Exercise: Texting Drivers

In order to reduce the number of fatal driving accidents caused by sending or receiving text messages many states in the U.S. have banned texting while driving. (Notice that this restriction is not the same as forbidding the use of any type of handheld device while driving, it only bans text messaging while driving.)

Goal: Your goal is to determine whether there was a reduction in fatal accidents following the state bans on text messaging.

Data: The data in 'txtbans.RData' comes from the Fatality Analysis Reporting System (FARS) of the NHTSA, an American census for all motor vehicle crash fatalities. It includes monthly information on fatal accidents. This data is merged with data on the enactment of text message bans and some state demographics.

The variables in the dataset are:

state: state id
time: month id
after: indicator variable (=1 if after treatment)
treated: indicator variable (=1 if text ban was implemented in that state)
laccsvobyinhab: log of (accsvobyinhab+1)
txmsban: indicator variable (=1 if text message ban is in place)
callban: indicator variable (=1 if cell-phone calls ban is in place)
bantime: id of the month when the ban was implemented
pop: state population
lpop: log of state population
unemp: state unemployment rate
lunemp: log of state unemployment rate
permale: percentage of males in the state
rgastax: gas (fuel) tax.
lrgastax: log of gas (fuel) tax
accsvobyinhab: number of fatal single-vehicle accidents with a sole occupant crashing into a non vehicular object per 1000 inhabitants.

Questions:

1. Explore the data set. How many months (and hence years) do you observe?
2. First, keep only the observations from April to run a simple 2-period diff in diff.

- Define the control and treatment group and write down your regression model in equation form.
- Create a table with the means of the variables of interest for the treatment and control groups before and after treatment.

- Use the table you created to calculate the difference-in-differences estimator over all periods.
- As seen in class, run a simple difference-in-differences regression model in order to obtain the same estimator as in 2).
- Interpret all the coefficients in the model. Was the text message ban effective in preventing car accidents?
- On top of the standard linear regression model assumptions, what is the additional assumption needed for difference-in-differences to estimate the average treatment effect?

3. For turning to the multi-period Diff-in-Diff, start again from the full dataset:

- Using R's plotting capabilities, explore whether or not the assumption in part 2 is reasonable. Justify your choice of plots.
- For running a multi-period diff in diff, return to the original dataset and create a new variable that measures "months since ban" (negative before the ban, set the "placebo-bantime" for untreated units in the middle of the dataset or (even better) randomly)
- Plot the treated and untreated units, but use time since ban as new time. Do your conclusions regarding the assumption in part 2 change?
- Create period-specific time dummies (1 month to treatment, month of treatment (months to treatment == 0)) , 1 month since treatment, 2 months since treatment, 3 months since treatment, etc.). Moreover create crossterms that capture month-specific treatment effects "(treated)*(1 month since treatment), (treated)*(2 months since treatment), (treated)*(3 months since treatment),... – **Write down the corresponding regression model.**
- Now estimate a multi-period difference-in-differences. Compare the estimated effects with the figure you plotted.
- Give examples of what could be credible violations of the assumption in the previous part that could help explain the results observed.

4. The variable *callban*: indicates whether a state has a law in place forbidding cell-phone calls while driving. What percentage of the states that implemented the text ban also had a call ban in place? How can the existence of a ban on cell-phone calls affect the enforcement of the text ban?

5. (*Hard (Bonus) Question*) There is a function called "linearHypothesis" in the package "car." Find out on your own how this function works. Try to test the assumption in part 2 using this function. *Hint:*: The function is similar to running the model that interacted the treated variable and the time period dummies. Think about the interpretation of these coefficients.