Applied Economics

Pooled Cross-Sections

Department of Economics
Universidad Carlos III de Madrid

Two Examples of Cross Sections

CIS surveys

- each month independently, the Centro de Investigaciones Sociológicas samples the Spanish population.
- they ask mainly about political attitudes and collects social and demographic information.

CPS

- each month independently, the US bureau of Labor Statistics quarterly samples the US population (60,000 households).
- main goal: labor situation. Estimation of the number of employed and unemployed people.

- data with this structure are known as pooled cross-sections
- we can take time effects into account

Pooled Cross Sections

- we may want to pool cross sections just to get larger samples: more precision and power.
- but we need to make assumptions about the value of the parameters in each period. It will be useful to pool data if the relationship of interest is constant in the period.
- Example: we will analyze the effect of education on wages with two cross sections from 1978 and 1985 (cps78_85.gdt from Wooldridge in `gretl`).

Pooling

- First, we may assume that parameters remain constant, and estimate the following model:

$$log(w_{it}) = \beta_0 + \beta_1 educ_{it} + u_{it}, t = 1978, 1985$$

Getting

$$\widehat{log(w_{it})} = \underset{(0.071)}{0.984} + \underset{(0.006)}{0.069}\, educ_{it}$$

- The return to education is about 7% both years.
- But the model assumes that the effect is constant.
- Alternatively, we may want to investigate the effect of time, for instance assuming that only the intercept changes

A more flexible model

- We can do that including a binary variable for each period, except one to avoid perfect multicollinearity. Usually, we exclude the first period.
- In our example we include a dummy variable for 1985: the variable $y85$ takes the value one if $t = 1985$ and 0 if $t = 1978$:

$$log(w_{it}) = \beta_0 + \beta_1 educ_{it} + \delta_1 y85_t + u_{it}$$

- In this model we allow that the dependent variable changes in the period but the effect of education is still constant.

A more flexible model: estimation

- The estimated coefficients on the dummy variables show the evolution of the dependent variable with respect to the base period.
- Running OLS:

$$\widehat{log(w)} = \underset{(0.068)}{0.886} + \underset{(0.005)}{0.063}\, educ + \underset{(0.029)}{0.348}\, y85$$

- With $\widehat{\delta_1}$ we capture the wage increase between 1978 and 1985, keeping education constant. Wages increased on average 34.8% between 1985 and 1978, keeping education constant.
- Estimated return to education is 6.3%. And of course the same in both years.

Time variations in the returns to education

- We can investigate whether the relations of interest change over time.
- We need to include interactions between the variable of interest and the time dummies:
- In our example we need interactions between *educ* and each time dummy ($educ \times y78$ and $educ \times y85$).
- Then we have two alternatives:

Alternative 1

- Run OLS of the dependent variable on all the interactions plus a constant: each slope measures the returns of education each period:

$$log(w_{it}) = \gamma_0 + \gamma_1 educ_{it} \times y78 + \gamma_2 educ_{it} \times y85_t + \gamma_3 y85 + u_{it}$$

- γ_1 represents the effect of education on wages in 1978 and γ_2 represents the effect of education on wages in 1985.

$$\widehat{log(w)} = \underset{(0.087)}{1.030} + \underset{(0.007)}{0.052}\,\text{educ78} + \underset{(0.008)}{0.077}\,\text{educ85} + \underset{(0.137)}{0.029}\,\text{y85}$$

- The estimated effect of education on wages in 1978 is 5.2%; in 1985 is 7.7%. To test if they are similar we test: $H0)\gamma_2 - \gamma_1 = 0$

Alternative 2

- Run OLS on a constant, *educ* and all interactions except one, for instance the one for the first year.
- Each slope for the interactions measures how the return for that year differ from the reference year (in this case 1978).

$$log(w_{it}) = \alpha_0 + \alpha_1 educ_{it} + \alpha_2 educ_{it} \times y85_t + \alpha_3 y85 + u_{it}$$

- OLS estimation:

$$\widehat{log(w)} = \underset{(0.087)}{1.030} + \underset{(0.007)}{0.052}\,\text{educ} + \underset{(0.011)}{0.025}\, educ \times y85 + \underset{(0.137)}{0.030}\, y85$$

- The estimated effect of education on wages in 1978 is 5.2%; in 1985 is 2.5 percentage points higher.
- In this case to test if the effect remained constant we test: $H0)\alpha_2 = 0$.

The Chow Test for Structural Change

All parameters can change

$$log(w_{it}) = \beta_{0t} + \beta_{1t} educ_{it} + u_{it}$$

- We can use an F test to check if the parameters are different in two different periods. This test is known as a Chow test.
- To compute the F test we need to compute the *SSR* (the sum of squares of residuals) from the restricted and non-restricted model.
- The pooled regression is the restricted model. The non-restricted SSR is obtained as the sum of the SSR for the separate models (one per period).

Chow (cont.)

- In our example: $log(w) = \beta_0 + \beta_1 educ + \delta_1 y85 + \delta_2 educ * y85 + u$
- In this equation $\delta_1 = 0$ implies that there is no difference in the constant, $\delta_2 = 0$ implies no difference in the return to education.
- The null hypothesis for the Chow test is: $H_0 : \delta_1 = \delta_2 = 0$ (no structural change).
- To obtain a heteroskedastic-robust test in `gretl`, run the pooled regression with interactions using robust standard errors and test the appropriate restrictions using `omit`.
- Or run the pooled regression with no interactions and use the command $chow\ d2 - -dummy$.

Example cont.

- Based on the previous model we want to analyze the evolution of wage differentials by gender.
- We estimate the following model:
  $log(w_{it}) = \beta_0 + \beta_1 y85 + \beta_2 educ_{it} + \beta_3 educ_{it} \times y85_t + \beta_4 exper + \beta_5 exper^2 + \beta_6 union + \beta_7 female + \beta_8 female \times y85_t + u_{it}$
- The estimated effect of education on wages in 1978 is about 7.5% and 9.3% in 1985, and the difference is significant.
- In 1978, holding other variables fixed, a woman earned 31.7% less than a man. This difference decreased around 8.5 pp in the period, but we reject the null that this difference is zero against the alternative that it is positive.

Example - some questions

- Interpret the coefficient on $y85$. Is it interesting?
- If we assume that all the other variables remain constant, Which is the estimated increase in the wage of a man with 12 years of education?
- If we use real wages, which coefficients will change with respect to the previous estimation? If you want to do it you can use 1978 dollars, the deflator for 1985 is 1.65.
- Modify the model to analyze if there is a wage differential for union members and if this differential has changed in this period.