

5.5 Interactions involving dummy variables

EXAMPLE 5.10 Do small firms discriminate against women more or less than larger firms? (file wage02sp)

$$\ln(wage) = \beta_1 + \delta_1 female + \theta_1 medium + \theta_2 large$$
$$+ \varphi_1 female \times medium + \varphi_2 female \times large + \beta_2 educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.027)}{1.624} - \underset{(0.034)}{0.262}\, female + \underset{(0.028)}{0.361} medium + \underset{(0.027)}{0.179} large$$
$$- \underset{(0.050)}{0.159}\, female \times medium - \underset{(0.051)}{0.043}\, female \times large + \underset{(0.0024)}{0.0497}\, educ$$
$$RSS = 359 \quad R^2 = 0.308 \quad n = 2000$$

$$H_0 : \varphi_1 = \varphi_2 = 0$$
$$H_1 : H_0 \text{ is not true}$$

$$F = \frac{\left[RSS_R - RSS_{UR}\right]/q}{RSS_{UR}/(n-k)} = \frac{\left[361 - 359\right]/2}{359/(2000-7)} = 5.55$$

5.5 Interactions involving dummy variables

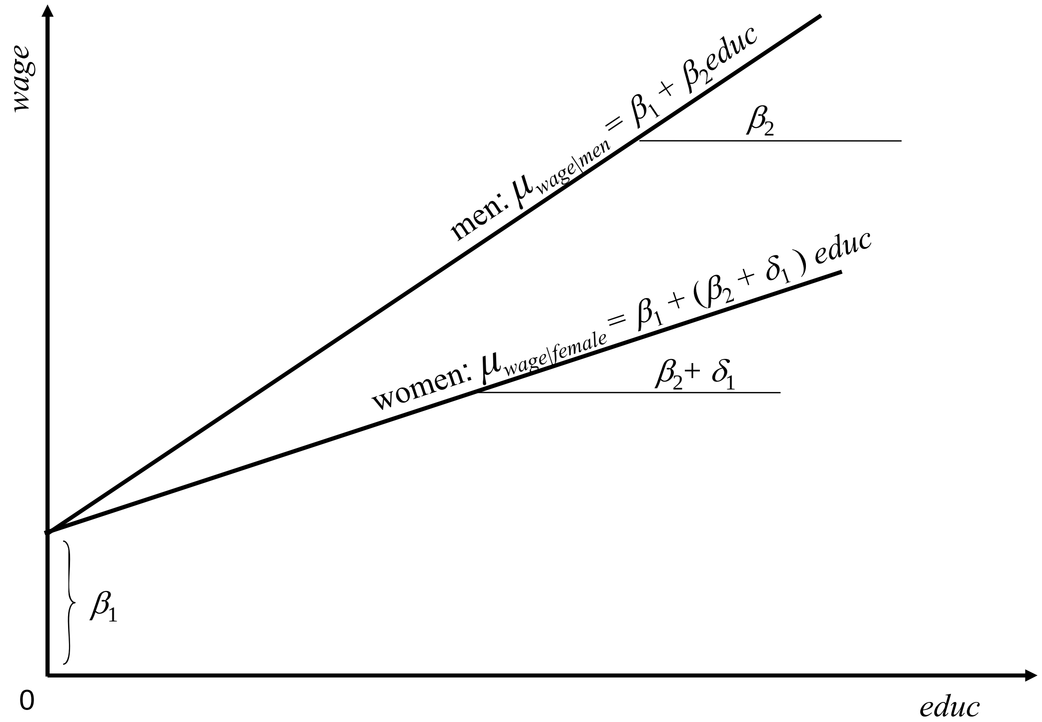

FIGURE 5.2. Different slope, same intercept.

