**READ ME**

This directory contains the data and code to replicate the results in

**“Task Specialization in U.S. Cities from 1880-2000”, (Guy Michaels, Ferdinand Rauch, Stephen Redding), Journal of the European Economic Association**

Stata do-files to reproduce Figures and Tables of the paper are located in the folder “do-files/Figures\_Tables\_Paper”. Stata do-files to reproduce Figures and Tables of the web appendix of the paper are located in the folder “do-files/Figures\_Tables\_Appendix”. The datasets that these do-files use are located in the “data” folder.

Do-files were last tested using Stata version 11.2.

Do-files partly rely on non-standard Stata commands “outreg2” and “run running”.

The main data contribution of this paper is summarized by these two datasets:

* VerbFreqvo: contains the verb frequencies for each verb and occupation
* TaskContentko: contains the task content measures for each thesaurus category and occupation

The following table gives a brief overview of the intermediate data files contained in this replication folder. A more detailed description of these data and their sources are given in the paper and its online appendix.

|  |  |
| --- | --- |
| class.dta | Thesaurus class scores by occupation. |
| indreg.dta | Broad industry regression output, generated in main tables do files. |
| interactiveness\_measure.dta | Mean interactiveness by occupation. |
| occreg.dta | Broad industry regression output, generated in main tables do files. |
| onet.dta | Onet measures by occupation. |
| panel\_county.dta | County data, including phone and highway data. |
| panel\_decomposition.dta | Industry decomposition by year. |
| panel\_decomposition\_metro.dta | Industry decomposition by year for metro sample. |
| panel\_decomposition\_nonmetro.dta | Industry decomposition by year for non-metro sample. |
| panel\_interactiveness.dta | This is the **main dataset**. It gives the main measure of interactiveness at the occupation-industry-year level. It was created by collapsing census micro data to the aggregate level described. |
| panel\_interactiveness\_man.dta | Main dataset constructed from subset of selected micro data only. Subset: Manufacturing. |
| panel\_interactiveness\_mar.dta | Main dataset constructed from subset of selected micro data only. Subset: Married. |
| panel\_interactiveness\_nomar.dta | Main dataset constructed from subset of selected micro data only. Subset: Non married. |
| panel\_interactiveness\_noskill.dta | Main dataset constructed from subset of selected micro data only. Subset: Low skill. |
| panel\_interactiveness\_serv.dta | Main dataset constructed from subset of selected micro data only. Subset: Services. |
| panel\_interactiveness\_skill.dta | Main dataset constructed from subset of selected micro data only. Subset: High skill. |
| panel\_metarea.dta | Interactiveness by metropolitan area. |
| panel\_occ\_ind | ALM measures and interactiveness by occupation, industry and year. |
| panel\_occ\_ind\_metro | Main dataset, including metropolitan area status. |
| results1939.dta | Verb regression output using the 1939 DOT. |
| thesaurus | Thesaurus verb counts. |
| thesaurus\_verb\_meaning | Verb share by thesaurus class. |
| thesaurus\_verb\_number | Verb counts by thesaurus class. |
| thesaurustaskcontent | Verb count indices by thesaurus section. |
| thesaurus1939 | Thesaurus verb counts for the 1939 DOTs. |
| verbregs | Coefficients from the verb level regressions. |
| wage-occ1950-year | Wage by occupation and year. |
| y1880\_coll | Micro data collapsed to level of occupation-industry. |
| y1900\_coll | Micro data collapsed to level of occupation-industry. |
| y1920\_coll | Micro data collapsed to level of occupation-industry. |
| y1940\_coll | Micro data collapsed to level of occupation-industry. |
| y1960\_coll | Micro data collapsed to level of occupation-industry. |
| y1980\_coll | Micro data collapsed to level of occupation-industry. |
| Y2000\_coll | Micro data collapsed to level of occupation-industry. |
| y1880 | Micro data by year. |
| y1900 | Micro data by year. |
| y1920 | Micro data by year. |
| y1940 | Micro data by year. |
| y1960 | Micro data by year. |
| y1980 | Micro data by year. |
| Y2000 | Micro data by year. |