Read-me File for the Replication Package for the paper:

"Financial Disruptions and the Organization of Innovation: Evidence from the Great Depression"

By Tania Babina, Asaf Bernstein, and Filippo Mezzanotti

This file reports background information on the replication package for the tables and figures that are enclosed in the main body of the paper "Financial Disruptions and the Organization of Innovation: Evidence from the Great Depression." Before working on the replication package, please read carefully this file and the paper. As we explain below, the package contains pseudo-data, where variables coming from the patent data are randomized.

The replication package is made up by three folders. The folder output contains the actual output for the tables and figures reported in the main body of the paper. Tables are reported in tex format and figures in PDF. If you plan to cite or use any of these figures, please cite the paper accordingly.

The folder code contains the Stata code to estimate the models reported in the main tables and figures in the paper, and to produce the output. In particular, we report three files. The file "FINAL_REPLICATION_MAIN_rep" contains the code to generate the main tables and figures, generally using data at county level. The file "FINAL_REPLICATION_INDIVIDUAL_rep" contains the code to produce table 10, which is estimated using individual level data. Lastly, the file "SUMMARY_REPLICATION_rep" contains the code to produce table 2, which is a summary statistics table, combining both patent level data and county level data. In principle, each file can be run automatically with the current folder structure of the replication package; however, the user would need to change the source directory (i.e., cd) at the top of the file. The installation of some external Stata package may be also necessary.

The folder input contains the data sets to run the code. As a general rule, each data sets contains the real content of a variable for those variables that were obtained from publicly available sources (e.g., bank data, county characteristics, etc.). However, we replace the content of the variable with a random number when the variable comes from the patent data put together by the authors. As we discuss in the data section of the paper, our patent data set is constructed starting from the Comprehensive Universe of US Patents (CUSP) (Berkes, 2018). To reconstruct our data from CUSP, you can follow the instructions reported in the data section of the paper and in Online Appendix A and B posted in Tania Babina website.[1] For any further question on data construction, please directly contact Tania Babina. For continuous variables, we have replaced the value with numbers drawn from a uniform between zero and one; for dichotomous variables we have randomized an indicator variable.

In particular, the folder input contains five files, all in Stata format. The file "final_data_county_rep" contains the main data set organized at county by decade level. Note that our county ID is constructed as a combination of a state FIPS code and county level ICPSR code.

[1] At the time of the submission, the link for Appendix A is: https://www.dropbox.com/s/lu6p9yby3ed2mlk/Online%20Appendix%20A.pdf?dl=0. The link for Appendix B is: https://www.dropbox.com/s/4yke1vaojk895c0/Online%20Appendix%20B.pdf?dl=0

The file “final_data_tech_year_level_rep” provides the data at county-level but with the breakdown of total patenting by technology. The file “final_data_year_level_rep” provides patent data at annual level. The file “final_individual_rep” contains the sample used for the individual level analysis. The file “final_patent_list_rep” contains the hypothetical patent list, which is used to create the first panel of Table 2. For all datasets except for the last one, the size of the sample is consistent with the sample size used in the paper.

Best,

Tania, Asaf, and Filippo