Supplemental Appendix: The Income-Achievement Gap and Adult Outcome Inequality

Eric Nielsen*
Federal Reserve Board

October 2022

*I thank Derek Neal, Gary Becker, Ali Hortaçsu, Sandy Black, Armin Rick, and seminar participants at the University of Chicago, the Federal Reserve Board, and the NBER Summer Institute for helpful comments and feedback. The views and opinions expressed in this paper are solely those of the author and do not reflect those of the Board of Governors or the Federal Reserve System.

Contact: Division of Research and Statistics, Board of Governors of the Federal Reserve System, Mail Stop 97, 20th and C Street NW, Washington, D.C. 20551. eric.r.nielsen@frb.gov. (202) 872-7591.

Disclosure Statement: The author has nothing to disclose.

Data Availability: The paper uses data publicly available at https://www.nlsinfo.org/ and http://www.fabianlange.ca/data.html.

1 Construction of Lifetime Income

Calculating pdv_labor for each respondent is complicated by four forms of missing data. First, not every respondent has their income recorded in a given year. Second, not every survey respondent is in the labor force in a given year. Third, interviews were conducted biennially after 1994, so income data is missing for odd-numbered years between 1994 and 2014. Fourth, the NLSY79 respondents can only be observed through ages 49-51, while the NLSY97 respondents can be seen only through ages 31-33. I address the first two kinds of missing data through the imputation rules described in Section 2 of the main paper (baseline, pessimistic, optimistic). I address the third form of missing data by linearly interpolating wage income for the odd-numbered years between 1995 and 2013 after applying one these imputation rules. I address the fourth form of missing data by using education-specific age-earnings profiles to extrapolate observed labor income profiles through to retirement in the NLSY79 and by not anchoring on NLSY97 incomes.

I build the various pdv_labor estimates using NLSY79 variables that measure total annual labor earnings and total hours worked across all jobs. I do not adjust for truncation, although the quantile regression-based estimates should not be sensitive to either truncation or the presence of outliers. For the "flexible" measures used in the robustness analysis, I estimate wage rates by dividing annual earnings by annual hours worked. I perform this division after imputing missing wages using either the optimistic or pessimist rule outlined in Section 2 of the main paper and after filling in the alternate-year data after 1994 using the linear interpolation procedure described above. This division results in a few unrealistically high wage estimates; I drop observations with implied wage rates above \$500 (in 2015\$); my estimates are not sensitive to this particular threshold choice. I convert hourly wages to annual earnings by assuming a full-time year of work consists of 2,087 hours, the U.S. Federal Government's Office of Personnel Management assumption for full-time work.

The age-earnings profiles of workers with different education levels are not simply log-level shifts of each other. Highly educated workers experience much more rapid wage income growth in percentage terms between the ages of 20 and 50. To account for these differences, I use the 2005 American Community Survey (ACS) to construct synthetic age-earnings profiles for men with different education levels. I focus on men because they have high rates of labor force participation, so their earnings should be less affected by selection. I use the mean earnings of men in several age buckets (18-24, 25-34, 35-44, 45-54, 55-64, and 65+) crossed with several education categories (<high school, high school, and college+).

I use data bucketed into 5- and 10-year increments. Let $m_{e,a,a+1}$ be the slope of the earnings line connecting the labor income in age buckets a and $a+1$ for education category $e \in \{< \text{high school}, \text{ high school}, \text{ college+}\}$, and let $\tilde{w}_{i,t,k}$ be the (imputed) annual wage income for respondent i in survey wave t using imputation rule $k \in \{\text{pess}, \text{ opt}\}$. I assume that each NLSY respondent will work until age 65 and then retire. I calculate the expected annual wage income of i in year 2014, $\hat{w}_{i,2014,k}$ using a regression of $w_{i,2014,k}$ on time trends and prior-year income estimates. I estimate $\hat{w}_{i,2014,k}$ in this way because I extrapolate out of the sample using this last "observed" wage and I want to ensure that what I take as the base for the extrapolation is not driven by missing data or transient shocks in the final period. I assume that i's yearly income increases and decreases from $\hat{w}_{i,2014,k}$ between the ages of 49 and 65 in accordance with the slopes $\{m_{e(i),a,a+1}\}$, where $e(i)$ is the education level of i. Putting all of this together, the estimated pdv_labor of a youth who was 15 (with analogous expressions for ages 16 and 17) at the start of the NLSY79 is given by