5.5 Interactions involving dummy variable

EXAMPLE 5.11 Is the return to education for males greater than for females? (file wage02sp)

$$wage = \beta_1 + \beta_2 educ + \delta_1 female \times educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.025)}{1.640} + \underset{(0.0026)}{0.0632}\, educ - \underset{(0.0021)}{0.0274}\, educ \times female$$

$$RSS = 400 \quad R^2 = 0.229 \quad n = 2000$$

$$H_0 : \delta_1 = 0$$
$$H_1 : \delta_1 < 0$$

$$t = -\frac{0.0274}{0.0021} = -12.81$$

5.6 Testing structural changes

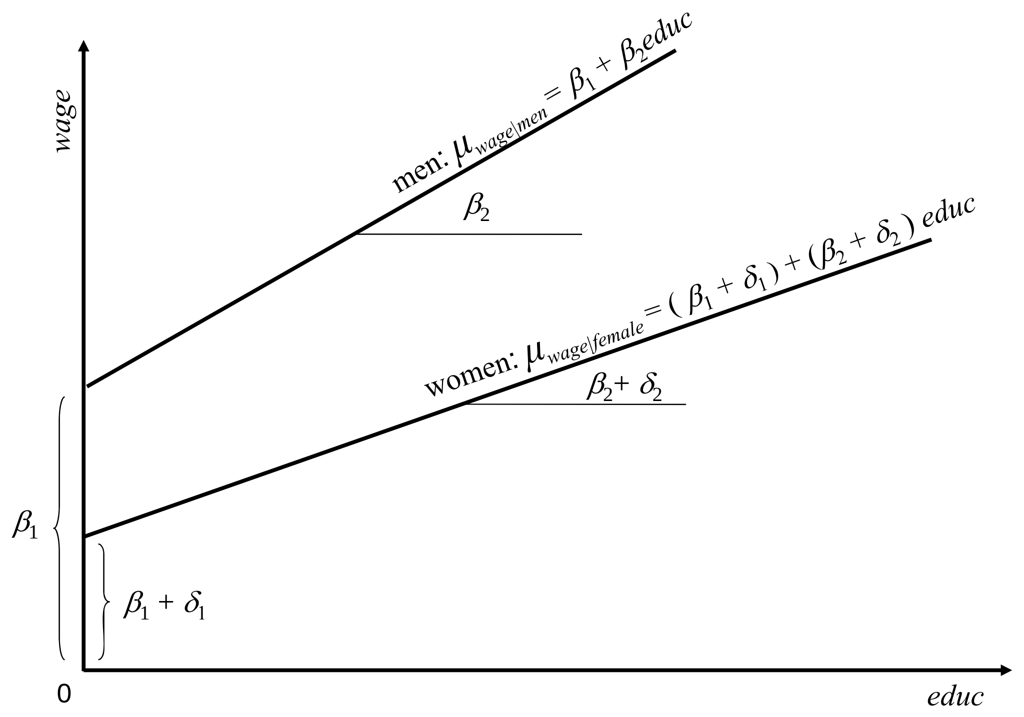

FIGURE 5.3. Different slope, different intercept.

5.6 Testing structural changes

EXAMPLE 5.12 Is the wage equation valid for both men and women? (file wage02sp)

$$wage = \beta_1 + \delta_1 female + \beta_2 educ + \delta_2 female \times educ + u$$

$$H_0: \delta_1 = \delta_2 = 0$$
$$H_1: H_0 \text{ is not true}$$

$$\widehat{\ln(wage)} = \underset{(0.030)}{1.739} - \underset{(0.0546)}{0.3319}\, female + \underset{(0.0030)}{0.0539}\, educ - \underset{(0.0054)}{0.0027}\, educ \times female$$
$$RSS = 393 \quad R^2 = 0.243 \quad n = 2000$$
$$\widehat{\ln(wage)} = \underset{(0.026)}{1.657} + \underset{(0.0026)}{0.0525}\, educ$$
$$RSS = 433 \quad R^2 = 0.166 \quad n = 2000$$

$$F = \frac{[RSS_R - RSS_{UR}]/q}{RSS_{UR}/(n-k)} = \frac{[433 - 393]/2}{393/(2000-4)} = 102$$

