**Assignment 5**

**Due: Friday, April 5th, 2024 11:59pm**

Please upload your assignment to Canvas under the Assignment 5 folder.

For all questions requiring you to use Stata or R, please include your output as part of your homework. You can highlight Stata output, right click, and select “copy as picture” to easily copy and paste Stata output into Word.

*Content*

The purpose of this exercise is to practice interpreting fixed effect regressions and difference-in-differences estimators.

*Collaboration*

Discuss the problem set and work on Stata code with your study group. Write up the content of the problem set alone, in your own words. Do not exchange answers with your study group. Do not work on a shared file.

**Part I: Estimating and interpreting fixed effects**

Approximately 40%-50% of driving fatalities occur in accidents in which at least one of the drivers involved has been drinking. There are approximately 40,000 highway traffic fatalities annually in the United States, so perhaps 20,000 of these deaths are alcohol-related. In the U.S., highway laws are under the jurisdiction of the states, and states have adopted different strategies to address the problem of drunk driving. In this problem set, you will use data (drunkdri.dta) on traffic death rates by state for 1982-1988 to evaluate the effectiveness of some of these laws.

**Output**

This assignment has you run a total of six regressions. You should put the results from all of your regressions into a single table (which you can create in excel or word or another spreadsheet software, see Table 7.1 in the S&W book for suggested format). The table does not need to display the estimated coefficients for any of the state and year fixed effects, but it should note which specifications include year and/or state fixed effects and should display coefficient estimates and standard errors for all of the other covariates. Each regression output should be presented in a different column within the table.

1. Use data for 1988 only
   1. Produce a scatterplot of the vehicle fatality rate against the beer tax.
   2. [Regression 1, Raw correlation]: Regress the vehicle fatality rate against the beer tax and the minimum legal drinking age. Interpret the coefficients on the beer tax and minimum legal drinking age.
   3. [Regression 2, control for “stuff”]: Add to the previous regression the percent of the population living in dry counties, average miles driven, the percent of drivers aged 15-24, and the per capita income in the state. This set of variables will be referred to for the rest of the problem set as the “state control variables.”
   4. Do the coefficients on the beer tax and drinking age change substantively from (b) to (c)? If so, explain why.
2. Use the full data set.
   1. [Regression 3, control for stuff]: Regress the vehicle fatality rate against the beer tax, the minimum legal drinking age and the state control variables.
   2. [Regression 4, add year fixed effects]: Add to the previous regression a full set of year dummies. Are the year dummies statistically significantly related to the fatality rate?
   3. Interpret the coefficients on the beer tax and minimum legal drinking age in the previous regression. Do the coefficients on the beer tax and minimum legal drinking age change substantively from 2(a) to 2(b)? Why or why not?
   4. [Regression 5, add state and year fixed effects]: Now, add to the regression in 2(b) a set of state fixed effects. Do the coefficients on the beer tax and minimum legal drinking age change substantively Why or why not?
   5. Write a paragraph in which you discuss whether you can confidently place a causal interpretation on the analysis you have undertaken. In your answer, discuss the biases that could threaten the analysis and how the control variables, year effects and state effects each attempt to deal with these biases.

**PART II: Difference-in-Differences**

For this assignment, you will read the following article and answer the questions below. As you read the article, do not worry about understanding all the technical details. Your goal is to get a big picture understanding of the study.

Akee, R.K., Copeland, W.E., Keeler, G., Angold, A. and Costello, E.J., 2010. Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, *2*(1), pp.86-115.

1. What is the causal question addressed in the study? (1 sentence)
2. What are the comparisons the authors are using in their difference-in-differences estimator (2-3 sentences).
3. What are the main assumptions the authors are making with their difference-in-differences strategy? Give two examples of how these assumptions could be violated in this specific scenario (1 paragraph).
4. What are the main outcome variables and how are they measured? (1-2 sentences)
5. What are the main findings? Make sure you discuss the economic significance (magnitude) as well as the statistical significance. (1 short paragraph)
6. The authors test a couple of different mechanisms through which the casino opening may have affected children’s long-run outcomes. Which of these mechanisms do you think is most plausible, and why? (2-3 sentences)
7. Briefly describe what you view as the key limitations or shortcomings of the study, and why. (1 paragraph)

## Stata Hints

*Analysis using only a subset of the data.*

The simplest way to restrict an analysis to a subset of the data is to use the restriction clause **if** in a STATA command. For example, to run a regression of vfrall on beertax for data from 1982 only, the command is

**regress vfrall beertax if (year==1982), robust**

The command executes the regression, using only the observations for which the statement in parentheses is true (the double equal sign denotes logical equality).

*Drinking age dummies*

The minimum drinking ages are *not* all integers. You should therefore define the drinking-age dummy variables as below.

gen da18=(mlda<19)

gen da19=(mlda>=19 & mlda<20)

gen da20=(mlda>=20 & mlda<21)

*Regressions with many dummy variables*

The most straightforward way to execute a regression with multiple dummy variables is to create them using **generate** statements, then include them in your regression. This is easy when there are only a handful of dummy variables, such as the year dummy variables in this exercise. However, creating all these variables can be quite tedious and can introduce typos and, in some situations, computer memory problems. Therefore, STATA provides a simple method for running fixed effect regressions in this circumstance, using the command **areg**. The syntax is:

**areg y x1 x2, absorb(x3)**

This command runs a regression of the variable y against x1 and x2, including dummy variables (fixed effects) that correspond to the variable x3. Note that x3 cannot be a continuous variable. In the context of this problem set, x3 would be the variable “state” (the state identification codes). The results from using **areg** are identical to those if you were to include a full set of state dummy variables, except that **areg** does not print out the coefficients on the state dummies. Verify this by running one of the dummy variable regressions in this exercise both ways – using **areg** and by explicitly creating and including the dummies.

*F-tests in stata*

The command “test” used after a regression lets you calculate an f-test of whether a combination of variables are significant. For instance, typing **test var1 var2 var3** conducts an F-test where the null hypothesis is that var1=0 and var2=0 and var3=0, with the alternative hypothesis that at least one of those variables is not equal to zero.

**Drunk Driving Data**

These data cover the 48 contiguous U.S. states for the years 1982 to 1988. Additional series are contained in the data set beyond those listed below (brief definitions have been incorporated into the .dta file). Traffic fatality data are from the U.S. Department of Transportation Fatal Accident Reporting System (FARS). This data set was graciously provided by Prof. Christopher Ruhm, University of Virginia.

vfrall = vehicle fatality rate (deaths per 10,000)

state = state ID code

year = year (1982, 1983,...,1988)

beertax = tax on case of beer (1988 dollars)

mlda = minimum legal drinking age

dry = percent of residents residing in “dry” counties, which do not allow sale of alcohol

yngdrv = percent of drivers aged 15-24

vmiles = average miles driven per driver

perinc = per capita personal income