Seminar 9 – Diff-in-diff and Matching methods

1. You are interested in estimating the effect of doing mathematics homework (yes/no) on mathematics test scores. You have data on all 10th graders in Oslo for the school year 2010/2011. The dataset contains the following variables

   - students: end year test scores, homework, nr. of missed classes, gender, age, test scores in 2009/2010
   - parents: education
   - teacher: gender, age, education

   About half of the students do their homework.

   (a) If you want to abolish homework, what effect would you want to estimate?
   (b) If you want to make homework mandatory what effect would you want to estimate?
   (c) You want to compare the effect of mandatory homework to the effect of an extra hour of math teaching. What effect of homework would you like to know?

   You want to estimate how well students that are currently not doing their homework would do, if they did their homework. You decide to use matching, and will therefore rely on a conditional independence assumption (CIA).

   (d) Explain your CIA. Be explicit about the counterfactual outcomes and the variables that you want to control for. Why might your CIA not hold? Can you think of examples where you get upward biased estimates? And downward biased estimates?
   (e) Explain how you use the CIA to estimate the counterfactual outcome, how you take into account that students that do their homework have different characteristics, and what support condition you need.
   (f) How would you estimate your effect using OLS?
   (g) You see in your data that boys never do their homework. What implications does this have for your research?

   You discover that not all teachers assign homework, and you get a new variable from Oslo municipality with information (0/1) on whether the teacher assigned homework or not. They tell you that teachers were assigned to give homework (or not) in a randomized experiment.

   (a) First you add this new information to your matching variables. What will happen to your estimates and standard errors?
   (b) How will you use this new data and what effects can you estimate?

2. We want to evaluate the impact of workers compensation for injuries at work on the duration of injury spells among workers, using data from Meyer, Viscusi, and Durbin, *American Economic Review* 85, 322-340, 1995.

(a) How would you expect compensations to affect injury duration? Consider estimating the following regression by OLS in a cross-section (at a given t):

$$duration_{it} = \alpha + \gamma benefits_t + \mathbf{x}'_{it}\beta + \epsilon_{it}$$

Why would we be skeptical about our estimate of γ?

Instead, Meyer et al. use a natural experiment where the US state of Kentucky raised the benefits for individuals with high earnings on July 15 1980 (Michigan also raised their benefits on January 1st 1982, but we only consider Kentucky in this exercise).

(b) Assume that $\epsilon_{it} = \mu_{high} + \zeta_t + u_{it}$ and that $E[u_{it}|\mathbf{x}_{it}] = 0$ where μ_{high} is a high earner fixed effect. Explain how you can then use the increase in the benefit level and a difference-in-differences (DD) model to identify the effect of the compensation level on the duration of injuries. Give an example where the necessary assumption(s) are violated.

(c) Using the means below, calculate a DD estimate of the impact of the minimum wage law on employment. Calculate the standard error of the DD estimate. How much does this suggest that the benefits level affects injury duration?

(d) What regression would you run to estimate the DD model including control variables?

(e) How could you evaluate whether the identifying assumption is likely to be satisfied?

```
. table post high, c(mean benefit)

------------------------------
=1 if     |
after     |
change in | =1 if high earner
benefits  |        0           1
----------+-------------------
        0 | 118.5757   151.0828
        1 | 118.2568   239.0925
------------------------------

. table post high, c(mean duration)

------------------------------
=1 if     |
after     |
change in | =1 if high earner
benefits  |        0           1
----------+-------------------
        0 | 6.271554    11.1766
        1 | 7.037328   12.89363
------------------------------

. table post high, c(sd duration)

------------------------------
=1 if     |
after     |
change in | =1 if high earner
benefits  |        0           1
----------+-------------------
        0 | 12.43191   28.98802
        1 | 16.12034   28.24566
------------------------------

. table post high, c(N duration)

------------------------
=1 if     |
after     | =1 if high
change in |    earner
benefits  |    0        1
----------+-------------
        0 | 1,705    1,233
        1 | 1,527    1,161
------------------------
```