ECON 102B - TA section

Beatriz Pousada

February 28, 2020

What we will cover in this section:

- OLS regression (linear, log-log log-linear, linear-log)
- Interactions to show heterogeneity

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 $e + 1$:

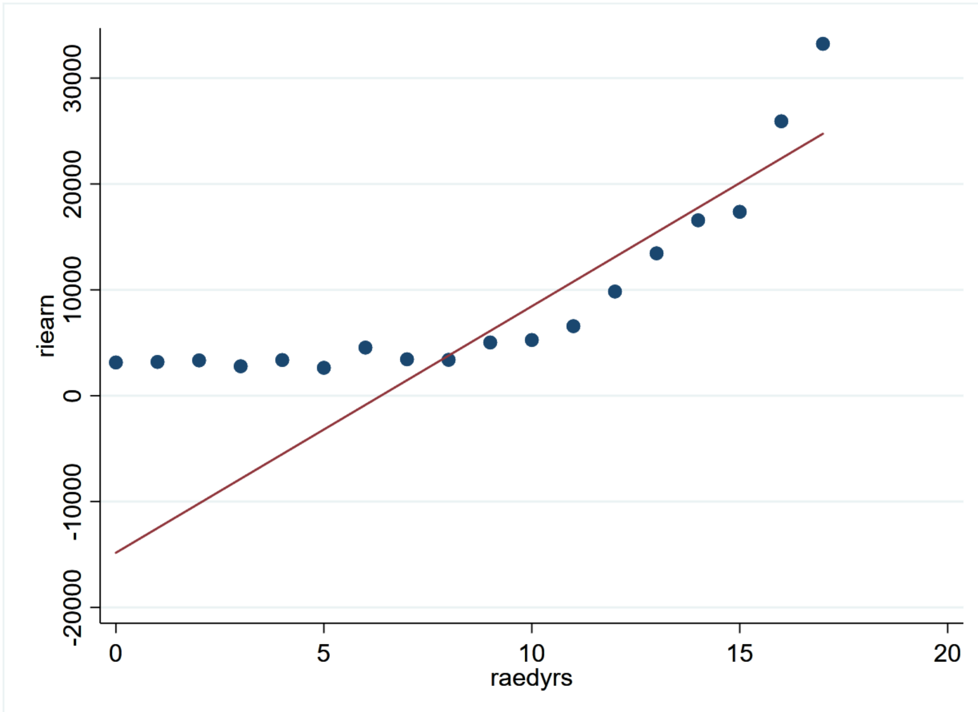
$$E(earn_{it}|educ_{it} = e) = \beta_0 + \beta_1 e E(earn_{it}|educ_{it} = e + 1) \qquad = \beta_0 + \beta_1(e + 1) \tag{4}$$

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.

Figure 1: Relationship between earnings and education - data from the HRS only individuals with more than 50 years old




```
. reg riearn raedyrs, robust

Linear regression                              Number of obs     =    233,396
F(1, 233394)      =    7333.22
Prob > F          =     0.0000
R-squared         =     0.0408
Root MSE          =      37873

------------------------------------------------------------------------------
|               Robust
riearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
raedyrs |   2328.381   27.18985    85.63   0.000     2275.089    2381.672
_cons |  -14836.95   275.2679   -53.90   0.000    -15376.47   -14297.43
------------------------------------------------------------------------------
```

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:

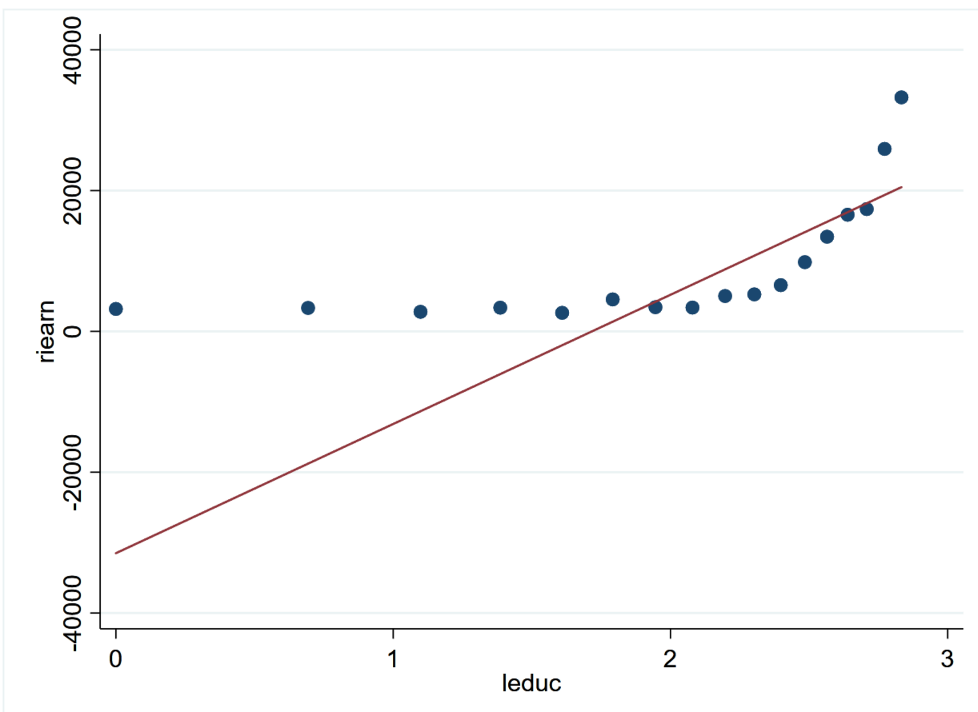
$$E(earn_{it}| \log(educ)_{it} = le + 0.01) - E(earn_{it}| \log(educ)_{it} = le) = 0.01\beta_1 \tag{9}$$

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.

Figure 2: Relationship between earnings and log-education - data from the HRS only individuals with more than 50 years old



```
. reg riearn leduc, robust

Linear regression                               Number of obs     =    231,423
F(1, 231421)      =    6868.41
Prob > F          =     0.0000
R-squared         =     0.0275
Root MSE          =      38268

------------------------------------------------------------------------------
|               Robust
riearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
leduc |   18351.09   221.4286    82.88   0.000      17917.1    18785.09
_cons |  -31505.23   492.8452   -63.93   0.000    -32471.19   -30539.26
------------------------------------------------------------------------------
```

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:

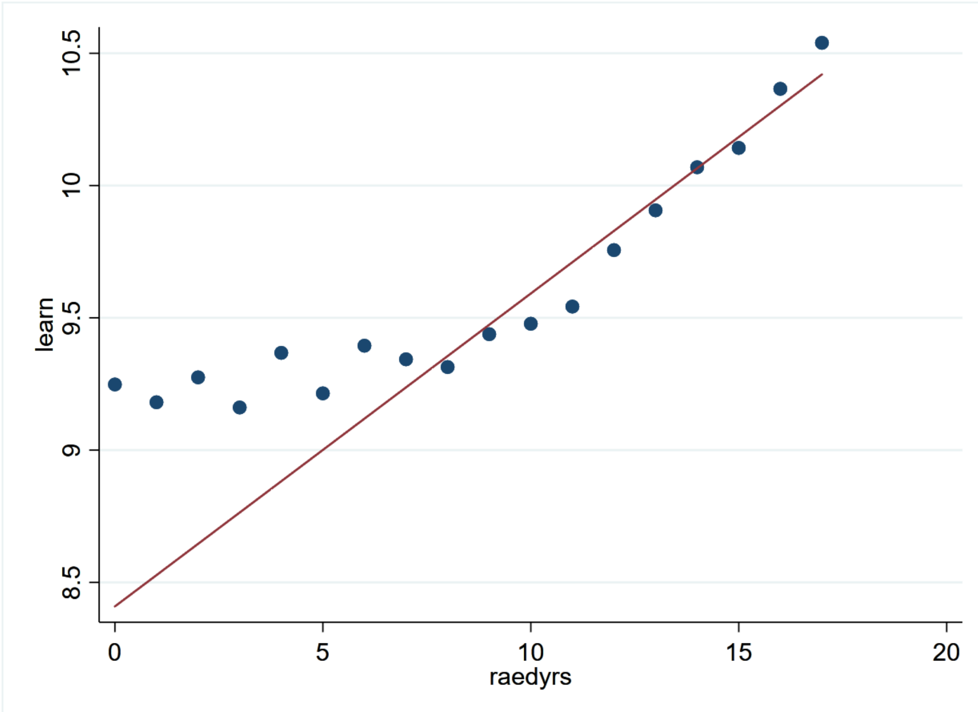
$$\log(earn_a) - \log(earn_b) = \beta_1 \Rightarrow \log\left[\frac{earn_a}{earn_b}\right] = \beta_1 \Rightarrow \frac{earn_a}{earn_b} = \exp(\beta_1) \tag{15}$$