$$\begin{aligned} PDV_{i,k} \equiv & \underbrace{\sum_{t=0}^{t=35} (0.95)^t \, \tilde{w}_{i,t,k}}_{\text{observed/imputed}} + \underbrace{\hat{w}_{i,2014,k} \sum_{j=1}^{5} (0.95)^{35+j} \left(1 + j m_{e(i),35,45}\right)}_{\text{projected, age 51-55}} \\ + & \underbrace{\hat{w}_{i,2014,k} \left(1 + 10 m_{e(i),35,45}\right) \sum_{j=1}^{10} (0.95)^{41+j} \left(1 + j m_{e(i),45,55}\right)}_{\text{projected, age 56-65}} . \end{aligned}$$

Constructing the school completion variables is much simpler. Both NLSY surveys record the highest grade completed for each respondent in each survey wave. Using these grade-completion variables, I construct a new variable for each survey wave t equal to the highest grade completed observed in any wave up to and including t. Occasionally, the highest grade completed for a respondent will decrease between one survey and the next. These data are difficult to interpret; my fill-in rule assumes that the lower value is incorrect. I only use the grade-completion variables up to 14 years after the start of the survey. Few respondents change their education status after age 30, so this restriction should have little effect on my estimates.

2 Stochastic Dominance Test Details

The Barrett and Donald (2003) test uses in combination multiple KS-like statistics that have known distributions under the null that one distribution weakly dominates the other. In more detail, given two independent samples of sizes N and M from populations X and Y with the same bounded support, Barrett and Donald show the statistic $\hat{S}_1 = \left(\frac{NM}{N+M}\right)^{\frac{1}{2}} \sup_z \left(\hat{F}_Y(z) - \hat{F}_X(z)\right)$ has probability $exp\left(-2\hat{S}_1^2\right)$ under the null $H_0 : \; F_Y(z) \leq F_X(z) \forall z$. To test whether the NLSY79 high-income score distribution dominates its NLSY97 counterpart one simply tests two nulls: $H_0 : F_{H,79}(s) \leq F_{H,97}(s), \forall s$ and $\tilde{H}_0 : F_{H,97}(s) \leq F_{H,97}(s), \forall s$ at some level α. There are four possible outcomes. If both nulls are rejected, then the cdfs cross, while if neither null is rejected, the cdfs are equal. If H_0 is not rejected while $\tilde{H}_0$ is, then $F_{H,79} \succ F_{H,97}$, while if $\tilde{H}_0$ is not rejected and H_0 is, $F_{H,97} \succ F_{H,79}$.

3 Monte Carlo Details

The first Monte Carlo explores the apparent difference in power between the ordinal method and the simple difference-in-difference (DiD) method which consists of just the change in the difference in mean achievement between the top and bottom income quintiles. In particular, I set the change in the low-income means (Δ_L) and high-income means (Δ_H) such that the true gap change is unambiguously negative. (i.e. $\Delta_L \geq 0 \; \wedge \; \Delta_H \leq 0$ with at least one inequality strict). I assume that each income group in each time period is normally distributed, all with the same variance set equal to the variance observed in the Reardon age 15-17 data. I then set the sample size for each survey/income bucket equal 600, close to the average group size in the actual data.[1] I set the significance level of each pairwise FOSD test as well as the significance level of

[1] An analogous Monte Carlo where each group's sample size is set equal to the corresponding (unweighted) size in the Reardon age 15-17 data produces very similar results.

the DiD test so that the overall rejection rate when the null is true is 0.05 for both methods (see the last row of Table 1).

