## Overview of files and folders in .\pred_wage_instrument

Files in this folder are used to create the predicted log wage instrumental variable based on data from the 5% 2000 US Census from IPUMS.

All folder references below begin with root folder: .\FINAL_Replication_CLMP_JPE

### Stata Code

- *main.do*
  - sets folder structure and executes all other .do files in this folder
  - Must set $root folder.

- *create_raw_data.do*
  - reads in raw Census ascii data file .\data\data_Census\usa_000009.dat
    - see .\data\data_Census\IPUMS_USA_5pct_extract_info.pdf for details on data extract request
  - formats and labels key variables
  - temporarily saves data in .\code\data_cleaning\pred_wage_instrument\temp_data\census_raw.dta

- *clean_census.do*
  - reads in .\code\data_cleaning\pred_wage_instrument\temp_data\census_raw.dta
  - cleans data, defines key variables and sample to be used
  - temporarily saves cleaned data in .\code\data_cleaning\pred_wage_instrument\temp_data\clean_census_5pct.dta

- *predict_census.do*
  - reads in .\code\data_cleaning\pred_wage_instrument\temp_data\clean_census_5pct.dta
  - predicted log wages are obtained from gender-specific regressions of log wages on personal characteristics (e.g., race, potential experience, educational attainment, industry, occupation, and state of residence).
  - saves predicted log wage measures used as our instrumental variable in data set .\data\data_derived\predicted_lnwage3.dta.

### Other folders

- .\code\data_cleaning\pred_wage_instrument\logs – Contains log files produced by .do files described above
- .\code\data_cleaning\pred_wage_instrument\temp_data – Used as a repository for temporary data files when running Stata .do programs