Assignment 3: Potential Outcomes and OLS

Juan Andrés Rincón†

Causal Inference and Research Design
June 2020

1 Potential Outcomes

	Variables					
Patient	Y^1	Y^0	Age	TE	D	Y
1	1	10	29	-9	0	10
2	1	5	35	-4	0	5
3	1	4	19	-3	0	4
4	5	6	45	-1	0	6
5	5	1	65	4	1	5
6	6	7	50	-1	0	7
7	7	8	77	-1	0	8
8	7	10	18	-3	0	10
9	8	2	85	6	1	8
10	9	6	96	3	1	9
11	10	7	77	3	1	10

Table 1.1: Perfect Doctor Example.

a) *Provide an example of how SUTVA might be violated for treatments of covid-19.*

Based on the stable unit-treatment value assumption, in the case of SARS-CoV-2 treatment, the homogeneous dose parameter can be violated conditional on the type of medical facility the patient is being treated. The treatment is $D = 1$ if having a ventilator and $D = 0$ if not (or given bed rest), so perhaps facility A has better ventilators and further care of the patient compared to facility B where they may have older ventilators and attention is not as good, so the dose is not homogeneous across individuals.

b) *Calculate each unit's treatment effect (TE).*

Since the Individual's Treatment Effect is defined as:

$$\delta_i = Y_i^1 - Y_i^0 \ , \tag{1.1}$$

then, Table 1 can be updated with the calculated values.

†Economics student at Universidad de los Andes, Colombia. URosario email: juana.rincon@urosario.edu.co. Uniandes email: ja.rincong@uniandes.edu.co.

c) *What is the average treatment effect for ventilators compared to bedrest? Which type of intervention is more effective on average?*

The Average Treatment Effect is defined as:

$$\begin{aligned} E[\delta_i] &= E[Y_i^1 - Y_i^0] \\ &= E[Y_i^1] - E[Y_i^0] \; . \end{aligned} \tag{1.2}$$

But, since we are working with data, the expectation of the parameter can be approximated with the arithmetic mean, so:

$$E[\delta_i] \approx \frac{1}{n}\left(\sum_{i=1}^{n} Y_i^1 - \sum_{i=1}^{n} Y_i^0\right) \; . \tag{1.3}$$

Where $n = 11$. Using this approach, we have that the average treatment effect is

$$E[\delta_i] \approx -0.\overline{54}.$$

Whereas the respective averages for each treatment are:

$$E[Y_i^1] = 5.\overline{45} \;\; < \;\; 6 = E[Y_i^0].$$

Meaning that, in average, bedrest is more effective than ventilators.

d) *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.*

For this particular case, that the *perfect doctor* knows what treatment is better for every patient, the treatment variable D is defined as:

$$D_i = \begin{cases} 1 & \text{if } \delta_i > 0, \\ 0 & \text{if } \delta_i \leq 0. \end{cases} \tag{1.4}$$

The patients where $D_i = 1$ will receive ventilators and the others will not.[1]

e) *Calculate the simple difference in outcomes. How similar is it to the ATE?*

The Simple Difference in Outcomes (SDO) is defined as:

$$\begin{aligned} SDO &= E[Y^1|D=1] - E[Y^0|D=0] \\ &= E_N[Y|D=1] - E_N[Y|D=0] \; . \end{aligned} \tag{1.5}$$

In arithmetic terms, the SDO can be calculated as:

$$SDO = \frac{1}{n_1}\sum_{i=1}^{n_1}(Y_i^1|D_i = 1) - \frac{1}{n_0}\sum_{i=1}^{n_0}(Y_i^0|D_i = 0) \; . \tag{1.6}$$

[1]Note that in (1.4), in the case of the same treatment effect, bedrest is preferred as it is cheaper.

Where $n = n_1 + n_0$. Thus, for this data-set, $SDO = 0.857$.

Particularly, $SDO > ATE$. This means that the sum of the selection bias and the heterogeneous treatment effect bias is negative.

f) Calculate the ATT and the ATU. How similar are each of these to the SDO? How similar are each of these to the ATE?

