Data lab #4

Incentives, Entrepreneurship, and Innovation in Science

Data source: Hvide, Hans K., and Benjamin F. Jones. 2018. "University Innovation and the Professor's Privilege." American Economic Review, 108 (7): 1860-98.

Question for discussion:	1. How did the abolition of the Professor's Privilege influence patenting by Norwegian professors? 2. How did it influence start-up growth
Submit	1) a **pdf** file with your report 2) your .R code. Two files separately, no zipping.
Deadline	Lecture on Tuesday, November 3, 11 am You should complete datacamp.com exercises by this deadline.

Introduction

This is the first data lab where we will focus on individual inventors and their incentives rather than on the aggregate number of inventions in a country.

In 2003, Norway ended the "professor's privilege," by which university researchers had previously enjoyed full rights to new business ventures and intellectual property they created. The new policy transferred two-thirds of these rights to the universities themselves. This policy regime is common in the United States and many other countries today. You can think of this change as an introduction of a 66% "tax" on inventions by the professors, collected by the universities that employ them.

Recommended Readings

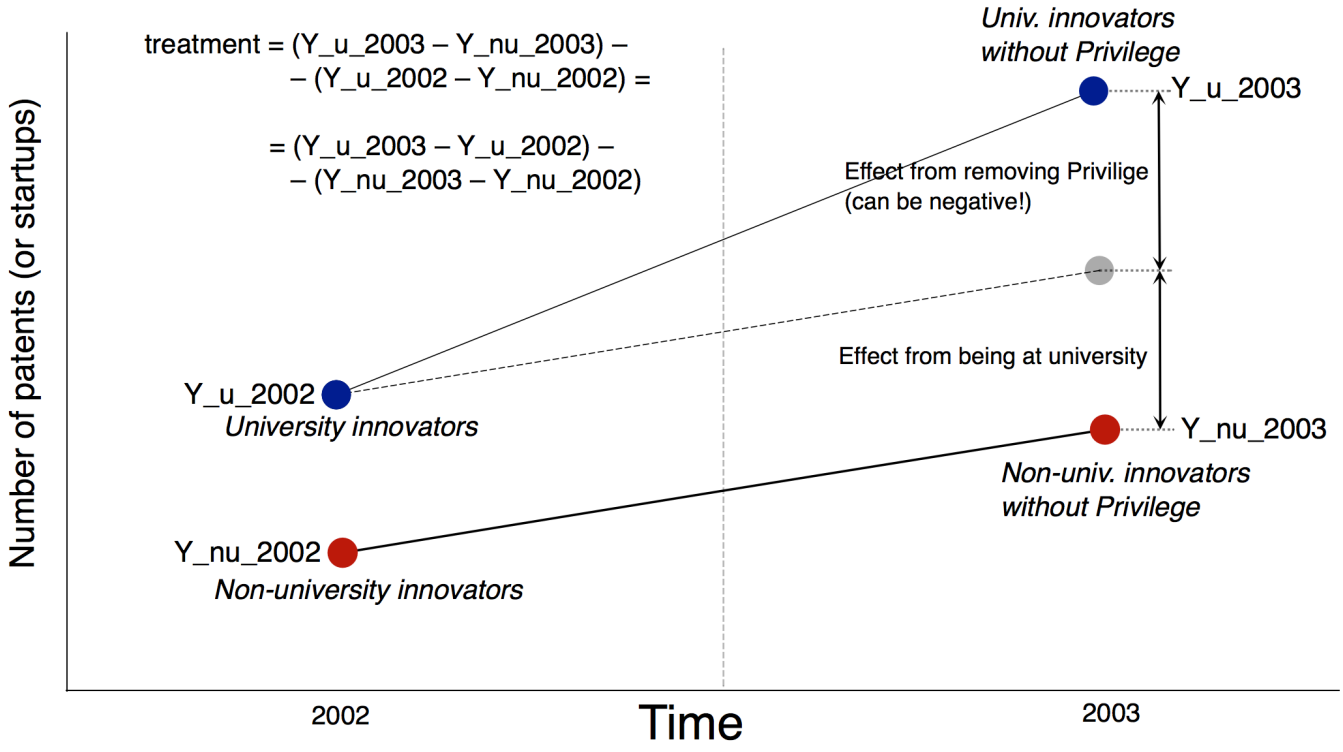
Review Chapter 5 of "Mastering Metrics" for a clear and concise econometric description of difference-in-differences analyses.

Instructions

0. Start by stating your research question. In examining the effects of a reform, like the professor's privilege, one could simply examine changes in the outcome variable (here patenting/ start-up entry for university professors) before and after the reform. In other words, one would examine the change in the outcome variable after a reform as a simple difference, without a control. Why could this approach lead you to the wrong conclusion? Explain and give an example to illustrate the potential problems with this approach. How does the use of a control group address this problem? State the identifying assumption that needs to be true if we want the difference-in-differences design to estimate the causal effect of the 2003 reform on the number of start-ups.

1. Reproduce Figures 1B and 2B from the paper (the per-capita patents and startups, 2 graphs in total). The figures should have all the key elements (including a clear legend and axes labels).

2. The difference-in-differences (henceforth diff-in-diff) design compares the changes in patenting after the reform for professors with the change in patenting for a control group of other Norwegians (or other Norwegian PhDs). With *individual-level* data we could run a t-test to see whether the difference between (Y_u_2003 – Y_u_2002) and (Y_nu_2003 – Y_nu_2002) is statistically significant. While we cannot access the original, individual-level data (for privacy reasons), we can examine the aggregate number of patents per group (professors, other PhDs, other Norwegians) and year. The graph below illustrates this test:



Make a 2x2 table that compares the means in this graph for both patents and start-ups in 2002 and 2003. Discuss the table. Was there a decline in patenting/start-ups for university professors, compared with the control?

3. While we cannot access individual-level data, we have industry-level data. Separately for patents and startups:

   a. Run a t-test to see if the **change** between 2002 and 2003 is statistically significant. You have to calculate the change in each category of patents/startups and then do a t-test. Report the relevant numbers in a table and in your text. (Do not take a screenshot of your results.)

b. What fields/industries drive the change in number of startups created? (Make sure to refer to the actual names of industries rather than their codes, see nace_codes.png for that)

4. Run the following diff-in-diff regression for the aggregate data. As a control group, you can use sample == 2 (all other workers in Norway) or some other sample (see the data description). You may run several specifications if you think it's useful

$$y_{ct} = \beta_0 Post_t + \beta_1 Treat_c + \beta_2 Treat_c \times Post_t + \varepsilon_{ct}$$

This is equivalent to running the t-test for the diff-in-diff (except this includes all years before 2002 and all years after 2003).

a. What do the coefficients β_1 and β_2 tell us in this regression? Do their estimates have the signs you expected?

b. Clearly state what the estimated effect is. For example, you can write "The estimated treatment effect is 2.2 more startups per 100,000 researchers" or "the regression suggests that the treatment effect is 10.1 fewer startups per 100,000 university workers". Compare it to your table from part 2.