**Data and Replication Files**

**“The Persistent Effect of Temporary Input Cost Advantages in Shipbuilding, 1850-1911.”**

**W. Walker Hanlon**

**September, 2019**

This folder contains the data and replication files behind the analysis in “The Persistent Effect of Temporary Input Cost Advantages in Shipbuilding, 1850-1911.” This word file describes the layout of this folder and the function of the files that it contains. Further questions about these files should be directed to [walker.hanlon@gmail.com](mailto:walker.hanlon@gmail.com). The analysis uses Stata 15.

**data** -- this folder contains the data used in the analysis.

The “master\_data\_all\_years” contains the core data set from which separate analysis data files are produced. This original data were entered from Lloyds and American Bureau of Shipping registers for 1871, 1889 and 1912. The raw entries were then processed to remove duplicate entries that appear in both registries whenever possible, as well as entries found in one year that also appeared in an earlier registry. They location information in the original data was also cleaned in several rounds of cleaning and latitude/longitude and region identifiers were added. Finally, only entries from Canada, the U.S. and the U.K. were retained. Entries for other countries, as well as the translation tables used to standardize the location information, are available upon request.

The four files with names beginning with “analysis” are produced by collapsing the master data by location and decade and formatting the data in various ways. Once the date are organized by location, the distance to the nearest navy shipyard is also added. These four files form the main input to the analysis. The ending “100t” indicates that ships of less than 100 tons have been omitted from the data in a file.

The three files with names beginning with “controls” contain the control variables used in the analysis. The sources and construction of these variables are described in the online appendix.

The three files with names beginning with “nearby\_tons” contain calculations of the tons within various distances of a location. These are calculated using the tonnage and latitude/longitude information in the master data.

The “homeport\_data\_1912” file contains data from the American Bureau of Shipping registry for 1912 for which additional details have been entered from the raw source data. Specifically, this file contains information on the homeport of the ship as well as ship draft and ship owner for a subset of ships.

**translation\_tables** – this folder contains a translation table that identifies shipbuilding areas. These areas are described in Online Appendix Figure A.7. These are used when clustering standard errors by shipbuilding region.

**temp –** this folderexists only to receive temporary files created during the process of preparing data for an analysis.

**code** – this folder contains the analysis code for each of the tables and figures in the main analysis.

**results –** this folder contains the results produced for by the analysis code corresponding to each of the tables and figures in the main analysis.

**appendix\_code** – this folder contains the analysis code for each of the additional analysis tables and figures included in the online appendix. Note that code is not included for figures or tables that are peripheral to the main analysis and not based on the ship registry data. In those cases, the sources of the data displayed in any tables are figures are described in the text.

**appendix\_tables\_figures –** this folder contains the results produced for by the analysis code in the “appendix\_code” file corresponding to each of the tables and figures in the online appendix.