

Readme for data files for “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure” by Dave Donaldson

Table 1 is generated by “Analysis/descriptives/T1_descriptives.do” in Stata.

Table 2 is generated by “Analysis/trade costs/TC estimation.do” in Stata. However, that do-file requires some .csv input files that are generated by “Analysis/trade costs/TC_est_prep.m” in Matlab that needs to be run first (and several times, as per the comment beginning on line 16), and which calls other .m files in this same folder. (The files in “matlab_bgl_4.0_osx64” come from the Boost Graph Library Matlab project.) This m-file takes a very long time to run and generates a large number of large files so is best run on a server (and ideally one that is set up for parallel processing) with a lot of storage space. The do-file also takes a long time due to the grid search and bootstrapped confidence intervals.

Table 3 is generated by “Analysis/trade flows/gravity estimation.do” in Stata. However, that do-file requires a .csv input file that is generated by “Analysis/trade flows/TF_est_prep.m” in Matlab that needs to be run first.

Table 4 is generated by the first part of “Analysis/income/income estimation.do” in Stata.

Table 5 is generated by the second part of “Analysis/trade flows/income estimation.do” in Stata. However, that do-file requires a .csv input file that is generated by “Analysis/simulation/T5_simulation.m” in Matlab that needs to be run first. This m-file calls other .m files and uses some data (.csv) files in this folder.

All of the above .do and .m files draw on data files that are all in “Data/”.