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.

Figure 3: Relationship between log-earnings and education - data from the HRS only individuals with more than 50 years old




```
. reg learn raedyrs, robust

Linear regression                              Number of obs     =     83,342
F(1, 83340)       =    6108.38
Prob > F          =     0.0000
R-squared         =     0.0748
Root MSE          =     1.2299

------------------------------------------------------------------------------
|               Robust
learn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
raedyrs |   .1182874   .0015135    78.16   0.000      .115321    .1212538
_cons |   8.409132   .0201424   417.48   0.000     8.369653    8.448611
------------------------------------------------------------------------------
```

1.4 log-log

Now let's do the log-log case:

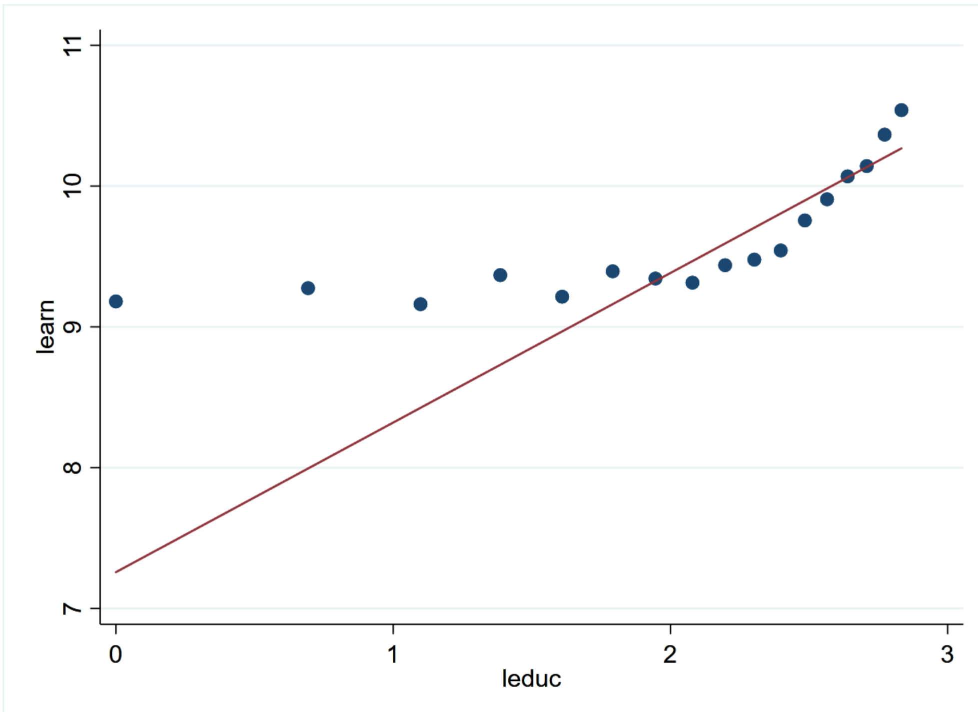
$$\log(earn)_{it} = \beta_0 + \beta_1 \log(educ)_{it} + u_{it} \tag{17}$$

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.

