

DATA AND PROGRAMS

December 11, 2008

David Deming

This memo documents the data files and programs used to generate the estimates found in AEJ Applied 2008-0217, “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start.”

I obtained the source data from the Bureau of Labor Statistics National Longitudinal Survey of Youth 1979 (NLSY-79) and the Child Supplement File (CNLSY.) I used the public version of these data, which are available for download at:

http://www.nlsinfo.org/ordering/display_db.php3#NLSY79-Children-Young%20Adults

I extracted the source data using the NLS database investigator software, which can be found here:

<http://www.nlsinfo.org/dbgator/index.php3>

The main data file which should be used for replication is “data_Deming_2008_0217.dta”, in STATA format. The variables in this file are taken raw from the NLSY, but they are renamed for easy use. I have included a file called “variable_names_2008_0217.do” which contains the code for renaming nearly all of the variables I ever extracted from the NLSY database. Not all of them are used in the paper. Users who wish to replicate results from the original data collection stage can refer to this file for the original variables.

The file “Deming_2008_0217.do” contains code to run the analysis that generates Tables 1 through 5 in the paper. It is annotated throughout. This program produces text files named “Table 1”, “Table 2” etc. that will save in the default storage directory. In order to generate these files, you need to install two user-generated programs in STATA. Links to both are below:

tabexport program (for Table 1):

<http://ideas.repec.org/c/boc/bocode/s456715.html>

estout program (for Tables 2-5):

<http://ideas.repec.org/c/boc/bocode/s439301.html>

I have also included a .dta and .do file that can be used similarly to generate the benefit-cost analysis in Appendix I.