Microeconometrics: Part I

-Tilburg University course 35M1C5-

Instructor: Dr. Alexandros Theloudis

Autumn 2022

Course assignment 1

The students should work on this assignment after Lecture 5 'Causal effects: examples and hands-on practice'. Please provide brief but complete answers. Hand in your answers just before Lecture 8, by **Monday October 3, 2022** at 2:45pm.

1. **(randomization; 60 points)** In this question, you will discuss a popular paper about a social experiment:

   - Banerjee, Cole, Duflo & Linden, 2007, Remedying Education: Evidence from Two Randomized Experiments in India, *Quarterly Journal of Economics*.

   The paper is available on Canvas. In this paper the authors study the effectiveness of two learning interventions. Please, read the paper carefully and answer the following questions. Keep your answers short but concise.

   (a) What are the two learning interventions? [10/60]

   (b) What are the potential outcomes of the interventions? [10/60]

   (c) Describe the experimental design. What treatment effects are the authors interested in? What empirical methods do they use to estimate them? [10/60]

   (d) Is random assignment successful in this experiment? How do the authors test if treatment assignment was random? [10/60]

   (e) Is selective attrition a problem? How do the authors test for it? [5/60]

   (f) Can Balsakhis have direct effects on students who get help from them and indirect effects on students who do not get help from them? How do the authors disentangle these effects? [10/60]

   (g) External validity is a common concern when drawing conclusions from social experiments. How does this paper attempt to address this concern? [5/60]

2. **(linear regression; 20 points)** Suppose the effect of x on y is described by the linear relationship

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Suppose the usual linear regression assumptions hold (LR.1-LR.4, lecture 1, slide 16). The ordinary least squares estimate of the slope coefficient β_1 is given by

$$\hat{\beta}_1 = \frac{n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\mathrm{Cov}(y_i, x_i)}{\mathrm{Var}(x_i)},$$

where $\bar{x} = \mathbb{E}(x)$ and $\bar{y} = \mathbb{E}(y)$ are the sample means.

Some textbooks use matrix notation and write this as $(\hat{\beta}_0, \hat{\beta}_1)' = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}$, where $\boldsymbol{x}$ is the $n \times 2$ matrix of $\big((1, x_i)\big)_{i=1,\ldots,N}$ observations and $\boldsymbol{y}$ is the $n \times 1$ matrix of $\big(y_i\big)_{i=1,\ldots,N}$ observations. Prove that the two expressions are equivalent.

3. **(omitted variables; 20 points)** Suppose y_i is determined by the linear relationship

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i,$$

where $\mathbb{E}(u_i|x_{1i}, x_{2i}) = 0$. x_{2i} is not observed in our data so one can only hope to estimate

$$y_i = \beta_0 + \beta_1 x_{1i} + \nu_i, \tag{1}$$

where $\nu_i = \beta_2 x_{2i} + u_i$. Derive an expression for the omitted variables bias that running OLS on (1) implies. Suggest a way to deal with this bias.

End of assignment 1