

# Summary of Replication Package For “The Growing Importance of Social Skills in the Labor Market”

David Deming

May 26, 2017

This document describes the contents of the replication package for “The growing importance of social skills in the labor market” (Deming, 2017), forthcoming at *The Quarterly Journal of Economics*. The first section explains how the included do files can be used to reproduce the tables and figures in the paper. The second section lists the data sources for the project.

## 1 Do Files and Results

- `0-master.do`: This is the master do file that runs all programs to clean the data, run the analysis and create the tables and figures in the paper. You can use this do file to run all programs at once, or you can run each program separately.
- `1-onet_import.do`: This do file reads the 1998 O\*NET data into Stata and creates the composite measures of task content that are used in the analysis.
- `2-census_acs_cleaning.do`: This do file cleans the raw 1980, 1990 and 2000 Census files and the 2005-2013 American Community Survey (ACS) files and prepares them for the analysis. Note that the raw Stata data files that are the inputs to this program are not included in the replication package due to their large size, so this program cannot be run without first downloading the data from the IPUMS USA website and reading the `.dat` files into Stata. See the description of the **import** folder below for details on how to obtain these data.
- `3-census_acs_analysis.do`: This do file runs the Census-ACS analysis and creates Figures 1, 3, 4, 5, A1, A2, A3, and A4 and Tables A1, A5 and A6 in the paper.
- `4-nlsy.do`: This do file imports and cleans the National Longitudinal Survey of Youth 1979 and 1997 (NLSY79 and NLSY97) data files, runs the analysis, and creates Tables 1, 2, 3, 4, 5, A2, A3 and A4 in the paper.
- **Note**: To run these programs, you must replace the words `YOURFILEPATH` in the global macro `topdir` at the start of each do file with the file path for the folder where you have saved the replication package.

## 2 Data Sources

The bullet points below refer to folders and sub-folders within the **Data** section of the replication package.

- **census-ac**s: This folder contains data from the Census and American Community Survey (ACS) and related files, as follows.
  - **clean**: The `.dta` files in this folder are created as output of `2-census_acs_cleaning.do` (except for `1980_2012_app_tables.dta`, which is created by `3-census_acs_analysis.do`).
  - **collapsed**: The `.dta` files in this folder are created as output of `3-census_acs_analysis.do`.
  - **import**: The `do` files in this folder read the raw 1980, 1990 and 2000 Census and 2005-2013 ACS `.dat` files into Stata.
    - \* As with the main `do` files, to run these programs you must replace the words `YOURFILEPATH` in the global macro `topdir` at the start of each `do` file with the file path for the folder where you have saved the replication package.
    - \* The raw `.dat` files are not included in the replication package due to their large size. These files can be downloaded from the IPUMS USA website (<https://usa.ipums.org/usa/>) using the “Select Data” tool, as follows.
      - Download the data separately for each of the following samples: 1980 5% state sample, 1990 5% sample, 2000 5% sample, 2007 ACS 3-year sample (which includes data from the 2005, 2006 and 2007 1% ACS samples), and 2010 ACS 3-year sample (which includes data from the 2008, 2009 and 2010 1% ACS samples). Finally, download data from the 2011, 2012 and 2013 ACS 1% samples as a single file. Save the `.dat` files in the **import** folder, and rename them `1980.dat`, `1990.dat`, `2000.dat`, `2006.dat`, `2009.dat`, and `2012.dat`, respectively.
      - Select the following variables: `year`, `multyear`, `perwt`, `sex`, `age`, `uhrswork`, `wkswork1`, `wkswork2`, `incwage`, `educ`, `classwkr`, `empstat`, `gqtype`, `occ`, `occ1990`, `ind`, and `ind1990`. Note that not all of these variables are present in each year (specifically, `wkswork1` is not available after 2007, and `multyear` is only available for the 2007 and 2010 3-year files). Also note that when you select `gqtype`, `educ`, `empstat` and `clsswkr`, the extract will automatically also include the version of these variables with “detailed” codes (`gqtyped`, `educd`, `empstatd` and `clsswkrd`), which are used in the data cleaning process.
    - \* Once you have downloaded the raw data and renamed the `.dat` files, you can run the `do` files in this folder, which will read the data into Stata and save it in the **raw** folder.
  - **raw**: This folder is empty, as the raw Census and ACS Stata files are not included in the replication package due to their large size. If you download the the Census-ACS data from IPUMS USA and read it into Stata (see instructions under previous bullet point), you will save the resulting `.dta` files in this folder. You can then run `2-census_acs_cleaning.do`, which takes the raw Census and ACS files as inputs.
  - **xwalk\_ind**: The `.dta` files in this folder consist of crosswalks between the industry codes in the 1970, 1980, 1990 and 2000 Censuses and a consistent set of aggregated industry codes (the “`ind6090`” codes) used by Autor, Levy and Murnane (2003) and Autor and Price (2013) and originally developed by Autor, Katz and Krueger (1998).
    - \* The `ind70.dta`, `ind80.dta` and `ind90.dta` files are from the Autor and Price (2013) replication package, provided to me by David Autor.