5.6 Testing structural changes

EXAMPLE 5.13 Would urban consumers have the same pattern of behavior as rural consumers regarding expenditure on fish? (file demand)

$$\ln(fish) = \beta_1 + \delta_1 urban + \beta_2 \ln(inc) + \delta_2 \ln(inc) \times urban + u$$

$$H_0 : \delta_1 = \delta_2 = 0$$

$$H_1 : H_0 \text{ is not true}$$

$$\ln(fish) = \beta_1 + \beta_2 \ln(inc) + u$$

$$\widehat{\ln(fish)} = \underset{(0.627)}{-6.551} + \underset{(1.095)}{0.678} urban + \underset{(0.087)}{1.337} \ln(inc) - \underset{(0.152)}{0.075} \ln(inc) \times urban$$

$$RSS = 1.123 \quad R^2 = 0.904 \quad n = 40$$

$$\widehat{\ln(fish)} = \underset{(0.542)}{-6.224} + \underset{(0.075)}{1.302} \ln(inc)$$

$$RSS = 1.325 \quad R^2 = 0.887 \quad n = 40$$

$$F = \frac{[RSS_R - RSS_{UR}]/q}{RSS_{UR}/(n-k)} = \frac{[1.325 - 1.123]/2}{1.123/(40-4)} = 3.24$$

5.6 Testing structural changes

EXAMPLE 5.14 Has the productive structure of Spanish regions changed? (file prodsp)

$$\ln(q)=\gamma_1+\alpha_1\ln(k)+\beta_1\ln(l)+\gamma_2 y2008+\alpha_2 y2008\times\ln(k)+\beta_2 y2008\times\ln(l)+u$$

$$\varepsilon_{Q/K(1995)}=\frac{\partial\ln(Q)}{\partial\ln(K)}=\alpha_1 \qquad \varepsilon_{Q/K(2008)}=\frac{\partial\ln(Q)}{\partial\ln(K)}=\alpha_1+\alpha_2$$

$$\varepsilon_{Q/K(1995)}=\frac{\partial\ln(L)}{\partial\ln(K)}=\beta_1 \qquad \varepsilon_{Q/K(2008)}=\frac{\partial\ln(L)}{\partial\ln(K)}=\beta_1+\beta_2$$

$$PEF(1995)=\gamma_1 \qquad PEF(2008)=\gamma_1+\gamma_2$$

$$H_0:\gamma_2=\ \alpha_2=\beta_2 \qquad H_1:H_0 \text{ is not true}$$

$$\ln(q)=\gamma_1+\alpha_1\ln(k)+\beta_1\ln(l)+u$$

$$\text{Unrestricted model}:\ \widehat{\ln(gva)}=\underset{(0.916)}{0.0559}+\underset{(0.185)}{0.6743}\ln(captot)+\underset{(0.185)}{0.3291}\ln(labour)$$

$$-\underset{(2.32)}{0.1088}\,y2008+\underset{(0.419)}{0.0154}\,y2008\times\ln(captot)-\underset{(0.418)}{0.0094}\,y2008\times\ln(labour)$$

$$R^2=0.99394 \quad n=34$$

$$\text{Restricted model}:\ \widehat{\ln(gva)}=-\underset{(0.200)}{0.0690}+\underset{(0.036)}{0.6959}\ln(captot)+\underset{(0.042)}{0.311}\ln(labour) \qquad R^2=0.99392 \quad n=34$$

$$F=\frac{(R^2_{UR}-R^2_R)/q}{(1-R^2_{UR})/(n-k)}=\frac{(0.99394-0.99392)/3}{(1-0.99394)/(34-6)}=0.0308$$

5.6 Testing structural changes

EXAMPLE 5.15 Another way to approach the question of wage determination by gender (file wage02sp)

Female equation

$$\ln(wage) = \beta_{11} + \beta_{21} educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.042)}{1.407} + \underset{(0.0041)}{0.0566}\, educ$$

$$RSS = 104 \quad R^2 = 0.236 \quad n = 617$$

Male equation

$$\ln(wage) = \beta_{12} + \beta_{22} educ + u$$

$$\widehat{\ln(wage)} = \underset{(0.031)}{1.739} + \underset{(0.0032)}{0.0539}\, educ$$

$$RSS = 289 \quad R^2 = 0.175 \quad n = 1383$$

$$F = \frac{\left[RSS_P - (RSS_F + RSS_M)\right]/k}{RSS_F + RSS_M)/(n-2k)} = \frac{\left[433-(104+289)\right]/2}{(104+289)/(2000-2\times 2)} = 102$$

The F statistic must be, and is, the same as in example 5.12.

