

Documentation for Do-files and Data Sets

Replication files for “Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program”

Martha J. Bailey, Hilary Hoynes, Maya Rossin-Slater, and Reed Walker

Last update: April 17, 2023

Overview

This document provides instructions on how to use the files contained in the **REPLICATION_PACKAGE** folder, which includes Stata code and other files used to construct figures and tables in Bailey, Hoynes, Rossin-Slater, and Walker (2023).

Data availability

Our main results are produced using confidential microdata from the U.S. Census Bureau and the Social Security Administration. Data from the Social Security NUMIDENT file, the 2000 Census, and the 2001-2013 American Community Survey contain restricted information and are available for access through the Census Bureau’s Research Data Center (RDC) network. To obtain access to these data, see the Census Bureau’s directions for writing a proposal: <https://www.census.gov/programs-surveys/ces/data/restricted-use-data/apply-for-access.html>. Your proposal must request access to each of the restricted-use datasets listed below.

Data sources and descriptions

1. The paper relies on public data and data build files from the **openicpsr-146361**: Bailey, Martha J., Sun, Shuqiao, and Timpe, Brenden. Data and Code for: Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency. Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-11-19. <https://doi.org/10.3886/E146361V1>
2. In addition to the publicly available data in openicpsr-146361, the paper relies on restricted data accessed via the Census Bureau’s Research Data Center network, project 1284. These data can be accessed by approved researchers in the RDC network.
 - a. **Social Security NUMIDENT file.** These data are derived from the Social Security Administration’s database of applications for Social Security cards. They include information on exact date of birth, place of birth, and exact date of death for

- individuals no longer living. [Note: there is no official citation for the NUMIDENT file].
- b. **Long-form Decennial Census 2000 and 2001-2013 American Community Survey.** These data are produced by official Census Bureau surveys and provide information on place of residence (including group quarters residence), human capital outcomes, economic self-sufficiency outcomes, and other information on large samples of the U.S. population. These data have been linked to the NUMIDENT to provide information on place of birth and, by extension, exposure to Head Start during childhood. [Note: there is no official citation for either the Decennial Census data or the ACS files].
3. In addition, in some results presented in the Online Appendix, the paper uses publicly available 1970 and 1980 Decennial Census and Panel Survey of Income Dynamics (PSID) data. One set of results relies on restricted data from the PSID: Panel Study of Income Dynamics, restricted use data. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI (data accessed in year 2019). These data can be accessed by approved researchers through the process outlined here: <https://simba.isr.umich.edu/restricted/ProcessReq.aspx>

Description of the code and other replication files

Figures 3, 4, 5, and 6, Tables 1-8, Appendix Figures 3-6, and Appendix Tables 1-7 are created from analyses conducted inside the RDC. All of the results contained in these figures were disclosed from the RDC as Excel (.xls) files, and then the .xls files were transformed into the figures and tables shown in the paper. Thus, (i) we provide the code we used to generate all of the results contained in these figures that generated the .xls files; and (ii) information on how these .xls files were manipulated to generate the figures and tables outside the RDC (either using Stata code or the Excel2Latex software).

Other figures and tables in the paper were created using data sources outside the RDC and we provide more details about them below.

Specifically, the **REPLICATION_PACKAGE** folder includes:

- (i) The disclosed .xls files containing the results in Figures 3, 4, 5, and 6, Tables 1-8, Appendix Figures 3-6, and Appendix Tables 1-7
- (ii) two do-files **make_graphs_final.do** and **construct_tables_final.do** that take these .xls files and generate some of the figures and tables in the paper, respectively
- (iii) a do-file and data set used to construct the map in Figure 1: **fsrollout.dta & colormap.do**
- (iv) a do-file used to create the schematic Figure 2 that is not based on data: **figure2schematic.do**

- (v) a sub-folder called **Manuscript_29108_DisclosedCode**, containing further subfolders with disclosed code from various disclosures from the RDC
- (vi) a sub-folder called **Census-PSID** containing replication materials for the Census/PSID work (also see README-Census-PSID in there)
- (vii) a file called **Guide_to_Code_for_Figures_Tables.pdf**, which provides a guide for where to find the code that constructs each figure and table in the paper.

To replicate the results, you will need Stata software, which must be purchased.

The following Stata commands must be installed (*using the code “ssc install [command name]”*):

- maptile
- spmap
- labutil
- sxpose
- gtools

To replicate the code, download and unzip the **REPLICATION_PACKAGE** folder. Note that, as described above, researchers must first download the build files from openicpsr-146361.

Then, reference “**Guide_to_Code_for_Figures_Tables.pdf**” which provides a guide for where to find the code that constructs each figure and table in the paper. The first column of the table in this document lists each figure and table number. The second column lists the names of the .xls files that contain the results disclosed from the RDC (for figures and tables that use RDC data). The third column provides information on how these disclosed .xls files were manipulated to create the tables and figures in the paper (either using **make_graphs_final.do** or **construct_tables_Final.do**, or with Excel2Latex). For figures and tables that do not rely on RDC data, the third column provides information on the code and data used. The fourth column contains the file paths and names of the do-files that were used inside the RDC for the analysis that generates each figure and table (located in the sub-folder **Manuscript_29108_DisclosedCode**).

To replicate the results from the Census and PSID, use the do-files in the subfolder **Census-PSID**.

Generating results using the RDC replication code and data takes about 2 weeks using the RDC computing environment. Generating results using the outside-RDC processing code takes about 1 minute in Stata v16.1 using a single processor on a computer with Intel® Core™ i7-8750H CPU @ 2.20 GHz with 8GB installed RAM.