**Assignment 4 – Potential outcomes and OLS**

**Due date: Thursday, June 11th, 2020 by 5:00pm**

**DIRECTIONS**:

**Potential outcomes and OLS**

1. Consider the simple hypothetical example in Table 1. This example involves eleven patients each of whom is infected with coronavirus. There are two treatments: ventilators and bedrest. Table 1 displays each patient’s potential outcomes in terms of years of post-treatment survival under each treatment.

Table 1: Perfect doctor example

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Patient** | **Y1** | **Y0** | **Age** | **TE** | **D** | **Y** |
| 1 | 1 | 10 | 29 |  |  |  |
| 2 | 1 | 5 | 35 |  |  |  |
| 3 | 1 | 4 | 19 |  |  |  |
| 4 | 5 | 6 | 45 |  |  |  |
| 5 | 5 | 1 | 65 |  |  |  |
| 6 | 6 | 7 | 50 |  |  |  |
| 7 | 7 | 8 | 77 |  |  |  |
| 8 | 7 | 10 | 18 |  |  |  |
| 9 | 8 | 2 | 85 |  |  |  |
| 10 | 9 | 6 | 96 |  |  |  |
| 11 | 10 | 7 | 77 |  |  |  |

* 1. Provide an example of how SUTVA might be violated for treatments of covid-19.
  2. Calculate each unit’s treatment effect (TE).
  3. What is the average treatment effect for ventilators compared to bedrest? Which type of intervention is more effective on average?
  4. Suppose the “perfect doctor” knows each patient’s potential outcomes and as a result chooses the best treatment for each patient. If she assigns each patient to the treatment more beneficial for that patient, which patients will receive ventilators and which will receive bedrest? Fill in the remaining missing columns based on what the perfect doctor chooses.
  5. Calculate the simple difference in outcomes. How similar is it to the ATE?
  6. Calculate the ATT and the ATU. How similar are each of these to the SDO? How similar are each of these to the ATE?
  7. Show that the SDO is numerically equal to the sum of ATE, selection bias and heterogeneous treatment effects bias. You will need to calculate the ATE, selection bias and heterogenous treatment effects bias.

1. The following two questions (h and i) ask you to estimate two regressions. Report your results in a “beautiful table” labeled Table 1 with a simple description based on parts (a) and (b). You may use this opportunity to learn outreg2 or estout. I have provided an example for using estout to do this in the /estout subdirectory on github in a file called ols.do.
   1. Using the dataset in part (1) based only on D, Age and Y, estimate the following equation:

Report the coefficient on . Is it equal to ATE, SDO, ATT or ATU?

* 1. Now run the following multivariate regression controlling for age.

Report the coefficient on . Is it equal to ATE, SDO, ATT or ATU? Did controlling for age recover the ATE?

* 1. Create a separate table labeled Table 2. This table should have three columns. The first equation is the multivariate regression. The second equation is the auxiliary regression of D onto Age. The third equation regresses Y onto which is the residual from the second equation. Compare the coefficient on D from the first equation to the coefficient on in the third equation. What does this tell you about how to interpret multivariate regressions?

**Directed acyclical graphs**

1. A 2005 article published in the Journal of Behavioral Medicine claims that forgiveness improves physical health outcomes.[[1]](#footnote-1) Assume that we are interested in evaluating the average causal effect of forgiveness (D) on health (Y) using observational data. We believe that forgiveness (*D*) causes health (*Y*), but our sample is restricted to patients in treatment (*patients)*. Individuals who are more open towards behavioral therapy in the first place (*openness*)select into treatment (*patients*) andare more likely to forgive (*D*). Wealth causes people to seek treatment (*patients*) because of their higher willingness to pay, and wealth also improves health outcomes, but wealth is not in your data. Wealth is also associated with insurance coverage, which also causes people to be in sample (*patients)* and which affects health outcomes.
   1. Assume that forgiveness is binary and we only have data on pre-existing patients, list all causal and non-causal pathways contained in the following estimator:

E[Y|D=1, *patients=1*] – E[Y | D=0, *patients=1*].

* 1. If you had a random sample of the population and not just pre-existing patients, would your answer to (a) change? Why/why not?

Forgiveness sample collision

Figure 1: Forgiveness-health study.

1. Use Figure 2 for the following questions. In all four DAGs (a-d), X is a binary treatment variable and Y is the outcome variable, U and V are unobservable. S, Z, X and Y are all observable (in your data).

