Problem Set 5

Francisco Brady

5 Apr 2024

### 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.

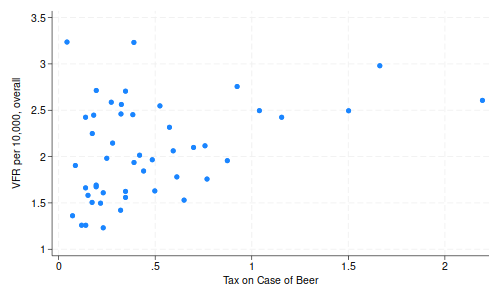
### Output

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*:

##### a) Produce a scatterplot of the vehicle fatality rate against the beer tax.

. use drunkdri.dta, clear  
  
. twoway scatter vfrall beertax if year == 1988  
  
. graph export vfr\_beer.png, width(500) replace  
file vfr\_beer.png saved as PNG format



VFR and Beer Tax

. qui reg vfrall beertax mlda if year == 1988, robust  
  
. eststo raw, title(Reg. 1)  
  
. qui reg vfrall beertax mlda dry vmiles yngdrv perinc if year == 1988, robust  
  
. eststo stuff, title(Reg. 2)  
  
. estout raw stuff, cells(b (star fmt(3)) se(par fmt(2))) starlevels(\* 0.10 \*\* 0.05 \*\*\* 0.01) le  
> gend stats(r2\_a r2 N, labels("Adj. R-Squared" "R-Squared" "Observations")) label collabels("")  
>   
  
────────────────────────────────────────────────────  
 Reg. 1 Reg. 2   
   
────────────────────────────────────────────────────  
Tax on Case of Beer 0.508\*\*\* 0.141   
 (0.12) (0.14)   
Minimum Legal Drin~e -0.945\*\*\* -0.345\*\*   
 (0.07) (0.13)   
% Residing in Dry ~s 0.004   
 (0.00)   
Ave. Mile per Driver 0.000\*\*\*  
 (0.00)   
% of Drivers Aged~24 0.869   
 (2.78)   
Per Capita Persona~e -0.000\*\*\*  
 (0.00)   
\_cons 21.645\*\*\* 8.454\*\*   
 (1.27) (3.16)   
────────────────────────────────────────────────────  
Adj. R-Squared 0.253 0.546   
R-Squared 0.285 0.604   
Observations 48.000 48.000   
────────────────────────────────────────────────────  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.01  
  
.

##### b) [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.

The coefficient on beertax is positive and significant, implying that holding the legal drinking age constant, a one dollar increase in the tax on beer is associated with an increase of .51 in the Vehicle Fatality Rate per 10,000. [Range is from .82 to 4.21, sd .57, so this is big-ish].  
The coefficient on the minimum legal drinking age is negative and significant. This implies that increasing the minimum legal drinking age by one year is associated with a .95 decrease in the Vehicle Fatality Rate per 10,000, holding beer taxes constant.

##### c) [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.”

##### d) Do the coefficients on the beer tax and drinking age change substantively from (b) to (c)? If so, explain why.

Yes. The coefficient on beer tax and drinking age both decrease substantively.  
The coefficient on beer taxes goes from 0.508 (significant at the 1% level) to 0.141 and loses significance at any conventional level. This suggests that the first regression suffered from omitted variable bias, and that including the state control variables helped to account for some of the differences in Vehicle Fatality Rates (VFR) that were not captured previously.  
The coefficient on minimum legal drinking age also decreases in magnitude from -0.945 to -0.345, however it is still significant (at the 5% level). This suggests again that including the state control variables helps to account for omitted variable bias, and that the naive regression was over-attributing some of the effects to minimum legal drinking age.