Table 1

Monte Carlo Estimates – Ordinal vs. Mean Differences

NLSY97-79 Mean Differences (sd)		Share Rejected		
Low-Income (sd)	High-Income (sd)	Ordinal	DiD	Disagreement
0.15	-0.15	0.79	0.94	0.18
0.1	-0.1	0.45	0.65	0.28
0.05	-0.05	0.19	0.22	0.17
0.15	0	0.54	0.41	0.28
0.1	0	0.27	0.22	0.21
0.05	0	0.12	0.09	0.12
0	0	0.05	0.05	0.07

Note: Each group/period's sample size = 600. The low-income and high-income columns give the mean differences in achievement (NLSY97-NLSY79) used in the data generation process. The share rejected columns give the fraction of simulated samples for which the ordinal (FOSD) and DiD (mean difference) methods reject the null of no change in the high-low achievement gap. The "Disagreement Share" column gives the fraction of samples in which the DiD and ordinal methods disagree, with one rejecting the relevant null and the other not.

The second Monte Carlo explores the apparent difference in power between the ordinal method and Reardon's method. Compared to the first Monte Carlo, the data generation step here is more involved, since Reardon's method uses the all income percentiles, rather than just the top and bottom quintiles. In detail, the steps of the Monte Carlo are:

1. Start with the Reardon data. Draw a bootstrap sample with replacement, subset to individuals aged 15-17, and estimate the deciles of parental income in both surveys.

2. Simulated NLSY79: Create a normal, iid random sample of reading scores of size N in each income decile with a mean equal to the corresponding mean from (1) and a variance of one. For each decile, set the income decile lower bound, upper bound, and midpoint using the corresponding data from (1).

3. Simulated NLSY79: Follow an identical procedure to the NLSY79 data. The only difference is that the mean achievement of each decile is given by the corresponding mean from the NLSY79 plus an adjustment factor which depends on the particular Monte Carlo being run.

4. Apply Reardon's method and my ordinal method to the simulated data. Repeat 2,000 times.

Table 2

Monte Carlo Estimates – Ordinal vs. Mean Differences and Ordinal vs. Reardon (2011) Method

Group	N	Ord	DiD	Ord/DiD	Reardon	Ord/Reardon	DiD Gap-Change	Reardon Gap-Change
1	2,400	0.15	0.17	0.15	0.11	0.16	-0.125	-0.123
1	10,000	0.38	0.49	0.24	0.07	0.34	-0.124	-0.120
2	2,400	0.22	0.27	0.19	0.17	0.22	-0.18	-0.18
2	10,000	0.60	0.78	0.25	0.20	0.46	-0.17	-0.18
3	2,400	0.16	0.13	0.15	0.09	0.17	-0.10	-0.10
3	10,000	0.47	0.35	0.28	0.05	0.44	-0.10	-0.10
4	2,400	0.25	0.19	0.21	0.13	0.24	-0.14	-0.14
4	10,000	0.74	0.57	0.28	0.11	0.64	-0.14	-0.14
5	2,400	0.25	0.19	0.21	0.20	0.24	-0.14	-0.19
5	10,000	0.74	0.57	0.28	0.26	0.53	-0.14	-0.19
6	1,200	0.14	0.18	0.16	0.21	0.22	-0.20	-0.22
6	2,400	0.26	0.34	0.22	0.28	0.26	-0.20	-0.22
6	10,000	0.71	0.87	0.21	0.36	0.44	-0.20	-0.22
7	600	0.26	0.32	0.21	0.34	0.26	-0.38	-0.37
7	1,200	0.35	0.54	0.28	0.46	0.33	-0.37	-0.36
7	2,400	0.67	0.83	0.21	0.63	0.29	-0.38	-0.37
7	10,000	1.00	1.00	0.00	0.89	0.11	-0.37	-0.36
8	1,200	0.13	0.10	0.13	0.09	0.15	-0.14	-0.11
8	2,400	0.25	0.19	0.21	0.09	0.24	-0.14	-0.11
8	10,000	0.74	0.57	0.28	0.05	0.69	-0.14	-0.11
8	200,000	0.98	1	0.02	0.02	0.96	-0.14	-0.11
9	1,200	0.13	0.10	0.13	0.11	0.23	-0.14	0.13
9	2,400	0.25	0.19	0.21	0.11	0.36	-0.14	0.13
9	10,000	0.74	0.57	0.28	0.09	0.75	-0.14	0.13
9	200,000	0.98	1	0.02	0.05	0.93	-0.14	0.13
10	2,400	0.05	0.06	0.08	0.03	0.07	0.00	0.00