Figure 4: Relationship between log-earnings and log-education - data from the HRS only individuals with more than 50 years old




```
. reg learn leduc, robust

Linear regression                                    Number of obs     =     83,032
F(1, 83030)       =    3421.90
Prob > F          =     0.0000
R-squared         =     0.0568
Root MSE          =     1.2408

------------------------------------------------------------------------------
|               Robust
learn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
leduc |    1.06233   .0181604    58.50   0.000     1.026736    1.097924
_cons |   7.258118   .0464536   156.24   0.000     7.167069    7.349166
------------------------------------------------------------------------------
```

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?

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

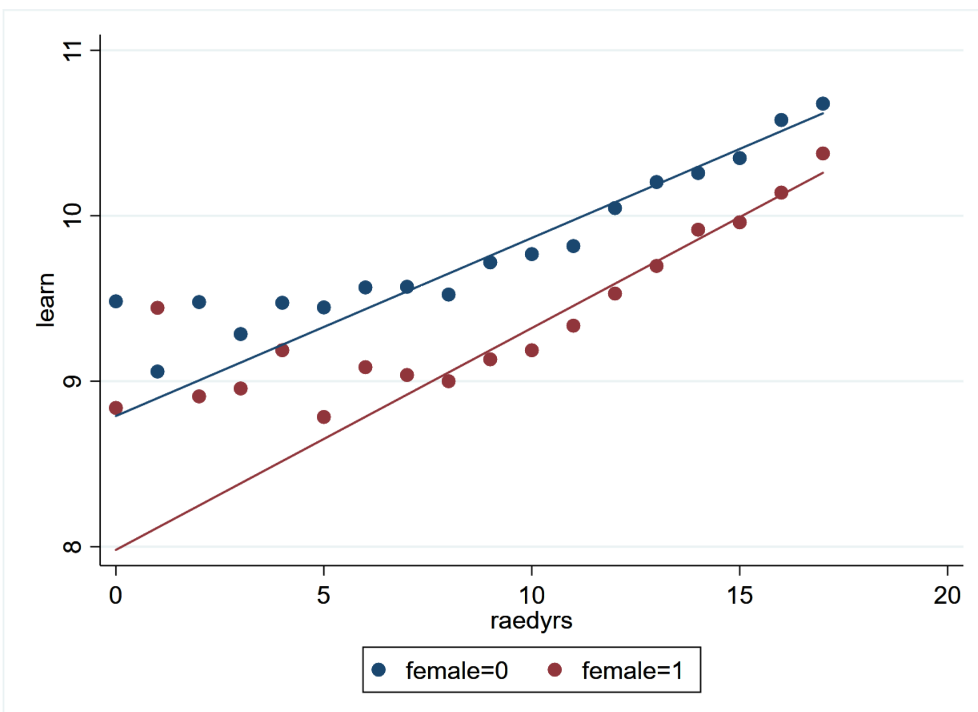
$$E(\log(earn)_{it}|educ_{it} = e, women_i = 1) = \beta_0 + \beta_3 + (\beta_1 + \beta_2)e \tag{23}$$

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.

Figure 5: Relationship between log-earnings and education by gender - data from the HRS only individuals with more than 50 years old



```
. gen educWomen = raedyrs*female
.
. reg learn raedyrs educWomen female, robust

Linear regression                               Number of obs     =     83,342
F(3, 83338)       =    3118.08
Prob > F          =     0.0000
R-squared         =     0.1084
Root MSE          =     1.2073

------------------------------------------------------------------------------
|               Robust
learn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
raedyrs |   .1075845   .0019482    55.22   0.000     .1037661    .1114029
educWomen |    .026473   .0030034     8.81   0.000     .0205864    .0323597
female |  -.8092315   .0398843   -20.29   0.000    -.8874044   -.7310587
_cons |   8.789897   .0258254   340.36   0.000      8.73928    8.840515
------------------------------------------------------------------------------
```