

IDEA · Econometrics I · 2021/22

Assignment 1 (Due January 27th, 2023)

Please carefully follow the instructions given at the end.

1. Consider the following figure (included in our class slides):

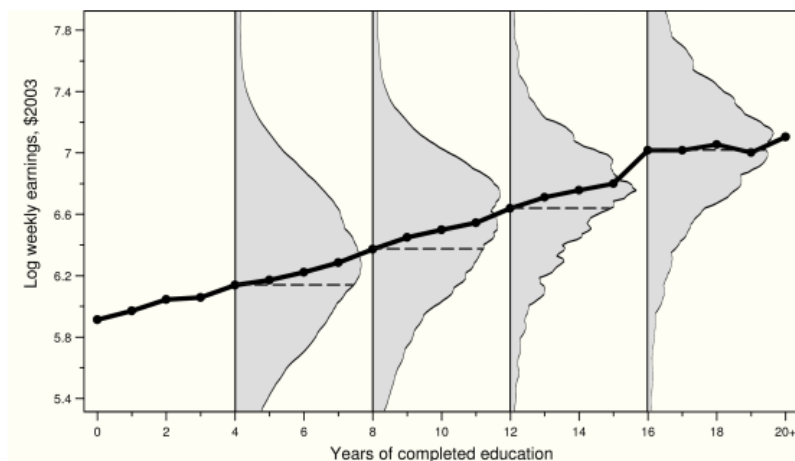


Figure 3.1.1: Raw data and the CEF of average log weekly wages given schooling. The sample includes white men aged 40-49 in the 1980 IPUMS 5 percent file.

The figure above, reproduced from Angrist&Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (page 31), summarizes more than 300,000 observations corresponding to a sample of white males in 1980s in the United States, regarding their years of education (*educ*) and the natural logarithm of their weekly earnings (*lwkllywge*). Datafile *Assig1.csv* includes the observations for variables *educ* and *lwkllywge* (i.e. $lwkllywge = \ln(wages)$) similar (not identical) to the ones used to produce the figure above.

- (a) Provide an estimate of average weekly pay for people with 12 years of education. This statistic would be an estimate of which parameter?
- (b) With the help of *Matlab*, reproduce the part of Figure 3.1.1 that includes the conditional sample means and the thick black function joining them. (Please notice you are only asked to reproduce the filled circles and the thick black line joining them, not the 4 Kernel density estimates of the conditional distributions included in the plot). Since the data file provided is slightly different than the one used by the authors, you will not get exactly the same values as those included in Figure 3.1.1. Include plot and Matlab code as your answer.
- (c) What does the thick black function you got allow you to say about the relationship between education and earnings? Explain in just a sentence. Try to be as rigorous as possible.
- (d) Consider that now, instead of focusing on conditional sample means, we focussed on conditional sample medians. With the help of *Matlab* calculate the conditional medians, and create a new plot including both, the function joining the conditional sample medians together with the function joining conditional sample means. Include the plot in your answer. Comment.

- (e) Let us focus our attention to the thick black function you got in (1b) joining all the conditional sample means (we will ignore now the conditional medians). Your job now is to fit a line through these sample means using *Matlab's polyfit* function. (i) Present a plot including the original sample means and the fitted line. (ii) What value did you get for the slope of the line?
- (f) Complete this sentence: "*According to the value of the slope of the line fitted in the question above, we can say that*" Be as specific as possible.

2. Consider the following simple linear regression model:

$$y = \beta_1 + \beta_2 x_2 + \epsilon,$$

where $\epsilon \equiv y - \beta_1 - \beta_2 x_2$ and

$$\beta_1 = E(y) - \beta_2 E(x_2) \quad \text{and} \quad \beta_2 = \frac{\text{cov}(x_2, y)}{\text{var}(x_2)}$$

Proof that $\beta_1 + \beta_2 x_2$ is the *BLP*, in *MSE* sense, of variable y given variable x_2 .

3. Proof the *Decomposition Theorem*, included in slide 1(12), that states that we can always decompose a variable y as:

$$y = E(y/x) + \epsilon$$

with (i) $E(\epsilon/x) = 0$ (ϵ mean independent of x), and (ii) ϵ uncorrelated with x and with any function of x .

Instructions for the assignment (*Please read them!*)

- Please hand in the assignment to Luis Ignacio at the beginning of TA session on Friday.
- Assignments are turn in individually.
- Pages need to be staples with your name on the top of the first page.
- Questions have to be answered in orderly manner. That is: Q1: 1.a, 1.b,..... Q2:....
- Requested *Matlab* outputs (statistics, plots,...) have to be inserted in the corresponding position in your answer (i.e., do not include them all at the end). *Matlab* code can be included as an appendix at the end.
- Answers, except for *Matlab* work, can be answered using a pen (i.e., not necessarily in latex).
- Original work is expected.