#### 2) *Use the full dataset:*

. gen y83 = (year == 1983)   
  
. gen y84 = (year == 1984)  
  
. gen y85 = (year == 1985)  
  
. gen y86 = (year == 1986)  
  
. gen y87 = (year == 1987)  
  
. gen y88 = (year == 1988)  
  
.   
. qui reg vfrall beertax mlda dry vmiles yngdrv perinc, robust  
  
. eststo col1, title(Reg. 3)  
  
. qui reg vfrall beertax mlda dry vmiles yngdrv perinc y83-y88, robust  
  
. eststo col2, title(Reg. 4)  
  
. qui reg vfrall beertax mlda dry vmiles yngdrv perinc y83-y88, absorb(state) robust  
  
. eststo col3, title(Reg. 5)  
  
. estout col1 col2 col3, cells(b (star fmt(3)) se(par fmt(2))) starlevels(\* 0.10 \*\* 0.05 \*\*\* 0.0  
> 1) legend stats(r2\_a r2 N, labels("Adj. R-Squared" "R-Squared" "Observations")) label collabel  
> s("")   
  
────────────────────────────────────────────────────────────────────  
 Reg. 3 Reg. 4 Reg. 5   
   
────────────────────────────────────────────────────────────────────  
Tax on Case of Beer 0.095 0.083 -0.571\*\*   
 (0.07) (0.06) (0.26)   
Minimum Legal Drin~e -0.008 -0.028 0.004   
 (0.03) (0.03) (0.02)   
% Residing in Dry ~s -0.000 -0.000 0.020\*\*   
 (0.00) (0.00) (0.01)   
Ave. Mile per Driver 0.000\* 0.000 0.000   
 (0.00) (0.00) (0.00)   
% of Drivers Aged~24 0.650 2.043 0.731   
 (1.31) (1.24) (1.03)   
Per Capita Persona~e -0.000\*\*\* -0.000\*\*\* 0.000\*\*\*  
 (0.00) (0.00) (0.00)   
y83 -0.074 -0.092\*\*   
 (0.11) (0.05)   
y84 -0.068 -0.153\*\*\*  
 (0.11) (0.05)   
y85 -0.036 -0.233\*\*\*  
 (0.12) (0.06)   
y86 0.085 -0.196\*\*   
 (0.14) (0.08)   
y87 0.106 -0.260\*\*\*  
 (0.16) (0.10)   
y88 0.135 -0.306\*\*\*  
 (0.18) (0.11)   
\_cons 2.445\*\*\* 2.680\*\*\* 0.103   
 (0.78) (0.80) (0.62)   
────────────────────────────────────────────────────────────────────  
Adj. R-Squared 0.372 0.374 0.909   
R-Squared 0.383 0.396 0.925   
Observations 336.000 336.000 336.000   
────────────────────────────────────────────────────────────────────  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.01  
  
.

##### a) [Regression 3, control for stuff]: Regress the vehicle fatality rate against the beer tax, the minimum legal drinking age and the state control variables.

##### b) [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?

When adding year fixed effects to the regression with state controls, none of the year dummies are statistically significant. To confirm this, we can run an F test for the year dummies.

. qui reg vfrall beertax mlda dry vmiles yngdrv perinc y83-y88, robust  
  
. test y83 y84 y85 y86 y87 y88  
  
 ( 1) y83 = 0  
 ( 2) y84 = 0  
 ( 3) y85 = 0  
 ( 4) y86 = 0  
 ( 5) y87 = 0  
 ( 6) y88 = 0  
  
 F( 6, 323) = 0.64  
 Prob > F = 0.6960

The output of the F test on years is not significant at any level, which means that we fail to reject the null hypothesis that all of the coefficients on years are zero.

##### c) 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?

The coefficient on beer tax in the full sample regression is 0.095 and decreases to 0.083 in the year fixed effects model. Because the estimate is not significant, it implies that beer taxes do not affect the VFR in this model, holding all else constant.

The coefficient on minimum legal drinking age in the full sample regression is -0.008, and is not significant. This increases to -0.028 in the year fixed effects model, but is also not significant. This implies that the minimum legal drinking age does not affect the VFR, holding all else constant.  
Overall, the changes from adding year dummies are small and do not affect the significance of the coefficients.

##### d) [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?

Yes, the coefficients on beer tax and minimum legal drinking age change substantively. For the beer tax, the coefficient increases in magnitude, and changes direction, from 0.083 in the year fixed effects model to -0.571 in the state-year fixed effects model. The coefficient is also significant at the 5% level, compared to not significant in the previous model. This new estimate suggests that a dollar increase in the beet tax is associated with a 0.571 decrease in the VFR. The change in direction and increase in significance implies that accounting for state-level factors should be included in the model.  
For minimum legal drinking age, the coefficient increases from -0.027 to 0.004, however it is still not significant. This implies that even after accounting for state-level fixed effects, the minimum legal drinking age does not have a significant impact on the VFR.

