README for "Taxation and Innovation in the 20th Century" by Akcigit, Grigsby, Nicholas, and Stantcheva

Contents of Packet

This replication packet contains the requisite code and data to replicate "Taxation and Innovation in the 20th Century" by Akcigit, Grigsby, Nicholas, and Stantcheva. The packet contains two folders, labeled Data and Programs. These are described below:

Data

This directory contains the data used in the construction of the paper's primary tables and figures. Specifically, it contains three Stata .dta datasets:

1) state_data – contains state-level aggregates of innovation outcomes, the tax variables used in the analysis, and the various controls included in the paper's state-level regressions (e.g. R&D tax credits, GDP per capita, population density, etc.). It contains information for the fifty U.S. states between 1940 and 2000. This is the source data for Figure 1-4, Tables 2 and 3, and some of Table 1.
2) micro_reg_data – this is an inventor-level panel dataset containing information on innovation output, corporate inventor status, inventors' home and resident state tax rates, and all controls used in the inventor-level OLS and IV regressions in the paper: Table 4, much of Table 6, and all the results of Appendix C.4. Note that a collapse of this micro_reg_data.dta to the state x year level will not correspond to the data in state_data.dta one for one. This is because some patents have multiple inventors listed. Such patents are counted only once in state_data.dta, but will be counted more than once if micro_reg_data is collapsed.
3) MLogit_data – this dataset is the input data to all location choice models estimated in the paper, the results of which are reported in Table 5 and the final row of Table 6. An observation in the dataset is an inventor x year x state. The state in which the inventor resides in a given year will have the variable choice = 1. This dataset additionally includes information on inventor effective tax rates and controls.

For clarity's sake, we only include the top corporate tax rate and a few variables describing the corporate tax base rules. Researchers interested in using more detailed information on state corporate tax rules may email the authors.

We request that those who use these data cite "Taxation and Innovation in the 20th Century" appropriately.

Programs

This directory contains all the programs used to construct the results of the paper's primary tables and figures. Specifically it contains 11 programs plus one master program:

0) 0_main.do: Runs all of the programs necessary to replicate the paper's tables and figures.
1) Table1_SummStats.do: Creates Table 1, which is a table of summary statistics.

2) Table2AB_StateRegs_OLS.do: Runs State-level OLS regressions used in Panels A and B of Table 2, in addition to many of the regressions in Appendix C.3
3) Table2C_StateRegs_IV.do: Runs state-level IV regressions used in Panel C of Table 2, as well as a few regressions in Appendix C.3
4) Table3_LongDifferences.do: Runs state-level long-difference regressions used in Table 3.
5) Table4_MicroRegs.do: Runs the inventor-level micro regressions used in Table 4 as well as many of the regressions in Appendix C.4
6) Table5_MultinomialLogits.do: Runs multinomial logit location choice models used in Tables 5 and the bottom row of Table 6.
7) Table6_CorpInvInteract.do: Runs OLS regressions interacting corporate inventor flags with tax rates, as in Table 6.
8) Fig1_BinnedScatters.do: Produces state-level binned scatter plots as in Figure 1.
9) Fig2_visualize_IV_variation.do: Visualizes state-level variation in the instrumental variables as in Figure 2.
10) Fig3_EventStudies.do: Produces event study plots as in Figure 3.
11) Fig4_DistributedLags.do: Produces distributed lag plots as in Figure 4.

A Note on Permanence

The publicly available data used in the paper (e.g. state personal income per capita and state population counts) were downloaded between 2018 and 2020 from Census and BEA websites. Any future revisions to these series may lead to minor discrepancies between the published numbers and those included in this replication packet.

Replicating Results of Paper

Software requirements

The coding in this paper was all carried out in Stata-mp 14.

Instructions

To replicate the results of the paper, follow these two steps:

1) Change the global macro “projdir” to point to the location in which the replication packet is saved in your system architecture.
2) Run 0_main.do in Stata which will run all of the programs necessary for the replication. Alternatively, one can run individual programs to replicate specific tables, as described above and commented in the codes.

The numbers to the tables will appear in the log files of these codes.