Problem Set 1

Himani Pasricha (20201143)

Dimitrios Argyros (20201790)
Tatiana Bezdenezhnykh (20201138)
Matthew Abagna(20201136)
Manuel Rodriguez()

January 26, 2021

Exercise 2.1

1. fff

    - Arrow 1 suggests being female (gender) results in being discriminated against.
    - Arrow 2 suggests being female results in being sorted to some specific jobs due to natural limitations or self selection. For example, female prefers to work as nurses because of being more socially sensitive compare to men.(citation).
    - Arrow 3 suggests discrimination affects occupational sorting. Females might be discriminated against some jobs or some occupation irrespective of physical limitation or self selection(For example, high level army position). The line is dust because theoretically the relation between occupational sorting and discrimination is unclear (stated as an assumption in a question).
    - Arrow 4 represents the main causal relationship we want to examine i.e. the relationship between discrimination and earnings. People discriminated against receive lower wages compare to the ones who are not discriminated.
    - Arrow 5 represents the relationship between occupational sorting and earnings. Some occupations yields higher earnings like doctors , judges and pilots.
    - Arrow 6 and Arrow 7 depicts the relationship between (unobserved) ability and occupational sorting and earnings respectively.

    Arrow from F to Y suggest that the marginal productivity of male and female is different and leads to different earnings (because of Marginal productivity). Therefore, the absence of arrow suggest by assumption male and female are equally productive.

    We can have an arrow through female to ability if nature doesn't treat male and female equally with regard to ability.

2. • Main Causal Effect : $D \longrightarrow Y$

   (a) Backdoor Path 1 : $D \longrightarrow O \longleftarrow A \longrightarrow Y$

   (b) Backdoor Path 2 : $D \longrightarrow O \longrightarrow Y$

   (c) Backdoor Path 3 : $D \longleftarrow F \longrightarrow O \longrightarrow Y$

   (d) Backdoor Path 4 : $D \longleftarrow F \longrightarrow O \longleftarrow A \longrightarrow Y$

   • Controlling for Female after controlling for ability doesn't add close path rather produces the same as Backdoor Path 2.

   • We can control for occupation if we can observed and control for ability as it satisfy the backdoor criterion.

   • Controlling both occupation and female closes the path satisfy backdoor criterion (We can see this through Backdoor Path 4).

3. In the case where ability is unobservable: Occupation is a collider , so controlling for occupation opens a new path $D \longleftarrow F \longrightarrow A \longrightarrow Y$. But if we can control for ability then this can lead to close path (Assuming that there is no direct relationship between gender and earnings)

Exercise 2.2

(a) • Arrow 1 depicts the main causal relationship we want to study i.e. the effect of connectedness on publication.

• Arrow 2 to Arrow 3 represents the path from unobserved ability to publication through perceived ability. Perceived ability can be position, tenure, institution of the author. For example, the more prestigious the institution author works for is the more perceived ability he has for the world this may affect his publication for particular journal (for given,quality of the paper). Also, perceived ability affects the publication through connectedness (Arrow 4). For example, better the position you are at increases the chance of being connected with the editor and that in turn, affects publication.

• Arrow 5 to Arrow 6 represents another path from unobserved ability to publication through quality of paper.

• Arrow 7 to Arrow 8 represents the last path through which unobserved ability affects connectedness through total number of publication and this affects the publication. The the higher ability authors may have more publications (We used the dashed line because that might not always be the case).The higher the number of publication the probability of collaboration with editor is more which may effect publication in the journal.(Schnider, 2020)

Regression Publication on Connectedness doesn't capture the a causal relationship because of the presence of the confounders like ability. Ability is unobserved and we can't control for it, what we can control for is perceived

ability but it close all the backdoor path.Ability is causually related to publication but we can't capture this relation as ability is unobserved.
A way to capture the relation is to use an instrument but that can't be perceived ability, because, ability affects publication also through paths other than perceived ability and thus the use perceived ability violates exclusion restrictions.

(b)
- Figure 2 represents the simplest version of the DAG, this can be augmented in the many ways.

  i. There is a source of sample selection as we are just considering the papers that have been published in the specific journal (BJE).Sample selection arises because of the fact authors choose papers to submit for the journal(BJE) based on the connection they have with editors.
  In case where the authors randomly submit their papers to all journal ,irrespective of journal characteristics, there will be no selection bias.

  ii. (DAG Manuel)

  iii. Following is the ways of mitigating the issue of sample selection:

  – *Heckman Two-stage Procedure*: To perform the regression, one of the essential assumptions is sample should be random. But in some cases, we only get a subsample that we called a selected sample. For instance, while analyzing the effect of publication on connectedness, the sample we focus on is of papers which authors themselves selected to submit for publication in BJE (that is not random). Thus, the sample is selected based on whether has been decided to be in BJE or not. This led to the result of OLS regression as inconsistent and biased. Heckman Correction is used to correct the bias occurring from non-randomly selected samples.
  Heckman correction uses a variable that identifies the probability that author choose a paper out of total working papers to publish in BJE. Then in second stage we can use these probabilities to study the effect of publication on connectedness.

For more information, refer to `https://github.com/gijs-pennings/latex-homework`.

Table 1: Regression Results

	Dependent variable:		
	Earnings		
	(1)	(2)	(3)
Discrimination	-2.978$^{\cdots}$	0.595$^{\cdots}$	
	(0.085)	(0.030)	
Female			-1.009$^{\cdots}$
			(0.028)
Occupation		1.785$^{\cdots}$	0.986$^{\cdots}$
		(0.006)	(0.010)
Abillity			2.030$^{\cdots}$
			(0.022)
Constant	10.965$^{\cdots}$	-6.857$^{\cdots}$	1.142$^{\cdots}$
	(0.061)	(0.064)	(0.097)
Observations	10,000	10,000	10,000
R^2	0.109	0.906	0.950
Adjusted R^2	0.109	0.906	0.950
Note:			$^{\cdot}$p<0.1; $^{\cdot\cdot}$p<0.05; $^{\cdots}$p<0.01