ECON 101 - TA section 2

Beatriz Pousada

October 16, 2019

What we will cover in this section:

- OLS regression (linear, log-log log-linear, linear-log)
- Interactions to show heterogeneity
- fixed effects
- diff in diff - only in the code
- RDD - only in the code

1 OLS regressions

Here we are going to review how to interpret regressions, to do so, we are using a very simple example of regressing earnings on years of schooling. We know there are several endogeneity problems with regressing earnings on years of education, but let's not discuss them now. Suppose we are not interested on the causal effect of earnings on education, we just want to know the relationship between those two variables.

1.1 linear-linear

Let's start supposing we are regressing those two variables linearly:

$$\text{earn}_{it} = \beta_0 + \beta_1 \text{educ}_{it} + u_{it} \tag{1}$$

What does the coefficients β_0 and β_1 mean? If we just apply the expectation operator to both sides of these equation we get:

$$E(earn_{it}|educ_{it}) = \beta_0 + \beta_1 \text{educ}_{it} \tag{2}$$

Let's calculate the average earnings for someone with zero education:

$$E(earn_{it}|educ_{it} = 0) = \beta_0 \tag{3}$$

Now let's calculate the average earnings for someone with education e and for someone with education $e + 1$:

$$E(earn_{it}|educ_{it} = e) = \beta_0 + \beta_1 e E(earn_{it}|educ_{it} = e + 1) \qquad = \beta_0 + \beta_1(e + 1) \tag{4}$$

If we take the difference:

$$E(earn_{it}|educ_{it} = e + 1) - E(earn_{it}|educ_{it} = e) = [\beta_0 + \beta_1(e + 1)] - [\beta_0 + \beta_1(e + 1)] = \beta_1 \tag{5}$$

So β_1 is how much more earnings individuals with one more year of education have on average in our data.

1.2 linear-log

Now suppose we are taking the log of years of education, but still running earnings linearly:

$$\text{earn}_{it} = \beta_0 + \beta_1 \log(\text{educ})_{it} + u_{it} \tag{6}$$

as before, let's calculate the expectation:

$$E(earn_{it}|\log(educ)_{it}) = \beta_0 + \beta_1 \log(educ)_{it} \tag{7}$$

If we calculate the average earnings for someone with one year of education we get that $log(1) = 0$:

$$E(earn_{it}|\log(educ)_{it} = 0) = \beta_0 \tag{8}$$

and if we calculate the difference between the average earnings of someone with $log(educ)_{it} = le+1$ and $log(educ)_{it} = le$ we also get the same as before:

$$E(earn_{it}|\log(educ)_{it} = le + 0.01) - E(earn_{it}|\log(educ)_{it} = le) = 0.01\beta_1 \tag{9}$$

How much more education someone with $log(educ)_{it} = le+0.01$, has than someone with $log(educ)_{it} = le$? Let's first define $log(educ_a) = le + 0.01$ and $log(educ_b) = le$ and then take the difference:

$$\log(educ_a) - \log(educ_b) = 0.01 \Rightarrow \log\left(\frac{educ_a}{educ_b}\right) = 0.01 \Rightarrow \frac{educ_a}{educ_b} = \exp(0.01) \approx 1.01 \tag{10}$$

If education increases by 1% we expect earnings to increase by $0.01\beta_1$ dollars.

1.3 log-linear

Now suppose we will take log-earnings, but keep years of education linear:

$$\log(earn)_{it} = \beta_0 + \beta_1 \text{educ}_{it} + u_{it} \tag{11}$$

What does the coefficients β_0 and β_1 mean? If we just apply the expectation operator to both sides of these equation we get:

$$E(\log(earn)_{it}|educ_{it}) = \beta_0 + \beta_1 \text{educ}_{it} \tag{12}$$

Let's calculate the average earnings for someone with zero education:

$$E(\log(earn)_{it}|educ_{it} = 0) = \beta_0 \tag{13}$$

Now let's take the difference, for people with one year of education apart:

$$E(\log(earn)_{it}|educ_{it} = e + 1) - E(\log(earn)_{it}|educ_{it} = e) = \beta_1 \tag{14}$$

What does it mean? Let's define the value of earnings such that $E(\log(earn)_{it}|educ_{it} = e + 1) = \log(earn_a)$ and $E(\log(earn)_{it}|educ_{it} = e) = \log(earn_b)$. So if we take the difference:

$$\log(earn_a) - \log(earn_b) = \beta_1 \Rightarrow \log\left[\frac{earn_a}{earn_b}\right] = \beta_1 \Rightarrow \frac{earn_a}{earn_b} = \exp(\beta_1) \tag{15}$$

let's transform this into percentages:

$$\left(\frac{earn_a}{earn_b} - 1\right) * 100 = 100 * (\exp(\beta_1) - 1) \approx 100 * \beta_1, \text{ if } -0.1 \leq \beta_1 \leq 0.1 \tag{16}$$

Individuals with one more year of education have $100 * \beta_1$ percent more earnings.

