ECO475 HW2[1]

1. Consider the following potential outcome model: $Y = DY_1 + (1-D)Y_0$. We add the following structure to the treatment selection by assuming

$$D = 1[Y_1 \geq Y_0],$$

that is, an individual choose $D = 1$ whenever his/her Y_1 is no smaller than Y_0. Furthermore, let

$$Y_d = X'\beta_d + U_d,$$

where X is observed characteristics, β_d is a vector of coefficient, and U_d is unobserved.

Let $(U_1, U_0)|X \sim N((0,0), \sigma_1^2, \sigma_0^2, \sigma_{01})$. So the two unobserved latent variables are independent with X, and follows joint normal distribution. Their marginal expectations are 0, marginal standard deviations are σ_1 and σ_0, respectively, and their covariance is σ_{01}.

The fix the idea, suppose there are two sectors and we can consider Y_d is the salary that an individual could earn in sector d, and D is the decision on which sector to work in. Of course, a rational people will choose to work in the sector where he/she can earn higher salary. However, only one of the pair (Y_1, Y_0) is observed, depending on which sector the individual is actually working in.

   (a) Consider linear regression of Y on (D, X). Show that the coefficient of D can not consistently estimate the average wage gap between sectors: $\mathbb{E}[Y_1 - Y_0]$.

   (b) Discuss the economic interpretation of $\mathbb{E}[Y_d|D = d, X = x]$ and $\mathbb{E}[Y_d|X = x]$, respectively.

   (c) One key parameter measures the consequence caused by self-selection. It is summarize best by $\Delta_d(x) = \mathbb{E}[Y_d|D = d, X = x] - \mathbb{E}[Y_d|X = x]$. Express $\Delta_d(x)$ explicitly as a function of X and model parameters.

   (d) Propose a way to estimate model parameters. Write down the sample objective function that you will maximize/minimize to calculate the estimates.

2. Read the paper "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records" by J. Angrist (1990), *American Economic Review (AER)* and write a small report (in about two pages), in which you will answer the following questions:

   (a) What is the research question this paper focuses on and why it matters?

   (b) To answer this research question, what do you think is the main difficulty in the econometric modeling? Does OLS work? Why?

   (c) How do you cast this research question into the potential outcome model (POM)? For instance, what is D, Y_0, and Y_1 etc.

   (d) If you want to use the POM to answer the research question, what assumptions you need to impose? Discuss the economic interpretation of your assumptions.

[1]Due by 11:59pm, March 29, 2021. Online submission through Quercus. Submit a PDF file.

(e) Under the framework of POM, what parameter that the 2SLS procedure would identify in this empirical context? Do you think it is useful information for policy makers, why or why not?

(f) Any other open comments that you have on this paper (Angrist, 1990, AER) and the POM discussed in Imbens and Angrist (1994, Econometrica).

Note: this is an individual assignment, work with yourself! Grades will be given based on the quality of the answers.