Respectively these are defined as:

$$\begin{aligned} ATT = E[\delta|D=1] &= E[Y^1 - Y^0|D=1] \\ &= E[Y^1|D=1] - E[Y^0|D=1] \ , \end{aligned} \tag{1.7}$$

$$\begin{aligned} ATU = E[\delta|D=0] &= E[Y^1 - Y^0|D=0] \\ &= E[Y^1|D=0] - E[Y^0|D=0] \ . \end{aligned} \tag{1.8}$$

And so, the arithmetic calculations are:

$$ATT = \frac{1}{n_1}\sum_{i=1}^{n_1}(\delta_i|D=1) \ , \tag{1.9}$$

$$ATU = \frac{1}{n_0}\sum_{i=1}^{n_0}(\delta_i|D=0) \ . \tag{1.10}$$

Thus, $ATT = 4$ and $ATU = -3.1429$.

Looking at the results, numerically, there is no evident relationship with the SDO, but theoretically, they make up the heterogeneous treatment effect bias, that added to the ATE and the selection bias, should equal the SDO.

g) 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, combine them in the appropriate way, and show that their sum is equivalent to the SDO.

So, theoretically, this expression should hold:

$$\begin{aligned} \underbrace{E[Y^1|D=1] - E[Y^0|D=0]}_{SDO} \quad = \quad & \underbrace{E[Y^1] - E[Y^0]}_{Average\ Treatment\ Effect} \\ & + \underbrace{E[Y^0|D=1] - E[Y^0|D=0]}_{Selection\ Bias} \\ & + \quad \underbrace{(1-\pi)(ATT - ATU)}_{Heterogeneous\ Treatment\ Effect\ Bias} \end{aligned} \tag{1.11}$$

As calculated previously, we have this values:

Parameter	Value
SDO	0.8571429
ATE	−0.5454545
ATT	4
ATU	−3.142857

Thus, the calculation for the Selection bias is:

$$Selection\ Bias = \frac{1}{n_1}\sum_{i=1}^{n_1}(Y_i^0|D=1) - \frac{1}{n_0}\sum_{i=1}^{n_0}(Y_i^0|D=0)\ . \tag{1.12}$$

$$Selection\ Bias = -3.142857.$$

And, for the heterogeneous treatment effect bias is:

$$Het.\ TE\ Bias = (1-\pi)(ATT - ATU)\ . \tag{1.13}$$

Since π is the proportion of treated ($D = 1$), $pi = 0.3636364$, and $ATT - ATU = 7.142857$.

Now, replacing every parameter in (1.11), then

$$\begin{aligned} SDO &= 0.8571429 \\ SDO &= (-0.5454545) + (-3.142857) + (1 - 0.3636364)(7.142857) \\ &= 0.8571429\ . \quad \square \end{aligned}$$

Then, the condition holds and it has been numerically proven.

2 Ordinary Least Squares

a) *Create a dataset based on the perfect doctor treatment assignment from part (1). This dataset should only contain D, Age and Y. Then estimate the following equation:*

$$Y_i = \alpha + \delta D + \varepsilon_i.$$

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

Running the simple linear regression model, the results are:

$$Y_i = (7.1429) + (\mathbf{0.8571})D_i.$$

Where $\hat{\delta} = 0.8571$, which is considerably different from the theoretical δ from the complete data-set. Moreover, statistically speaking, with 11 observations, the regression has near to no statistical power, and so the coefficient is not statistically different from zero.

Nevertheless, $\hat{\delta} = SDO$, which means that there is indeed a bias term that prevents the estimator from being equal to the Average Treatment Effect. Without controls, the coefficient is overestimating the real parameter due to bias.

b) Now run the following multivariate regression controlling for age.

$$Y_i = \alpha + \delta D_i + \beta Age_i + \varepsilon_i.$$

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

Estimating the multivariate regression controlling by Age, the results are:

$$Y_i = (6.35543) + (\mathbf{0.01419})D_i + (0.02019)Age_i.$$

In this case, having the same problem as before, there is not statistical significance of the estimator, but controlling did change drastically the value of the estimator, although not towards a particular parameter of the ones mentioned.

Yet, regression results are shown in Table 2.1, where D is the δ estimator. As previously mentioned, there is not enough statistical power to make this main coefficient sttistically significative.

