README FILE FOR THE REPLICATION OF:

Selection into Identification in Fixed Effects Models, with Application to Head Start

Doug Miller, Na'ama Shenhav, Michel Grosz

Directions:

1. Download the raw data (listed below).
2. Run 0*makepsiddata.do* to produce the PSID analysis data
3. Run the following do-files to produce the results in the paper:
   - *1psidresults.do* (estimates using the PSID analysis data)
     i. Note: this calls *convexhull.do*, which makes the convex hull graphs (Figures B.4, B.5)
     ii. For the p-score reweighting, see lines 118-210 (creation of weights) and 771-909 (implementation of two-step reweighting)
   - *1cnlsyresults.do* (estimates using the CNLSY data)
     i. For the p-score reweighting, see lines 1077-1159 (creation of weights) and 1170-1326 (implementation of two-step reweighting).
   - *2montecarlo.do* (Monte Carlo results in Table 5)

Description of datasets:

- **psid19682011.dta:** 1968-2011 PSID panels with renamed variables (but no other changes)
  - Code for renaming variables available on request
- **data_Deming_2008_0217.dta:** CNLSY79 data downloaded from replication files for Deming (2009)
- **litreview.dta:** Data on papers from the fixed effects literature review
- **pov_thresh_for_psidfe.dta:** Data on family-specific poverty thresholds for PSID.
  - For raw data, download "Poverty_Thresholds_Wide.dta" from ICPSR
- **psid_for_mcs:** subset of variables from the PSID to run Monte Carlo replications
- **cpi99:** CPI for converting earnings/income to 1999 dollars.
- **cps_1995weight:** Population-representative weights created from the 1995 CPS, following Garces, Thomas, and Currie (2002).