David Deming

"Why Do Wages Grow Faster for Educated Workers?" README file

(last updated December 2024)

This README file explains the contents of the replication folder and provides instructions for how to replicate the results from the paper and the appendix. The do files allow you to reproducing all tables and figures, using Stata (version 16.1).

To perform the replications, first run the "preprocess" files for the NLSY79 and NLSY97. They are made as separate files in case you wish to focus on only one of the two surveys. The files "nlsy79_preprocess" and "nlsy97_preprocess" call on .csv files that were created directly from the NLS data investigator, which can be accessed here - https://www.nlsinfo.org/investigator/pages/home. The files nlsy79_value-labels.do and nlsy97_value-labels.do give the complete variable list, which allows you to recreate the data query on the NLS investigator site. They also rename and label the variables so that they can be more easily used.

The file "nlsy_analysis_indiv_panel" calls on the datasets created by both preprocess files above, and generates the results for main Figures 1, 2, 3, 7 and 8, and Appendix Figures A1-A5, Figures A9-A11, and Appendix Table A5. These are the results that use only the individual level panel data (e.g. they are unique at the person ID by year) and do not focus on within- vs. between-job sources of wage growth.

The files "nlsy79_analysis_job_panel" and "nlsy97_analysis_job_panel" call on the datasets created by both preprocess files and generate the results for main Figures 4-6 and Table 1, as well as Appendix Figures A6-A8 and Appendix Tables A1-A4. These results convert the data into a panel of job spells (e.g. they are unique at the person ID by job level).

There is also a file called "cps_wage_projections.do" that constructs the average wage path by age of different occupations using the CPS ASEC, for use in Figure 8. The CPS data were downloaded from IPUMS-CPS. The data file was too large to include directly in the replication file, but instructions for replicating the IPUMS query are included in the .do file.

The programs above call on a few other datasets, which are included in the replication file:

1. I use the age-adjusted AFQT score created by Altonji, Bharadwaj, and Lange (2009), which was designed to make it comparable across survey waves. That dataset is afqt_adjusted_final.dta and was downloaded directly from their replication materials.
2. I use the 1998 O*NET measure of occupation routineness, which I take directly from the replication files in Deming (2017). Autor and Dorn (2013) and Deming (2017) both convert the O*NET task routineness measure (which is on a Likert scale, ranging from 1-7) and convert it into a 0-10 scale where occupations are given the value that reflects their labor-supply weighted percentile rank in the employment distribution. The file routine_convert_2017.dta contains this rescaling, which is done on the 1% files of the 2017-2019 American Community Survey with a sample restricted to working adults (ages 18-65).

3. Finally, I include a series of crosswalks created by David Dorn to link occupations and industries across Census years. Those original files can be found on his data page - https://www.ddorn.net/data.htm.