5.6 Testing structural changes

EXAMPLE 5.16 Is the model of wage determination the same for different firm sizes? (file wage02sp)

$$small: \ln(wage) = \beta_{11} + \delta_{11} female + \beta_{21} edu + u$$
$$medium: \ln(wage) = \beta_{12} + \delta_{12} female + \beta_{22} edu + u$$
$$large: \ln(wage) = \beta_{13} + \delta_{13} female + \beta_{23} edu + u$$

$$H_0: \begin{cases} \beta_{11} = \beta_{12} = \beta_{13} \\ \delta_{11} = \delta_{12} = \delta_{13} \\ \beta_{21} = \beta_{22} = \beta_{23} \end{cases} \qquad H_1: \text{No } H_0$$

$$small \quad \widehat{\ln(wage)} = \underset{(0.034)}{1.706} - \underset{(0.031)}{0.249} female + \underset{(0.0038)}{0.0396} educ \quad RSS = 121 \quad R^2 = 0.160 \quad n = 801$$

$$medium \quad \widehat{\ln(wage)} = \underset{(0.051)}{1.934} - \underset{(0.039)}{0.422} female + \underset{(0.0046)}{0.0548} educ \quad RSS = 123 \quad R^2 = 0.302 \quad n = 590$$

$$large \quad \widehat{\ln(wage)} = \underset{(0.046)}{1.749} - \underset{(0.039)}{0.303} female + \underset{(0.0044)}{0.0554} educ \quad RSS = 114 \quad R^2 = 0.273 \quad n = 609$$

$$F = \frac{\left[RSS_P - (RSS_S + RSS_M + RSS_L)\right]/2k}{(RSS_S + RSS_M + RSS_L)/(n-3k)} = \frac{\left[393 - (121+123+114)\right]/6}{(121+123+114)/(2000-3\times 3)} = 32.5$$

5.6 Testing structural changes

EXAMPLE 5.17 Is the Pinkham model valid for the four periods? (file pinkham)

1907-1914 $sales_t = \beta_{11} + \beta_{21} advexp_t + \beta_{31} sales_{t-1} + u_t$ 1915-1925 $sales_t = \beta_{12} + \beta_{22} advexp_t + \beta_{32} sales_{t-1} + u_t$

1926-1940 $sales_t = \beta_{13} + \beta_{23} advexp_t + \beta_{33} sales_{t-1} + u_t$ 1941-1960 $sales_t = \beta_{14} + \beta_{24} advexp_t + \beta_{34} sales_{t-1} + u_t$

$$H_0: \begin{cases} \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} \\ \beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} \\ \beta_{31} = \beta_{32} = \beta_{33} = \beta_{34} \end{cases} \qquad H_1: \text{No } H_0$$

$$sales_t = \beta_1 + \beta_2 advexp_t + \beta_3 sales_{t-1} + u_t$$

1907-1914 $\widehat{sales_t} = \underset{(603)}{64.84} + \underset{(1.025)}{0.9149}\, advexp + \underset{(0.425)}{0.4630}\, sales_{t-1}$ $SSR = 36017$ $n = 7$

1915-1925 $\widehat{sales_t} = \underset{(190)}{221.5} + \underset{(0.557)}{0.1279}\, advexp + \underset{(0.300)}{0.9319}\, sales_{t-1}$ $SSR = 400605$ $n = 11$

1926-1940 $\widehat{sales_t} = \underset{(112)}{446.8} + \underset{(0.115)}{0.4638}\, advexp + \underset{(0.0827)}{0.4445}\, sales_{t-1}$ $SSR = 201614$ $n = 15$

1941-1960 $\widehat{sales_t} = \underset{(134)}{-182.4} + \underset{(0.241)}{1.6753}\, advexp + \underset{(0.111)}{0.3042}\, sales_{t-1}$ $SSR = 187332$ $n = 20$

$\widehat{sales_t} = \underset{(95.7)}{138.7} + \underset{(0.156)}{0.3288}\, advexp + \underset{(0.0915)}{0.7593}\, sales_{t-1}$ $SSR = 2527215$ $n = 53$

$$F = \frac{\left[SSR_P - (SSR_1 + SSR_2 + SSR_3 + SSR_4)\right]/3k}{(SSR_1 + SSR_2 + SSR_3 + SSR_4)/(n-4k)}$$

$$= \frac{\left[2527215 - (36017 + 400605 + 201614 + 187332)\right]/9}{(36017 + 400605 + 201614 + 187332)/(53 - 4\times 3)} = 9.16$$