Exercise: Comparative Case Studies & Process Tracing

Let's simulate some fake data and see whether we are able to recover the correct treatment effect using comparative case study methods. Our research question is whether development causes democratization.

1. First, let's generate some confounder variables for 100 countries (we have only cross-section data).

(a) The variable 'Inequality' should be drawn randomly from the normal distribution with mean 0.5 and standard deviation 0.05.
(b) The variable 'Economic Growth' should be drawn randomly from the normal distribution with mean 3 and standard deviation 1.
(c) The variable 'Continent' should be randomly drawn with equal probability from four possibilities: Europe, Americas, Asia, and Africa.

2. Our outcome is going to be a binary variable for whether the country is a democracy or not. Use the expressions below to simulate potential outcomes, with a treatment effect which boosts the chances of democratization by 50% (so our treatment effect is 0.5). (`Europe` is an indicator for whether the country is in Europe).

$$y_0 = \begin{cases} 1 \text{ if } 2 * inequality - \frac{growth}{3} + Europe > 0 \\ 0 \text{ else} \end{cases}$$

$$y_1 = \begin{cases} 1 \text{ if } y_0 = 1 \\ 1 \text{ if } y_0 = 0 \ \& \ Binom(0.5) = 1 \\ 0 \text{ else} \end{cases}$$

3. Treatment D is the level of income of the country. It depends on the confounding variables and we will simplify to High versus Low income based on whether the country has an income over USD10,000. Imagine we know the treatment assignment mechanism so that binary (1/0) treatment is determined by the following expression:

$$D = \begin{cases} 1 \text{ if } N(10000, 2000) + 2000 * Europe - 1000 * inequality - growth * 500 > 10000 \\ 0 \text{ else} \end{cases}$$

4. Calculate observed outcomes based on potential outcomes and treatment.

5. As always, as a benchmark, let's run the 'naive' regression of the outcome on the treatment with no controls.

6. Now we will implement a comparative case study methodology for just two units. Select one treated (high income) unit that is a democracy. What are the values of the three confounding variables, inequality, growth and country?

7. From all of the control (low income) units that are NOT democracies, pick one at random.

(a) Is it possible that inequality explains why your treated unit is a democracy and the control unit is not?
(b) Is it possible that economic growth explains why your treated unit is a democracy and the control unit is not?
(c) Is it possible that being European explains why your treated unit is a democracy and the control unit is not?
(d) Can we conclude anything about the role of income level (treatment) on democrcay?

8. Identify the control unit that has values of the confounders that ensure you can answer "No" to questions (a)-(c) in Question 7. It may not be possible - if not, try picking a different treated unit and then search for an appropriate control unit.

9. Once you succeed in finding an appropriate pairing of treated and control units, what can you conclude about the role of development in explaining democracy?