

Readme: Programs for “Evaluating Public Programs with Close Substitutes: The Case of Head Start”

This paper uses data from the Head Start Impact Study (HSIS). Access to the HSIS data is restricted, so the data set cannot be provided. Researchers interested in accessing the data can submit applications to Research Connections. The application is available at:

<http://www.researchconnections.org/childcare/resources/19525>

The folder “program_archive.zip” contains the computer programs used to process the data and produce the estimates in the paper. The contents of this file are:

- setup: This folder contains the STATA programs used to process the raw data and construct the analysis data set, analysis_file.dta. It includes two subfolders:
 - do_files: This folder contains 5 setup programs. They should be run in the following order: tests.do, parents.do, centers.do, weights.do, combine.do
 - dictionaries: This folder includes a set of data dictionaries with variable definitions, referenced in the setup programs.
- stata: This folder contains one STATA program, analysis.do, which uses analysis_file.dta to compute the estimates in all tables in the article. The “Switches” section at the top of the file defines a set of macros that determine which part of the analysis is executed. The “Model” and “Restrictions” sections define macros that determine which variables are included and restrictions are imposed. The “two_step_choice” subsection of the program saves the file sitedata.csv, which is used in the Matlab section of the code.
- matlab: This folder contains Matlab programs that run the multinomial probit estimation routine, with or without group fixed effects. The estimate.m program is the master Matlab program that runs the estimation code. Estimate.m program pulls in “sitedata.csv” and calls all the other programs, and saves the file “output.csv”, which is called in the latter sections of the stata analysis file.
- bootstrap: This folder contains the programs used to obtain bootstrap standard errors, confidence intervals and p-values for two-step models in the paper. The key programs are:
 - bootstrap.m: This Matlab program draws bootstrap samples and reestimates the multinomial probit model on these samples. When executed from a shell it takes an argument labeled “i” which determines the random number seed and the names of files to be saved. This allows the trials to run in parallel. The results are saved in files with names of the form output_coreX_trialY.csv. All other Matlab programs in the bootstrap folder are called by bootstrap.m.
 - bootstrap_process.do: This STATA program pulls in the choice model estimates for each bootstrap trial and uses them to construct test statistics. The “Switches” section at the top determines the set of statistics that will be reported.