Replication Package for "Toward an Understanding of the Returns to Cognitive Skill Across Cohorts"

Authors: Judith Hellerstein, Sai Luo, and Sergio Urzúa

Contact: Sai Luo, luosaiecon@gmail.com

[A] Data Availability Statement

This replication package provides all necessary final datasets and code to reproduce the results presented in the paper, including the Online Appendix. Instructions for original data access inquiries are provided below.

[B] Data Provenance and Access

1. National Longitudinal Survey of Youth (NLSY) 1979 and 1997: The primary dataset for the main analysis.
Instructions: The final dataset used is provided in the replication package as "nlsy_final.dta." Anyone interested in constructing their own NLSY dataset can access the data via the NLS data investigator (https://www.nlsinfo.org/investigator). More information on the NLSY data is provided in the paper and on BLS website (https://www.bls.gov/nls/data-overview.htm).

2. National Education Longitudinal Study of 1988 (NELS88) and the Educational Longitudinal Study of 2002 (ELS02): Used for the analysis in Section 3.1 of the paper and Online Appendix C.
Instructions: The final dataset used is provided in the replication package as "nels_final.dta." Anyone interested in constructing their own NELS/ELS dataset can access the data via the NCES data lab (https://nces.ed.gov/datalab).

3. National Assessment of Education Progress (NAEP) long-term trend (LTT) assessments: Used for Table 5 of the paper.
Instructions: The statistics used in Table 5 are obtained directly from NAEP's report on LTT assessments (available at https://nces.ed.gov/nationsreportcard/ltt/).

[C] Software and Hardware Requirements

- Stata Version: The code was last tested and run on Stata 18 (SE/MP). We recommend using Stata 18 or newer for compatibility.
- Stata Packages: To run the analysis, several user-written packages are required. Please install them by running the following commands in the Stata command window:

```
ssc install grc1leg, replace
```

ssc install schemepack, replace

[D] Instructions for Replication

Each dofile in the replication package can be executed separately. There is no need to run files sequentially. Detailed instructions are provided below.

1. **nlsy_analysis.do**:
    - **Input:** NLSY dataset (data/nlsy_final.dta).
    - **Function:** Generate summary statistics, distribution figures, and regression tables using the NLSY dataset.
    - **Output:** Table 1, 2. Figure 1, 2. Online Appendix Table A1, A2, E1, H2, H3, G1. Online Appendix Figure A1, A2, B1, B2, E1.
2. **nlsy_bootstrap.do**:
    - **Input:** NLSY dataset (data/nlsy_final.dta).
    - **Function:** Calculate bootstrap statistics for the distribution differences between cohorts.
    - **Output:** Statistics used in Table 3.
3. **nlsy_yitzhaki.do**:
    - **Input:** NLSY dataset (data/nlsy_final.dta).
    - **Function:** Conduct the Yitzhaki decomposition using the NLSY dataset.
    - **Output:** Figure 8, 9, 10. Online Appendix Figure A3, E2, E3, E4, F1, F2. Counterfactual OLS estimates.
4. **nlsy_yitzhaki_nonlinear.do**:
    - **Input:** NLSY dataset (data/nlsy_final.dta).
    - **Function:** Conduct the Yitzhaki decomposition allowing for nonlinear returns, as introduced in Online Appendix G.
    - **Output:** Figure G1.
5. **nels_analysis.do**:
    - **Input:** NELS/ELS dataset (data/nels_final.dta).
    - **Function:** Generate summary statistics, distribution figures, and regression tables using the NELS/ELS dataset.
    - **Output:** Table 4. Figure 3. Online Appendix Table A3, C1.
6. **nels_yitzhaki.do**:
    - **Input:** NELS/ELS dataset (data/nels_final.dta).
    - **Function:** Conduct the Yitzhaki decomposition using the NELS/ELS dataset.
    - **Output:** Online Appendix Figure C1, C2.
7. **simulation.do**:
    - **Input:** N/A.
    - **Function:** Simulate data and conduct the Yitzhaki decomposition using the simulated data.

- **Output:** Figure 4, 5, 6, 7.

How to Run

1. **Install** all required Stata packages listed above.
2. **Set the path:** Open each .do script and set the global path variable at the top to the location of the replication package folder on your computer.
3. **Run scripts:** Execute the scripts separately.

[E] Data Dictionary

The **NLSY dataset** (data/nlsy_final.dta) include the following variables:

Variable Name	Variable Label
pubid	NLSY ID
race	Race and Ethnicity: Hispanic=1, Black Non-Hispanic=2, White Non-Hispanic=3
hgc_mom	Mother's Highest Grade Completed
bothparents_age14	Living with Both Parents as Child
sample	Sample: NLSY-79 or NLSY-97
weight	BLS Sample Weight
soc_nlsy2_std	Social score (Deming)
noncog_std	Noncog score (Deming)
female	Gender: Female=1, Male=0
avg_age	Avg. Age over non-missing wage observations
ln_wage	100 * Ln(Avg. Wages over 25-39)
ln_wage_noschl3133	100 * Ln(Avg. Wages over 31-33)
ar	Arithmetic Reasoning score (Theta)
ar_se	S.E. of Arithmetic Reasoning score (Theta)

Variable Name	Variable Label
wk	Word Knowledge score (Theta)
wk_se	S.E. of Word Knowledge score (Theta)
pc	Paragraph Comprehension score (Theta)
pc_se	S.E. of Paragraph Comprehension score (Theta)
group	Group: Cohort*Gender*Race
group2	Group: Gender*Race
aafqt	AAFQT score (Altonji et al.)
ar_altonji	Arithmetic Reasoning score (Altonji et al.)
wk_altonji	Word Knowledge score (Altonji et al.)
pc_altonji	Paragraph Comprehension score (Altonji et al.)
no_altonji	Numerical Operation score (Altonji et al.)
aafqt_bin3	AAFQT score bin
educ	Edu
educ_group	RECODE of educ
educ_hs	high school graduate (incl. GED)
educ_sc	some college
educ_co	bachelor and above
res_aafqt	Residualized AAFQT score (by Noncog, Social & Educ)
abovep50	Dummy: AAFQT above median
belowp50	Dummy: AAFQT below median
aafqt_abovep50	AAFQT * (AAFQT>p50)

Variable Name	Variable Label
aafqt_belowp50	AAFQT * (AAFQT<p50)
aafqt_avg_age	AAFQT * Average Age
aafqt_sample	AAFQT * NLSY97
soc_nlsy2_std_sample	Social * NLSY97
noncog_std_sample	Noncog * NLSY97
educ_sample	Edu * NLSY97
aafqt_belowp50_sample	AAFQT * (AAFQT<p50) * NLSY97
aafqt_abovep50_sample	AAFQT * (AAFQT>p50) * NLSY97
abovep50_sample	(AAFQT>p50) * NLSY97
belowp50_sample	(AAFQT<p50) * NLSY97
aafqt_avg_age_sample	AAFQT * Average Age * NLSY97

The **NELS/ELS dataset** (data/nels_final.dta) include the following variables:

Variable Name	Variable Label
id	Student ID
female	Gender: Female=1
ln_hrinc	100 * Ln(hourly earnings 7yrs after highsch graduation
wt	Sample Weight
math	Math (Grade 12)
math_g10	Math (Grade 10)
sample	Sample: NELS88=0, ELS02=1

Variable Name	Variable Label
group	Group: Cohort*Gender*Race
group2	Group: Gender*Race
math_sample	Math * ELS02