##### e) 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.

I don’t think I would be confident in a causal interpretation for this analysis. I could be confident about claiming a strong association between VFR and the beer tax, but there are other things that could potentially impact the vehicle fatality rate. This is the reason for including the state control variables (% of dry counties, average miles driven, % of young drivers, and income). Leaving these out, biased the coefficients on both beer tax and driving age upwards, overestimating their impacts.  
The inclusion of year fixed effects is to account for any time trends that could affect all of the states in the analysis, like a recession or new national law related to driving.  
Including state fixed effects attempts to control for any differences across states, and including them produced significant changes magnitude and direction of the coefficients.

### 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)

The study tries to estimate the effectiveness of cash transfer programs on a number of outcomes for children in treated households.

#### 2. What are the comparisons the authors are using in their difference-in-differences estimator (2-3 sentences).

The pre- and post groups in this study are the cohorts of children in households with at least one American Indian parent that received cash transfers, with the younger children (age 9 or age 11 cohort) acting as the after-treatment group (post), and the older children (age 13 cohort) in the household serving as the before-treatment group (post). The comparison group is children in households with no American Indian parent in the household that were not eligible for cash transfers, both before-treatment and after-treatment.

#### 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).

Parallel Trends: The authors assume that macroeconomic conditions are similar across the groups being differenced. To address this, they show that labor market conditions are not statistically different across Native and non-Native households. An example of how this could be violated is if there were employment opportunities that only tribal members were eligible for, which could translate into better child outcomes for both treated groups, in excess of the cash transfer payments.  
Confounding Factors: The second assumption is that there are no exogenous policy changes that occur over the period of study that could impact the outcomes. An example of how this could be violated is if a crime prevention program was instituted within the tribal jurisdiction at the outset of the cash transfer program.

#### 4. What are the main outcome variables and how are they measured? (1-2 sentences)

1. Educational attainment, measured as length of schooling, and probability of graduating high school.
2. Justice involvement, measured from survey responses of drug dealing by the children in the households, as well as arrest data.

#### 5. What are the main findings? Make sure you discuss the economic significance (magnitude) as well as the statistical significance. (1 short paragraph)

The authors find that an exogenous increase in income increases the measured years of schooling at age 21, but this interaction for the first cohort (age 9) is not significant. The probit models estimated a statistically significant interaction for the first cohort, implying that the cash transfers are increased the probability of being a high school graduate by age 19 for the treated households by almost 15 percent. Additionally, the authors find that for households previously in poverty, the interaction term for the first cohort, exposed to the treatment for longer was positive and significant at the 5% level, indicating that the treated households on average increase education by an additional year.  
In the crime regression, the probit regression estimates imply that the cash transfers are associated with a 22% decrease in the probability of having commited a crime by age 16-17. The final regression uses self-reported drug dealing, with the estimates showing that the cash transfers are associated with a decrease in that probability of 6.5%.

#### 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)

The authors mention the possibility that the extra income could be used to purchase better educational materials, but are not able to measure this. In my opinion, the most plausible mechanism is the potential for the extra income to decrease the probability of parental arrest. Parent arrest can have lasting effects on children, [elevating stress levels](https://www.irp.wisc.edu/resource/how-witnessing-a-parents-arrest-can-get-under-the-skin-of-a-child/), which may lead to poor behavioral outcomes.

#### 7. Briefly describe what you view as the key limitations or shortcomings of the study, and why. (1 paragraph)

The study may have limited external validity because of the distinct position of Native American populations in America. It’s possible that the baseline characteristics of the population are much different than the general population, including rates of poverty, over-policing, and other social determinants of both criminal justice involvement and educational attainment. This fact may limit the generalizability of the results. In addition, as the authors note, they are unable to observe educational expenses, which could directly impact one of their main outcomes (educational attainment). Some of the observations used to measure parental quality, such as positive interactions with children, are measured using survey response data, which might be subject to response bias. Finally, while the results show more modest results on the older treated cohort (age 11), the authors do not mention the impacts of any spillover effects on the younger cohorts.