Data lab #6

Biasi, Barbara, and Petra Moser "Effects of Copyright on Science. Evidence from the WWII Book Replication Program" http://ssrn.com/abstract=2542879.

Question for discussion: How do book prices influence the creation of new science?

Deadline: Tuesday, December 1, 11am

Introduction

In 1942, the Book Republication Program (BRP) allowed US publishers to reprint the exact content of science books with German-owned copyrights. Similar to compulsory licensing, the BRP improved access to German-owned knowledge by reducing the price of German-owned science books. After the BRP, the price of the average BRP book declined by 25 percent. In this lab, we examine how the BRP influenced the creation of new knowledge building on the knowledge in BRP books. You will also learn a new technique to build a better control group by selecting observations for the control group that are similar to the observations in the control group.

To measure changes in the creation of new knowledge building on BRP books, we use changes in the number of new articles and books that cite the knowledge in BRP books. More specifically, we will compare changes in citations to BRP books in chemistry and mathematics with changes in citations to Swiss books, as a control. Changes in citations to Swiss books are a good control for changes in citations to German books because, like German researchers, Swiss researchers were leaders in chemistry and mathematics. Yet, unlike Germany, Switzerland was neutral during WWII, so that Swiss books were not subject to the BRP.

The data structure is a panel of citations to BRP and Swiss book. For each book, you know the number of citations received per year. You also know the language of each citing publication.

Please note that you do not observe German/Austrian books that were not part of BRP. Thus, in your data, "being a part of BRP" is equivalent to "being a German/Austrian book".

Instructions

1. Please create the following summary statistics.
    a. What is the total number of books that became subject to the BRP? What is the most common field of BRP books?
    b. Create a histogram of the total number of citations for BRP books.

c. Create a histogram of the total number of citations for BRP books after 1942.

2. For a difference-in-difference regression to estimate the causal effect of the BRP, the following identification assumption has to be true: In the absence of the BRP, changes in citations to the treated group (the BRP books) and the control group should be similar. Challenges with this test are that 1) BRP books were not selected at random and 2) much of Europe was destroyed after WWII. We therefore need to take an extra step to build a comparison group.

   a. Your panel data include citations by authors writing in English, German, and other languages. Among these three groups, which authors do you expect to benefit most from the BRP? Why? Explain how you could use this feature of the empirical setting to develop an alternative difference-in-differences test that avoids the problem with selection into the BRP.

3. Instead of comparing the citations to BRP books in different languages, you can compare the changes in the citations to BRP books vs Swiss books (Switzerland was neutral during WWII).

   a. Based on the variables in your dataset, construct a suitable comparison group for the BRP books by finding Swiss books with similar characteristics. For example, you can compare BRP and Swiss books in the same fields or with similar number of citations. Explain how you define the comparison group and why.

   b. Create a similar to Figure 3: Plot the number of citations for BRP books and Swiss books per year (use the control group you constructed in part a.)

   c. Compare changes in citations to BRP books and Swiss books you created for question 3. Run the regression of the form:

$$cite_{it} = \beta BRP_i * post_t + book_i + \tau_t + \varepsilon_{it}$$

   where $cite_{it}$ is the citations measures you chose, BRP_i is 1 if the book is a part of BRP $post,$ is 1 for post-1942 and $book_i$ a and τ_t are book fixed-effect and year fixed effect

   d. *(optional)* Read about Mahalanobis nearest neighbor algorithm and pick a comparison group using it. You will still have to decide which variables to use for matching.