Replication files for:

“Coal Smoke, City Growth, and the Costs of the Industrial Revolution”

W. Walker Hanlon

March 2, 2019

This folder contains replication files for the main results tables and figures. The file is organized as follows. The data used in the analysis are in the “input\_data” folder. The sources for these data can be found in the text of the paper or the accompanying online appendix.

The “prepare\_key\_variables” includes information on the construction of key variables where the procedure may not be clear from the paper text.

The remaining folders contain the replications files corresponding to each of the main analysis figures and tables included in the published version of the draft. Each folder contains one or more .do files which were prepared using Stata version 15 (64 bit).

To run an individual do file, simply replace the global at the top with the correct filepath for your system. Note that maintaining the overall file structure is important for allowing each individual do file to run smoothly.

Replication files for the more extensive set of robustness checks described in the online appendix are available on request by contacting the author at [walker.hanlon@gmail.com](mailto:walker.hanlon@gmail.com).

**Input files index** – a brief description of the files contained in the “input\_data” folder. For further details on the sources of each data series see the Online Appendix.

**bot\_data\_all** – this file contains a cross section of local rental, goods prices, and wage data produced by the Board of Trade in 1905. It contains the following data series:

* rent – price of rental dwellings
* prices – index of local goods prices
* combined – local cost of living price constructed by the Board of Trade from rent and goods prices

The next few series present the wages for various workers, such as skilled builders, unskilled builders, skilled engineers, etc.

**CCG\_data\_agg\_analysis\_industries** – this is the main city-industry database which is derived from the Census of Population occupation data. For further details on the construction of this database see <http://walkerhanlon.com/>.

**city\_characteristics** – this file contains characteristics such as latitude and longitude, patents, rainfall, air frost days, and a region identifier for the cities included in the main analysis database. See the online appendix for data sources.

**city\_characteristics\_qol\_analysis** – this file contains the same city characteristics but for the extended set of cities used in the quality of life analysis.

**city\_mortality\_prepared** – this file contains series reflecting violent deaths based on data from the Registrar General’s Reports.

**city\_pop ­–** This file includes data on overall city population from the Census of Population

**coal\_use\_adjustment\_factors --** this file contains coal use adjustment factors. The construction of this variable is described in the “prepare\_key\_variables/coal\_use\_adjustment\_factors” folder

**industry\_coal\_use --** this file contains industry coal use per worker values which have been constructed using the data on coal (and coke) use by industry and employment by industry from the 1907 Census of Production.

**industry\_demands –** These data describe the division of industry demand across households, investment, government, and exports using data from the input-output table constructed in Thomas (1987)

**industry\_salaried\_shares** – this file contains data on industry employment of salaried and wage workers from the 1907 Census of Production. See the Online Appendix for further discussion.

**input\_shares\_for\_estimation** – this file contains industry labor, capital and intermediate cost shares based on data from the 1907 Census of Production. See the Online Appendix for further discussion.

**prepared\_city\_carb50 --** this file identifies the share of the area within 50km of the district center that is composed of carboniferious (coal-bearing) rock strata. Data on the geology comes from the UK Geological Survey and the measures were constructed from this using GIS.

**prepared\_connections\_matrices\_controls\_lag20\_IO** – this file contains controls for local input-output connections. These are based on the input-output matrix constructed by Thomas (1987) following the approach used in Hanlon & Miscio (2017). The files constructing these variables are in the “prepare\_control\_variables\connections\_matrices\_controls” folder.

**prepared\_connections\_matrices\_controls\_lag20\_OCC** – this files contains controls based on the local occupational and demographic similarity of industries. These controls are from Hanlon & Miscio (2017) and are based on data from the Census of Population. The files constructing these variables are in the “prepare\_control\_variables\connections\_matrices\_controls” folder.

**prepared\_industry\_demographics** – this file contains the female and youth employment share in each industry based on Census of Population data

**prepared\_national\_firm\_size\_data –** this file contains a data series showing the firm size of the firm employing the median worker (ordered by firm size) in each industry. This is calculated using data on number of employees reported by firm owners in the 1851 Census of Population.

**QOL\_analysis\_data** – this file contains employment by city and industry for the extended set of cities used in the Quality of Life analysis using Census of Population data from 1901.

**seaports\_controls\_cities –** this file contains information on whether each city was a major seaport using data from the Board of Trade.

**workpop\_data --** this file contains city working population including both the analysis industries as well as workers in other sectors (e.g., government). The data were constructed from the Census of Population reports.