

# hw9

Zeren Li

11/6/2019

**This problem set is due at 8:30 am on 11/14**

**Please upload both Rmd and PDF files on Sakai**

**Do not show the code in the pdf, show outputs, and write-up only**

**Total points: 10**

## Data management

Load `moms.dta` as `moms`, `dads.dta` as `dads`, `faminc` as `faminc`, `kids3.dta` as `kids3` 1. Append two datasets, stacking them one atop the other, show the entire data to see if this works correctly. Next, show the average income of each family.

2. In the appended dataset, create a dummy variable called `momdad`, show the average income of moms and dads using `group_by()`.
3. Combine the `dads` with the `faminc` by variable `famid`, having the dads information and the family information side by side.
4. Remove family 1 in `dads`, remove family 3 in `faminc`, merge `dads` and `faminc` in the following ways:  
4.1 return all rows from `faminc`, and all columns from `faminc` and `dads`. 4.2 return all rows from `dads`, and all columns from `dads` and `faminc`. 4.3 return all rows and all columns from both `dads` and `faminc`.
5. One-to-many match merging: merge `dads` with `kids3`. To be noted, there can be multiple kids per dad and hence this is a one-to-many merge

## Difference-in-Differences

In this homework, we replicate Holzer, Block, Cheatham, Knott (1993) using `jtrain.dta`. “The paper explores the effects of a state-financed training grant program for manufacturing firms in Michigan. Using a three-year panel of data from a unique survey of firms that applied for these grants, the authors estimate the effects of receipt of a grant on total hours.” We will replicate the key findings of this paper as follows:

Description of the policy treatment: "The data used for this analysis are from a unique survey of firms in Michigan that applied for training grants under the state's MJOB (Michigan Job Opportunity Bank Upgrade) program during 1988 and 1989.

For questions 6–7, only use the data from 1987 and 1988. Construct the Difference-in-Differences estimator in two different ways:

6. Construct the 4 means (pre-treatment control group, pre-treatment treatment group, post-treatment control group, post-treatment treatment group). Then, conduct a T-test on the mean difference of `hrsemp` between the treatment and control groups in two periods (Year1987, Year1988).
7. Run the following regression

$$hrsemp_{it} = \beta_0 + \beta_1 grant_{it} + \beta_2 1(year = 1988) + \beta_3 E_i + u_{it}$$

where  $1(\text{year} = 1988)$  is time dummy variable, coded as “1” if **year** equals 1988 and “0” otherwise.  $E_i$  is a dummy variable for being a treatment (i.e. someone who would receive the grant in 1988). (Hints:  $\text{grant}_{it}$  is the interaction of  $1(\text{year} = 1988)$  and  $E_i$ )

8. Run the fixed effect regression:

$$\text{hrsemp}_{it} = \theta_i + \beta_1 \text{grant}_{it} + \beta_2 1(\text{year} = 1988) + u_{it}$$

Do you get the same answer, why or why not?

For question 9–10, we use all the data rather than 2 years

9. Show the time trend of  $\text{hrsemp}_{it}$ . can we tell that the DID model satisfies the parallel trend assumption from the time trend figure? If not, what kind of empirical evidence do we need to verify the assumption? (Discuss the data and the way of visualization/statistical test)
10. Run a DID model with time and firm fixed-effects, cluster the standard error at firm-level, interpret the result.

$$\text{hrsemp}_{it} = \beta_1 \text{grant}_{it} + \theta_i + \gamma_t + u_{it}$$