Note: The N shown is the the whole sample. The number of observations per decile/sample bucket is given by $N/20$. The Ord/DiD column shows the share of simulations for which ordinal and DiD methods disagree, while the Ord/Reardon column shows the corresponding share for the ordinal and Reardon methods. The DiD Gap-Change column shows the average DiD estimate, while the Reardon Gap-Change column shows the average Reardon estimate of the 90/10 gap change. The mean difference vectors are as follows: Group 1 = {0.075, 0.05, 0.025, 0.025, 0, 0, -0.025, -0.025, -0.05, -0.075}; Group 2 = {0.1, 0.075, 0.05, 0.05, 0.025, -0.025, -0.05, -0.05, -0.075, -0.1}; Group 3 = {0, 0, 0, 0, 0, -0.025, -0.05, -0.075, -0.10, -0.10}; Group 4 = {0, 0, 0, 0, 0, -0.025, -0.075, -0.1, -0.125, -0.15}; Group 5 = {0, 0, 0, 0.5, 0.5, -0.025, -0.075, -0.1, -0.125, -0.15}; Group 5 = {0, 0, 0, 0.5, 0.5, -0.025, -0.075, -0.1, -0.125, -0.15}; Group 6 = {0.10, 0.10, 0.10, 0.10, 0.10, -0.1, -0.1, -0.1, -0.1, -0.1}; Group 7 = {0.225, 0.15, 0.075, 0.05, 0.025, -0.025, -0.05, -0.075, -0.15, -0.225}; Group 8 = {0, 0, -0.10, -0.15, -0.2, -0.025, -0.075, -0.10, -0.125, -0.15}; Group 9 = {0, 0, -1, -1, -1, -0.025, -0.075, -0.10, -0.125, -0.15}; Group 10 = {0, 0, 0, 0, 0, 0, 0, 0, 0, 0}.

4 Additional Empirical Results

Table 3

Additional High- and Low-Income FOSD Tests

Math	NLSY79-97 Comparison	Age 16	Cutoffs First	Regression Adj.	Non-Crosswalked Scores
Full Sample	Low Income	ambiguous	ambiguous	79 dominates	ambiguous
	High Income	equal	79 dominates	79 dominates	79 dominates
	High-Low Gap	ambiguous	ambiguous	ambiguous	ambiguous
Non-Black, Non-Hisp.	Low Income	97 dominates	equal	79 dominates	equal
	High Income	equal	79 dominates	79 dominates	79 dominates
	High-Low Gap	decrease	decrease	ambiguous	decrease
Black	Low Income	79 dominates	79 dominates	79 dominates	79 dominates
	High Income	97 dominates	equal	equal	97 dominates
	High-Low Gap	increase	increase	increase	increase
Reading					
Full Sample	Low Income	97 dominates	97 dominates	equal	equal
	High Income	equal	79 dominates	79 dominates	79 dominates
	High-Low Gap	decrease	decrease	decrease	decrease
Non-Black, Non-Hisp.	Low Income	97 dominates	97 dominates	equal	equal
	High Income	equal	79 dominates	79 dominates	equal
	High-Low Gap	decrease	decrease	decrease	no change
Black	Low Income	equal	79 dominates	equal	equal
	High Income	97 dominates	equal	equal	equal
	High-Low Gap	increase	increase	no change	no change
AFQT					
Full Sample	Low Income	97 dominates	ambiguous	79 dominates	79 dominates
	High Income	79 dominates	79 dominates	79 dominates	79 dominates
	High-Low Gap	decrease	ambiguous	ambiguous	ambiguous
Non-Black, Non-Hisp.	Low Income	97 dominates	equal	equal	equal
	High Income	79 dominates	79 dominates	79 dominates	79 dominates
	High-Low Gap	decrease	decrease	decrease	decrease
Black	Low Income	79 dominates	79 dominates	79 dominates	79 dominates
	High Income	97 dominates	equal	equal	97 dominates
	High-Low Gap	increase	increase	increase	increase

Note: Tests run at $\alpha = 0.05$. The "Age 16" column uses only age-16 respondents. The "Cutoffs First" column subsets on race only after defining the high- and low-income categories. The "Regression Adj." column uses test-score residuals after regressing out age, sex, urban/rural status, parental education, and race. The "Non-Crosswalked" column uses non-crosswalked test scores, standardized separately by year. "79 dominates" and "97 dominates" mean that indicated distribution dominates, while "ambiguous" means that the distributional comparison or gap change is ambiguous ordinally. "Equal" means that the null of distributional equality cannot be rejected, while "no change" means that the indicated achievement gap did not change.