	Dependent variable:	
	Y	
	OLS	
	(a)	(b)
D	0.857	0.014
	(1.430)	(2.340)
Age		0.020
		(0.043)
Constant	7.143***	6.355**
	(0.862)	(1.907)
Observations	11	11
R^2	0.038	0.064
Adjusted R^2	−0.068	−0.170
Residual Std. Error	2.282 (df = 9)	2.388 (df = 8)
F Statistic	0.359 (df = 1; 9)	0.274 (df = 2; 8)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2.1: Regression results for models (a) and (b). Note: standard errors are in parentheses. Made in R with the extttstargazer package.

c) 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 $\tilde{D}$ which is the residual from the second equation. Compare the coefficient on D from the first equation to the coefficient on $\tilde{D}$ in the third equation. What does this tell you about how to interpret multivariate regressions?*

$$Y_i = \alpha + \delta D_i + \beta Age_i + \varepsilon_i$$
$$D_i = \beta_0 + \gamma_1 Age_i + \epsilon_i$$
$$Y_i = \alpha + \delta \tilde{D}_i + \epsilon_i$$

Running these equations gives Table 2.2. What should be noted is that the coefficient of model (3) is the same D coefficient as in model (1) since the residuals of model (2) are "cleaned" from the effects of the age variables, thus, running the outcome variable against these residuals is equivalent to running model (1) in terms of the treatment variable.

	Dependent variable:		
	Y	D	Y
		OLS	
	(1)	(2)	(3)
D	0.014		
	(2.340)		
Age	0.020	0.014^{***}	
	(0.043)	(0.004)	
$\tilde{D}$			0.014
			(2.280)
Constant	6.355^{**}	-0.403	7.455^{***}
	(1.907)	(0.236)	(0.702)
Observations	11	11	11
R^2	0.064	0.591	0.00000
Adjusted R^2	-0.170	0.546	-0.111
Residual Std. Error	2.388 (df = 8)	0.340 (df = 9)	2.327 (df = 9)
F Statistic	0.274 (df = 2; 8)	13.004^{***} (df = 1; 9)	0.00004 (df = 1; 9)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2.2: Regression results for the linear model equations in (c).

Although the coefficients are the same, the intercept and the standard errors change. This is because the constant term in a simple linear regression depends on the means of the dependent and independent variable and the value of the estimator. In terms of efficiency, the standard error for the last regression is smaller than the first model, making it more difficult to eventually reject the null hypothesis and taking a toll on the efficiency of the estimator.

3 Directed Acyclical Graphs

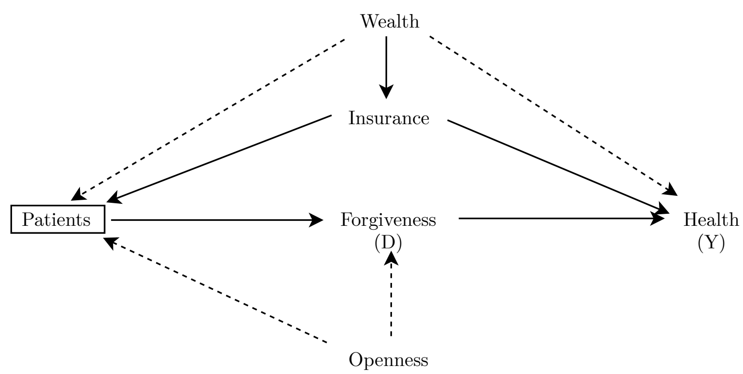

Figure 3.1: Forgiveness-health study DAG.

3.1 Forgiveness on Health

a) *Write down all backdoor paths between D and Y. Mark whether they are open or closed.*

Since a *backdoor path* is any path between the treatment, D, and the outcome, Y, these are the paths:

#	*Backdoor Path*	*Status*
1	Forgiveness $\leftarrow \boxed{Patients} \leftarrow$ Insurance $\rightarrow$ Health	Closed
2	Forgiveness $\leftarrow \boxed{Patients} \leftarrow$ Wealth $\rightarrow$ Insurance $\rightarrow$ Health	Closed
3	Forgiveness $\leftarrow \boxed{Patients} \leftarrow$ Wealth $\rightarrow$ Health	Closed
4	Forgiveness $\leftarrow \boxed{Patients} \leftarrow$ Insurance $\leftarrow$ Wealth $\rightarrow$ Health	Closed
5	Forgiveness $\leftarrow$ Openness $\rightarrow \boxed{Patients} \leftarrow$ Insurance $\rightarrow$ Health	Opened
6	Forgiveness $\leftarrow$ Openness $\rightarrow \boxed{Patients} \leftarrow$ Insurance $\leftarrow$ Wealth $\rightarrow$ Health	Opened
7	Forgiveness $\leftarrow$ Openness $\rightarrow \boxed{Patients} \leftarrow$ Wealth $\rightarrow$ Insurance $\rightarrow$ Health	Opened
8	Forgiveness $\leftarrow$ Openness $\rightarrow \boxed{Patients} \leftarrow$ Wealth $\rightarrow$ Health	Opened

Table 3.1: Backdoor paths of Figure 3.1.

