Econometrics 1: Microeconometric Methods of Impact Evaluation.
Exercise 4: Difference-In-Differences

1. Use the database injury.dta available in the Lecture 4.

   The goal of this exercise is to evaluate the impact of workers compensation for injuries at work on injury duration. For more details, see B.D. Meyer, W.K. Viscusi, and D.L. Durbin (1995), "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment," American Economic Review 85, 322-340. What is the expected effect of compensations on the injury duration? Can we do a simple regression of the compensations on the duration an employee receive them? Why?

   In fact, the authors use a natural experiment: on July 15 1980, Kentucky (and on January 1st 1982 for Michigan), raised the benefit amount for high-earnings individuals.

   (a) Explain where there should be different responses for low-earners or high-earners. By writing down an equation, show how this can identify the desired effect.

   (b) Regress the log of duration on a dummy indicating wether the observation is before or after the reform, a dummy indicating whther the individual is a high-earner or not, and the interaction of both. Interpret the results. Why would certain omitted variables bias the estimation?

   (c) Introduce all the control variables that seem adequate. Try to justify why. Interpret the results.

   (d) If you had more precise information on the year of observation., which test could you undertake? (Think about Michigan). Show a specification that could test this.

   (e) Evaluate this strategy according to the problems explained in class: common time effects, endogeneity of changes, serial correlation. How would you do to solve these issues?

2. The goal of this exercise is to study the problem explained in Bertrand, M., Duflo, E. and S. Mullainathan (2004), "How Much Should We Trust Differences-In-Differences Estimates?", The Quarterly Journal of Economics, V.119, N.1, 1 February 2004, pp. 249-275(27).

   (a) Use the cigarette database under Lecture 3.

   (b) The goal is now to generate artificial laws against cigarette consumption: they should have no impact on consumption!!

      1. Keep only the year 1985
      2. Generate for each state a random number between 1 and 11 (which will be the year during which a law was passed, call it law). Hint: use the "uniform" command in stata.
      3. Keep randomly half of the 48 states (these steps follow exactly the paper in question). Hint: use again the "uniform" command in stata.
      4. Merge with the original database.

5. Generate a cumulative index of the law, called law_cum (equal to 1 every year after the law is passed, 0 otherwise).

(c) Regress the cigarette consumption (lgpackpc) on the cumulativ index of the law (law_cum). What do you find? Repeat the procedure several times just to be sure. Ideally, you could write a loop (with the "forvalues" comand in stata).

(d) Regress the cigarette consumption (lgpackpc) on the temporary index of the law (law). What do you find? What does this prove?

(e) Test for the presence of autocorrelation: with a graph, by regressing 1) lgpackpc on lgpackpc lagged, 2) law_cum on law_cum lagged, 3) the errors from a simple pooled OLS on these errors lagged. Is there a problem of autocorrelation?

(f) Regress the cigarette consumption (lgpackpc) on the cumulative index of the law (law_cum), accounting for an AR(1) disturbance term. Hint: xtregar.

(g) Use the cluster command, to add cluster(state). Why does it solve the problem? What do you find?

(h) If enough time, do a block bootstrap. Hint: bsample draws a sample with replacement from the existing data.