Table 4

Lifetime Income Mean-Anchored Estimates, Alternative Income Outcomes and Models

Log Gap Changes

Model	Income Measure	Math	Reading	AFQT
cubic	baseline	-0.01	-0.09**	-0.06
		(0.04)	(0.04)	(0.04)
local linear	baseline	-0.02	-0.08*	-0.06
		(0.04)	(0.04)	(0.04)
equation (1)	pessimistic-observed	-0.03	-0.09**	-0.07*
		(0.04)	(0.04)	(0.04)
equation (1)	pessimistic-full	-0.03	-0.07**	-0.06**
		(0.03)	(0.03)	(0.03)
equation (1)	optimistic-observed	-0.04	-0.08**	-0.07**
		(0.03)	(0.04)	(0.04)
equation (1)	optimistic-full	-0.02	-0.04**	-0.04**
		(0.02)	(0.02)	(0.02)

Level Gap Change ($1,000s)

Model	Income Measure	Math	Reading	AFQT
cubic	baseline	-25	-41***	-38**
		(15)	(14)	(16)
local linear	baseline	-21	-39***	-35**
		(15)	(14)	(15)
equation (1)	pessimistic-observed	-17	-30***	-28**
		(12)	(12)	(12)
equation (1)	pessimistic-full	-16	-31***	-28**
		(11)	(11)	(12)
equation (1)	optimistic-observed	-30	-52**	-48**
		(22)	(21)	(23)
equation (1)	optimistic-full	-22	-40**	-37**
		(17)	(16)	(17)

Note: None of the estimates in this table adjust for test score reliability. Standard errors based on 1,000 bootstrap iterations shown in parentheses. Estimates for alternative lifetime income measures are based on anchored scales from regressions of the form given by equation (1) of the main paper. For these income measures, "pessimistic" and "optimistic" refer to pessimistic and optimistic imputations of missing income data, respectively. The "observed" measures use observed annual earnings to construct pdv_labor, while the "full" measures use full income. The "cubic" models for log(pdv_labor) use a cubic polynomial in s instead of a linear term in equation (1) of the main paper. The "local linear" estimates based on local linear regressions (bandwidth=0.5 sd) estimated on non-black, non-Hispanic men. *** p<0.01, ** p<0.05, * p<0.1.

Table 5

School Completion Mean-Anchored Estimates, Alternative Models and Anchor Surveys

High School

Model	Anchor Survey	Math	Reading	AFQT
cubic	NLSY79	-0.004	-0.025*	-0.018
		(0.014)	(0.014)	(0.015)
local linear	NLSY79	-0.022	-0.045***	-0.043**
		(0.018)	(0.017)	(0.018)
local linear interacted	NLSY79	-0.004	-0.019	-0.016
		(.013)	(.0149)	(.015)
cubic	NLSY97	-0.010	-0.029***	-0.025**
		(0.012)	(0.011)	(0.012)
local linear	NLSY97	-0.011	-0.036***	-0.034***
		(0.012)	(0.011)	(0.012)
local linear interacted	NLSY97	-0.006	-0.028***	-0.023**
		(0.011)	(0.011)	(0.011)

College

Model	Anchor Survey	Math	Reading	AFQT
cubic	NLSY79	-0.023	-0.048***	-0.045**
		(0.022)	(0.018)	(0.023)
local linear	NLSY79	-0.030	-0.059***	-0.056***
		(0.021)	(0.018)	(0.021)
local linear interacted	NLSY79	-0.023	-0.051***	-0.048**
		(0.022)	(0.017)	(0.022)
cubic	NLSY97	-0.026	-0.056***	-0.051**
		(0.022)	(0.019)	(0.021)
local linear	NLSY97	-0.034	-0.068***	-0.064***
		(0.023)	(0.021)	(0.022)
local linear interacted	NLSY97	-0.026	-0.055***	-0.051**
		(0.022)	(0.019)	(0.021)