1.4 log-log

Now let's do the log-log case:

$$\log(earn)_{it} = \beta_0 + \beta_1 \log(educ)_{it} + u_{it} \tag{17}$$

similar to before, if we calculate the difference between the average earnings of someone with $log(educ)_{it} = le + 1$ and $log(educ)_{it} = le$ we also get the same as before:

$$E(\log(earn)_{it} | \log(educ)_{it} = le + 0.01) - E(\log(earn)_{it} | \log(educ)_{it} = le) = 0.01\beta_1 \tag{18}$$

we saw before that this is an 1% increase in education. Combining with what we saw in the log-linear case, we expect to observe a 1% increase in education to reflect in a $100 * 0.01 * \beta_1 = \beta_1$ percent increase in earnings.

2 Interactions to show heterogeneity

Suppose now that we are keeping the log-linear example (this is the one that made more sense in the graph), but we want to see if there is any heterogeneity by gender on the earnings difference by years of education.

$$\log(earn)_{it} = \beta_0 + \beta_1 \text{educ}_{it} + u_{it} \tag{19}$$

To do so, we run the following regression:

$$\log(earn)_{it} = \beta_0 + \beta_1 \text{educ}_{it} + \beta_2 \text{women}_i * \text{educ}_{it} + \beta_3 \text{women}_i + u_{it} \tag{20}$$

For men we have the same interpretation, men with one more year of education have $100 * \beta_1$ percent more earnings.

$$E(\log(earn)_{it} | educ_{it} = e + 1, women_i = 0) - E(\log(earn)_{it} | educ_{it} = e, women_i = 0) = \beta_1 \tag{21}$$

What about for women?

$$E(\log(earn)_{it} | educ_{it} = e + 1, women_i = 1) = \beta_0 + \beta_3 + (\beta_1 + \beta_2)(e + 1) \tag{22}$$

$$E(\log(earn)_{it} | educ_{it} = e, women_i = 1) = \beta_0 + \beta_3 + (\beta_1 + \beta_2)e \tag{23}$$

Taking the difference:

$$E(\log(earn)_{it} | educ_{it} = e+1, women_i = 1) - E(\log(earn)_{it} | educ_{it} = e, women_i = 1) = \beta_1 + \beta_2 \tag{24}$$

Women with one more year of education have $100 * (\beta_1 + \beta_2)$ more earnings. The percent increase in earnings for women with one more year of education is $100 * \beta_2$ percentage points bigger than for men.

3 Fixed-Effects

There are many reasons why regressing earnings on years of education is not causal, but some of those reasons we can control for. One example of those variables is age, older individuals might have less years of education because when they were young staying school was less common. Moreover, older individuals might have smaller earnings because they are not able to work anymore. One way of tackling this issue is to simply include an age variable in the regression.

$$\log(earn)_{it} = \beta_0 + \beta_1 \text{educ}_{it} + \beta_3 \text{age}_{it} + u_{it} \tag{25}$$

However, when we do this, we are assuming that the relationship between age and earnings and the relationship between age and education is linear. To relax the linearity assumption we can simply include age fixed effects in the regression.

$$\log(earn)_{it} = \beta_0 + \beta_1 \text{educ}_{it} + \sum_{a=\underline{a}}^{\bar{a}} \delta_a \mathbb{1}(age_{it} == a) + u_{it} \tag{26}$$

When including fixed effects it's important to make sure you are not including one dummy for each observation in the data-set and that all the variables in your regression vary within the fixed effect groups. That's why when we add household id fixed effects we don't get a result, because education doesn't vary within household id.

Notice that another problem with our earnings regressions is that we didn't deflate the earnings variable. One easy way to account for that without having to go to the cpi is to include year fixed effects in the regression.

$$\log(earn)_{it} = \beta_0 + \beta_1 \text{educ}_{it} + \sum_{a=\underline{a}}^{\bar{a}} \delta_a \mathbb{1}(age_{it} == a) + \sum_{w=0}^{T} \theta_w \mathbb{1}(t == w) + u_{it} \tag{27}$$