

ECON 453
In-Class Exercise 8
November 2, 2023

Please download the file “IC8.gdt”, a gretl data file. This dataset comes from the 2000 United States Census. During this exercise, we will replicate the methodology used in Dynarski (2008). The dataset includes information for people born in Alabama, Florida, Georgia, South Carolina, and Texas that were between the ages of 22 and 34 when the 2000 Census was conducted. The study that we are replicating uses a difference-in-difference methodology and analyzes the impacts of the Georgia Hope Scholarship program, which began for students that graduated high school in 1993.

1. Let's begin by conducting a descriptive analysis.
 - a. Make a dummy variable for those born in Georgia (call it **georgia**).
 - b. Make a dummy variable for those *not* born in Georgia (call it **control**).
 - c. Make a dummy variable for those that graduated high school in 1993 or more recently (call it **after**). We assume everyone graduates high school at age 18.
 - d. Make a dummy variable for those that graduated high school before 1993 (call it **before**).
 - e. Make a dummy variable for those that have *at least an associate degree* (call it **college**).
 - f. Use sample restrictions to fill out the simple difference-in-difference table below. For each of the cells, find the percentage that have at least an associate degree (using summary statistics for your **college** variable).

	18 before 1993	18 after 1993	Difference
Born in Georgia			
Not born in Georgia			
Difference	N/A	N/A	

- g. Discuss the what the results indicate about the potential impact of the scholarship program on the level of college attainment in Georgia.
2. Next, we want to formally estimate the effect with the regression form of the difference-in-difference model.
 - a. Create an interaction term between your **georgia** and **after** variables.
 - b. Run a regression using **college** as the dependent variable and **georgia**, **after**, and the **interaction** as the regressors.
 - c. Discuss what each of the coefficients tells us:

Georgia:

After:

Interaction

- d. Are the results different from the simple analysis above? Does the regression improve our study?

3. Run the model again, but this time add the following regressors: **female**, **black**, **Hispanic**, and **age**.
 - a. Comment on what the coefficients on these new control variables tell us
 - b. Comment on what happened to the estimated impact of the program once we added the additional control variables
4. Create a new dummy variable (called **bach**) indicating whether an individual earned at least a bachelor's degree. Run the same regression as question 3, but use this new dummy variable as our dependent variable.
 - a. What is the estimated effect of the program if we analyze this outcome? Is this a better outcome to look at?
5. Let's investigate whether the program had heterogeneous impacts on different groups. For each of the following groups, restrict the sample and run the simplified version of the model from question 2 (no control variables). Report the coefficient on the interaction term in each case. Use **college** as your dependent variable.

Female:

Male:

Black:

Hispanic:

White:

Comment on what we learn about the effects of the program on different groups:

6. The control group in this dataset includes people born in Alabama, Florida, South Carolina, and Texas. Choose which of these you believe might be the best comparison to Georgia and restrict the sample to compare just the two states. Run the same specification as in question 2.
 - a. What is the estimated impact of the program, and how does this compare to the effect we found using the broader group of control states?