Note: None of the estimates in this table adjust for test score reliability. Standard errors based on 1,000 bootstrap iterations shown in parentheses. The "cubic" models use a cubic polynomial in s instead of a linear term in the baseline probit model. The "local linear" estimates are based on local linear regressions (bandwidth=0.5 sd) estimated on non-black, non-Hispanic men. The "local linear interacted" estimates are based on separate local linear regressions (bandwidth=0.5 sd) for each group defined by the interaction of indicators for Black and female. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

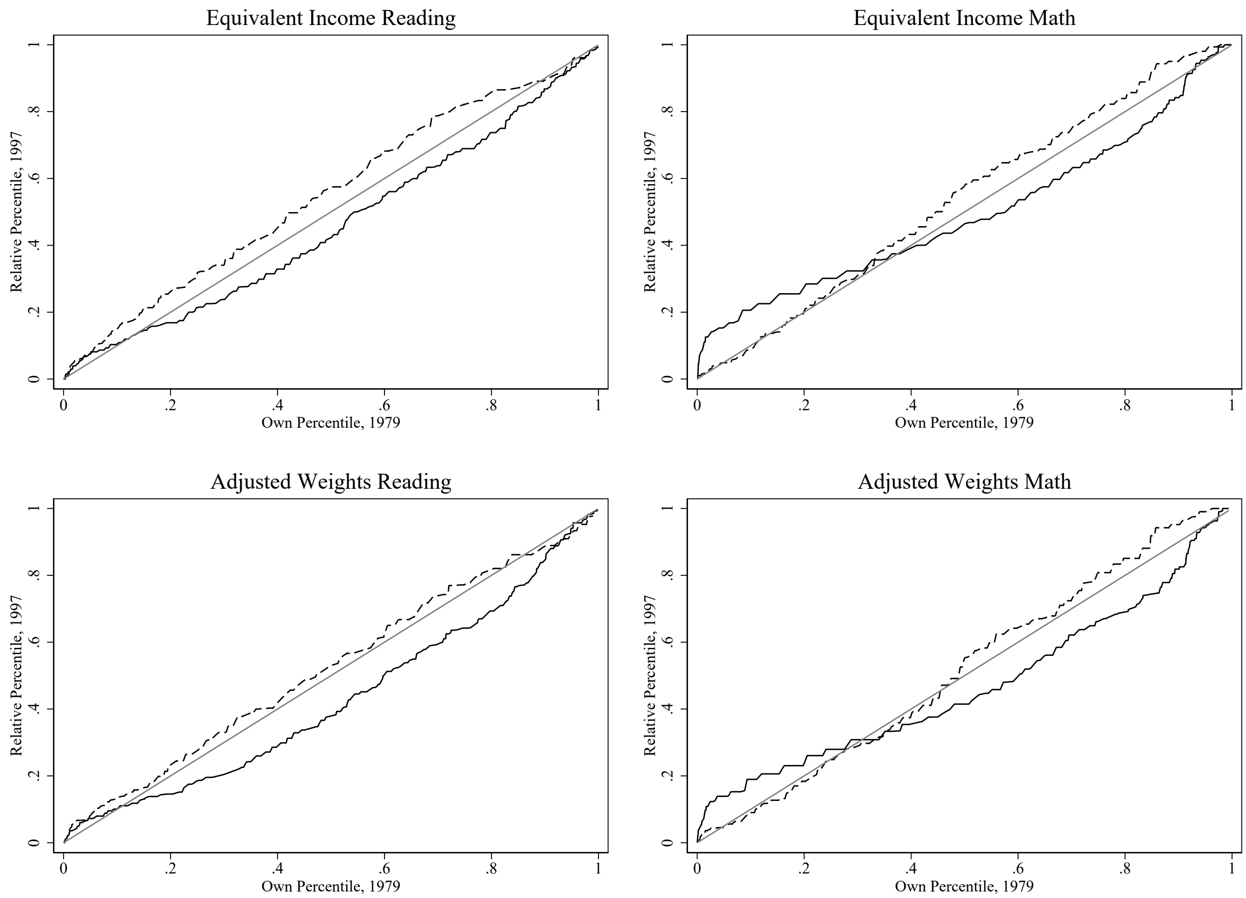


Figure 1

Full Sample Percentile-Percentile Curves, Alternative Income Definitions and Sample Weights

Note: Figures compare low- (high-) income youth in the NLSY79 to low- (high-) income youth in the NLSY97. Dotted 45-degree line of equality plotted for reference. Curves estimated using the empirical high- and low-income score cdfs.

