DID Homework

Jennifer Hill, Ray Lu & Zarni Htet

## Objective

The goal of this exercise is to simulate data that may or may not satisfy the assumptions of a difference in differences design.

##### PART A: DID

The setting here is similar to the last assignment (RDD).

You will simulate hypothetical data collected on women who gave birth at any one of several hospitals in disadvantaged neighborhoods in New York City in 2010. This time we are envisioning a government policy that makes available job training for teen mothers. This program is only available for women in households with income below $20,000 at the time they gave birth. The general question of interest is whether this program increases annual income 3 years later. You will generate data for a sample of 1000 individuals. For this assignment we will make the unrealistic assumption that everyone who is eligible for the program participates and no one participates who is not eligible.

#### Question 1. God role: simulate income.

Simulate the “assignment variable” (sometimes referred to as the “running variable”, “forcing variable”, or “rating”), income, in units of thousands of dollars. Use the following model:

Then plot using a histogram.

#### Question 2. Policy maker role: Assign eligibility indicator.

Create an indicator for program eligibility for this sample. Call this variable “eligible”. (You can use the same code as the previous assignment.)

#### Question 3: God role.

For question 3 you will simulate income at 3 years post treatment in thousands of dollars. You will assume linear models for both E[Y(0) | X] and E[Y(1) | X]. The *expected* treatment effect for everyone should be 4 (in other words, E[Y(1) - Y(0) | X] should be 4 at all levels of X). The residual standard deviation of each potential outcome should be *2*.  
a) You will simulate using the following model

$$
Y(0) \sim N(6 + .3\*\text{income}\_{\text{pre}}, 2^2)\\
Y(1) \sim N(6 + .3\*\text{income}\_{\text{pre}}+4, 2^2)
$$

b) You will save two datasets:  
(1) fullA should have the forcing variable and both potential outcomes (2) obsA should have the forcing variable, the eligibility variable, and the observed outcome.

#### Question 4. Researcher and god role. Plot your data!

Make a scatter plots of pre-treatment income (x-axis) versus observed post-treatment income (y-axis). Plot eligible participants in red and non-eligible participants in blue.

Now plot the full response surface.

Now make a plot like the ones in the DID lecture with time on the x axis and income on the y axis. Plot observed means as open circles; blue corresponds to the control and red corresponds to the treatment. In the same figure plot

using different red symbol (square, triangle, filled in circle).

Based on what you know about the DID assumptions, will the DID estimate of E[Y(1)-Y(0) | Z=0] be close to the truth? Why or why not?

#### Question 5. Researcher role. Estimate the treatment effect using all the data using two approaches:

1. a regression discontinuity approach (linear model only)
2. a DID approach

#### Question 6. Researcher and god roles. Thinking about the assumptions?

Can you think of a way of altering the simulation setup so that the DID assumptions would hold? Make this change, rerun (3) and (5) above, and comment on what you find.