Question 5

Kai Liao

November 2022

Question: Applied linear regression. You have a large cross section of data on wages along with all the demographic information you could have - call these variables X, your vector of controls. The only education information you have is number of years of schooling. You speak with a two economists and both agree there are breaks in average wages at 13 years of education (high school graduation) and at year 16 (college graduation). Also, both agree that wages increase with years of education, but at a decreasing rate. However, economist 1 argues that the rate of decrease in the value of education is smaller for women than men. Economist 2 disagrees and suggests that the returns to education decrease at the same rate for women and men, but looks like it is less unconditionally because women have a lower wages on average. Write down one linear regression model to test who is right. Be specific about hypothesis tests.

Solution: We can run the following fixed effect regression, written in simple OLS form, as follow

$$\begin{aligned} wages = {} & \alpha_0 + \alpha_1 * female + \beta_0 * \mathbb{1}[educ \leq 13] + \beta_1 * \mathbb{1}[13 < educ \leq 16] \\ & + \beta_2 * \mathbb{1}[educ > 16] + \beta_3 * \mathbb{1}[educ \leq 13] * female + \beta_4 * \mathbb{1}[13 < educ \leq 16] * female \\ & + \beta_5 * \mathbb{1}[educ > 16] * female + \delta X + \varepsilon \end{aligned} \tag{1}$$

Using the F test to jointly test the null hypothesis that β_3, β_4, and β_5 are equal to 0.

$$F = \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})/q}{SSR_{\text{unrestricted}}/(n-k-1)}$$

where $SSR_{\text{unrestricted}}$ is the SSR for model (1) and $SSR_{\text{restricted}}$ is the SSR for model (2).

$$\begin{aligned} wages = {} & \alpha_0 + \alpha_1 * female + \beta_0 * \mathbb{1}[educ \leq 13] + \beta_1 * \mathbb{1}[13 < educ \leq 16] \\ & + \beta_2 * \mathbb{1}[educ > 16] + \delta X + \varepsilon \end{aligned} \tag{2}$$