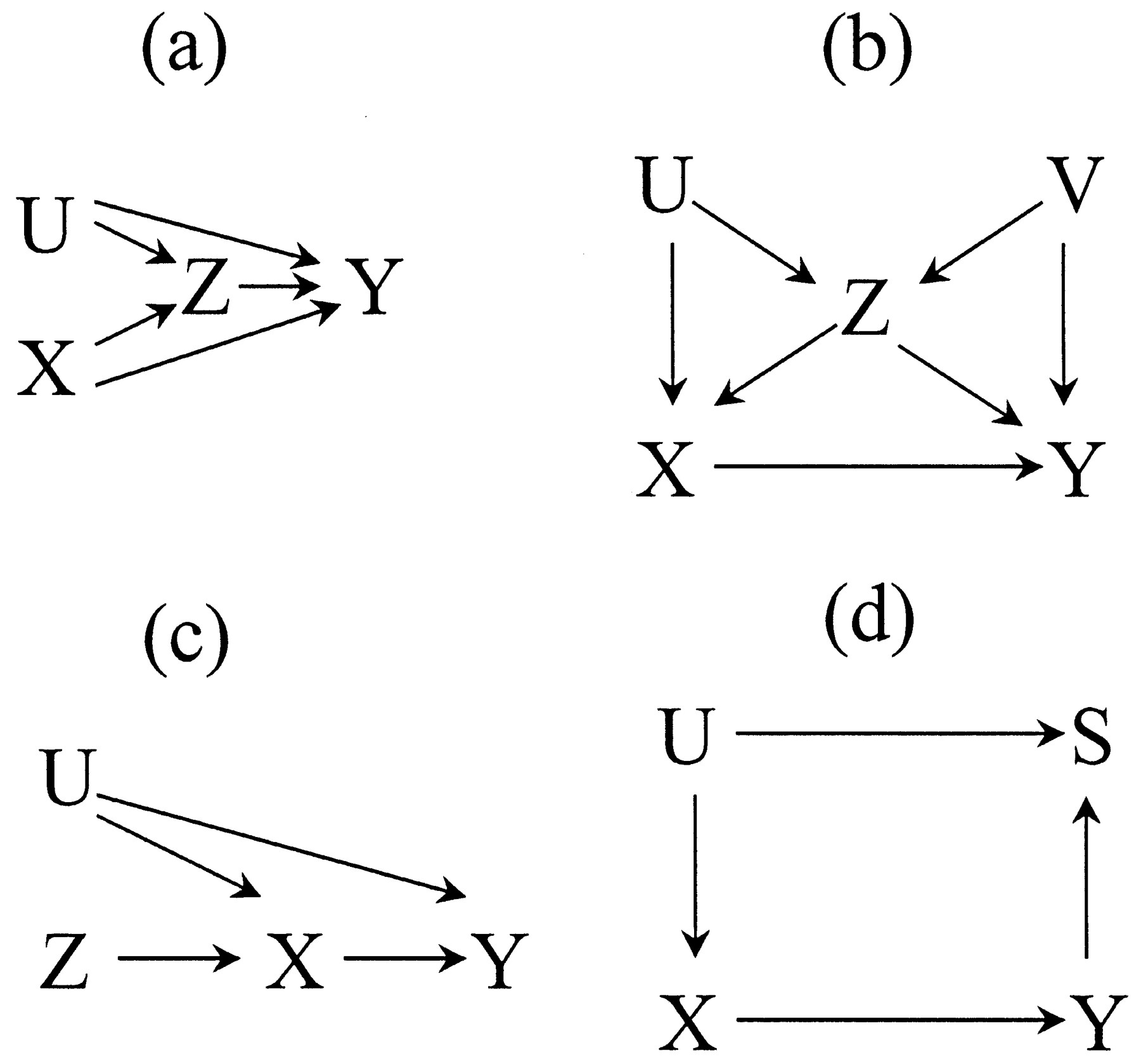
**

Figure 2: Four DAG examples

* 1. If you estimated E[Y | X=1] – E[Y | X=0] for each diagram, what causal and statistical pathways would be contained in that estimator?
  2. For each DAG, write down all backdoor paths from X to Y and indicate whether they are open or closed. Write down a conditioning strategy that satisfies the backdoor criterion. If one does not exist, why does it not exist?

1. Lawler, et al. (2005), “The Unique Effects of Forgiveness on Health: An Exploration of Pathways”, Journal of Behavioral Medicine, vol. 28 (2) April, pp. 157-167. [↑](#footnote-ref-1)