The backdoor paths from 1 to 4 are non-colliders that have been conditioned on the observed variable Patients, then the paths are blocked and the Backdoor criterion is satisfied.

Although, for the paths from 5 to 8, the conditioning on Patients opens new paths, spurius correlations.

b) *What identification strategy would allow you to estimate the causal effect of forgiveness on health? Assume you aren't limited to merely data on patients.*

With the same available variables an identification strategy could be conditioning Insurance instead of Patients. Since Insurance is never a collider, conditioning it closes all backdoor paths.

In other case, if Wealth were observable, it could as well be conditioned.

c) *Now assume you only have data on patients. Assume that forgiveness is binary and you calculate the following simple difference in outcomes:*

$$Y = \alpha + \delta D + \gamma Insurance + \varepsilon$$

But in this regression, you only use data that you have on patients. Will your estimate of δ *identify the ATE? Why/why not? Your answer should indicate whether this control strategy opened up in any backdoors or closed any backdoors.*

Since D is binary, taking the value of 1 and 0, the expected values of estimation should correspond to:

$$E[Y|D=1] = \alpha + \delta + \gamma Insurance \ , \tag{3.1}$$
$$E[Y|D=0] = \alpha + \gamma Insurance \ . \tag{3.2}$$

Note that $E[\varepsilon] = 0$ due to OLS assumptions.

So, calculating the simple difference between (3.1) and (3.2) should yield:

$$\begin{aligned} SDO &= E[Y|D=1] - E[Y|D=0] \\ &= \alpha + \delta + \gamma Insurance - (\alpha + \gamma Insurance) \\ &= \delta \ . \end{aligned}$$

Now, taling into account what was mentioned in the last clause *b*), controlling with *Insurance* instead of *Patients* should close all backdoor paths but #3 which doesn't include variable

3.2 Theorical Approach of DAGs

a) *Write down all backdoor paths from X to Y and indicate whether they are open or closed.*

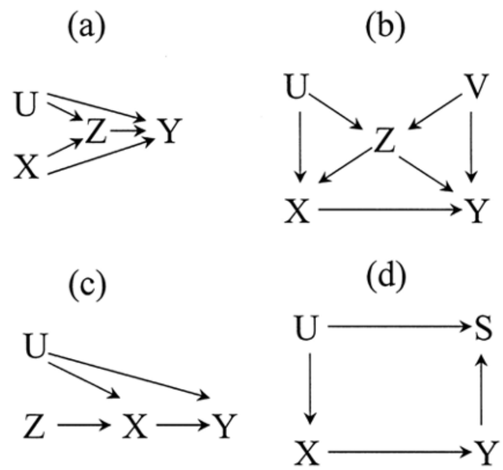

Figure 3.2: Caption

Fig	*#*	*Backdoor Path*	*Status*
(a)	1	X → Z → Y	Open
	2	X → Z ← U → Y	Closed
(b)	1	X ← U → Z → Y	Open
	2	X ← U → Z ← V → Y	Closed
	3	X ← Z → Y	Open
	4	X ← Z ← V → Y	Open
(c)	1	X ← U → Y	Open
(d)	1	X ← U → S ← Y	Closed

Table 3.2: Backdoor paths and states of Figure 3.2.

b) *Write down a conditioning strategy that satisfies the backdoor criterion. If one does not exist, what is stopping it?*

Fig	*#*	*Description*	*Solution*
(a)	1	Condition Z. Since it's a noncollider, conditioning blocks the path. But making this creates a new path with a spurius correlation because of (a)2, since that path is already closed.	X $\rightarrow \boxed{Z} \rightarrow$ Y
(b)	1	Condition Z. Since it's a noncollider, conditioning blocks the path.	X $\leftarrow$ U $\rightarrow \boxed{Z} \rightarrow$ Y
	3	Condition Z. Since it's a noncollider, conditioning blocks the path.	X $\leftarrow \boxed{Z} \rightarrow$ Y
	4	Condition Z. Since it's a noncollider, conditioning blocks the path. The problem with all these is struggling with (b)2, where conditioning Z would condition a collider and creating a spurius correlation.	X $\leftarrow \boxed{Z} \leftarrow$ V $\rightarrow$ Y
(c)	1	U is unobserved, then theoretically conditioning it could solve the path, but it cannot be conditioned.	

Table 3.3: Identification strategies to block backdoor paths based on Table 3.2.