- \* The file `ind00.dta` is created by `4-nlsy.do`.
- **xwalk\_occ**: The files in this folder consist of crosswalks between the Census and ACS occupation codes and a consistent set of aggregated occupation codes (the “occ1990dd” codes) originally developed by Autor and Dorn (2013) and updated for this project.
  - \* Specifically, the files with names ending in “\_update.dta” are the updated crosswalks that map the 1980 Census, 1990 Census, 2000 Census, 2005-2009 ACS and 2010-2014 ACS codes, respectively, to the occ1990dd codes.
  - \* The `occ1990dd_titles.csv` and `occ1990dd_titles.dta` files contain the titles of the updated occ1990dd codes.
  - \* The `occ70-80cw.dta` file is a crosswalk between the 1970 and 1980 Census occupations from the Autor and Price (2013) replication package.
  - \* Finally, `occ1970_occ1990dd.dta` is a crosswalk between the 1970 Census codes and the occ1990dd codes that is created by `4-nlsy.do`.
  - \* See Section 2 of the data appendix for details of the occ1990dd updates.
- **dot**: This folder contains data from the Fourth Edition of the Dictionary of Occupational Titles (DOT), published in 1977.
  - Specifically, `dot77-70.dta` contains data on the four DOT variables used Autor, Levy and Murnane (2003), aggregated by 1970 Census occupation.
  - The file called `treiman.dta` is a sample of 1980 Census respondents coded with both 1970 and 1980 Census occupation codes by Donald Treiman.
  - Both `dot77-70.dta` and `treiman.dta` are from the Autor and Price (2013) replication package. Also see Section A.2 of the Data Appendix in Autor, Levy & Murnane (2003) for more details on these files.
  - Lastly, `dot77-occ1990dd-Tr.dta` is created by `2-census_acs_cleaning.do`.
- **nlsy**: This folder contains data from the NLSY79 and NLSY97.
  - **import**: This folder contains two sub-folders with the raw NLSY79 and NLSY97 data downloaded using the NLS investigator (<https://www.nlsinfo.org/investigator/pages/login.jsp>).
    - \* The files ending in `.csv` are the raw data that are loaded into Stata by `4-nlsy.do`, while the `.do` files provide value labels (and are also run by `4-nlsy.do`).
    - \* The files ending in `.cdb` are codebooks for the extracts.
    - \* Finally, the files ending in `.NLSY79` and `.NLSY97` are “tagsets” that include a list of the variables in the extract, which can be uploaded to the NLS investigator to recreate the files in this folder.
  - **altonjietal2012**: This folder contains a file, `afqt_adjusted_final.dta`, from the replication package for Altonji, Bharadwaj and Lange (2012).
  - The files `nlsy_merged.dta`, `nlsy79_clean.dta` and `nlsy97_clean.dta` are created by `4-nlsy.do`.

- **onet**: This folder contains data from the 1998 O\*NET.
  - **text\_files**: This folder contains the raw O\*NET data in text file format.
    - \* Data are available from the National Crosswalk Service Center (<https://www.xwalkcenter.org/index.php/component/content/article/83-onetinfo/103-onet98>) as a Microsoft Access Database.
    - \* Data can be exported from the database as text files. Do not select the option to export with formatting and layout.
    - \* The database objects named “means\_ab”, “means\_gw”, “means\_in”, “means\_kn”, “means\_ov”, “means\_sk”, “means\_wc”, “xwalk” and “occupat” correspond to the text files named `Ability.txt`, `WorkActivity.txt`, `Interest.txt`, `Knowledge.txt`, `WorkValue.txt`, `Skills.txt`, `WorkContext.txt`, `Crosswalk.txt` and `Titles.txt`, respectively.
  - The 1998 O\*NET Data Dictionary (`onet98_datadict.pdf`) can also be downloaded from the National Crosswalk Service Center at the same link as above.
  - The files `onet98_occ90_xwalk.dta`, `onet98_occ1990dd` and `onet1998.dta` are created by `1-onet_import.do`.
  - The file `onet98_occ1990dd_pct.dta` is created by `4-nlsy.do`.

## References

- Altonji, J., Bharadwaj, P. and Lange, F.: 2012, Changes in the characteristics of american youth: Implications for adult outcomes, *Journal of Labor Economics* **30**(4), 783–828.
- Autor, D., Katz, L. and Krueger, A.: 1998, Computing inequality: Have computers changed the labor market?, *The Quarterly Journal of Economics* **113**(4), 1169–1213.
- Autor, D., Levy, F. and Murnane, R.: 2003, The skill content of recent technological change: An empirical exploration, *Quarterly Journal of Economics* **118**(4), 1279–1333.
- Autor, D. and Price, B.: 2013, The changing task composition of the US labor market: An update of Autor, Levy, and Murnane (2003), *Working paper*, Massachusetts Institute of Technology, Cambridge, MA.