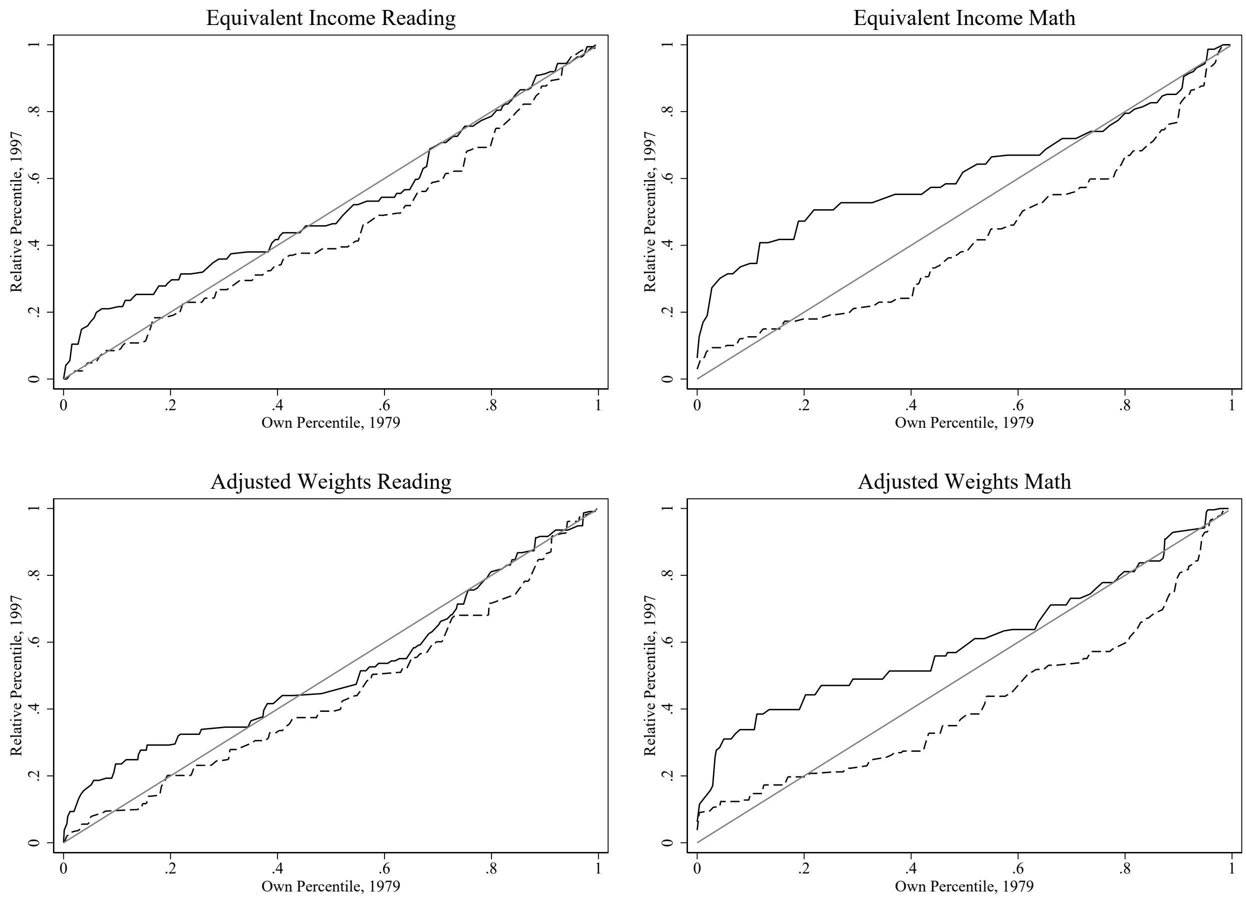
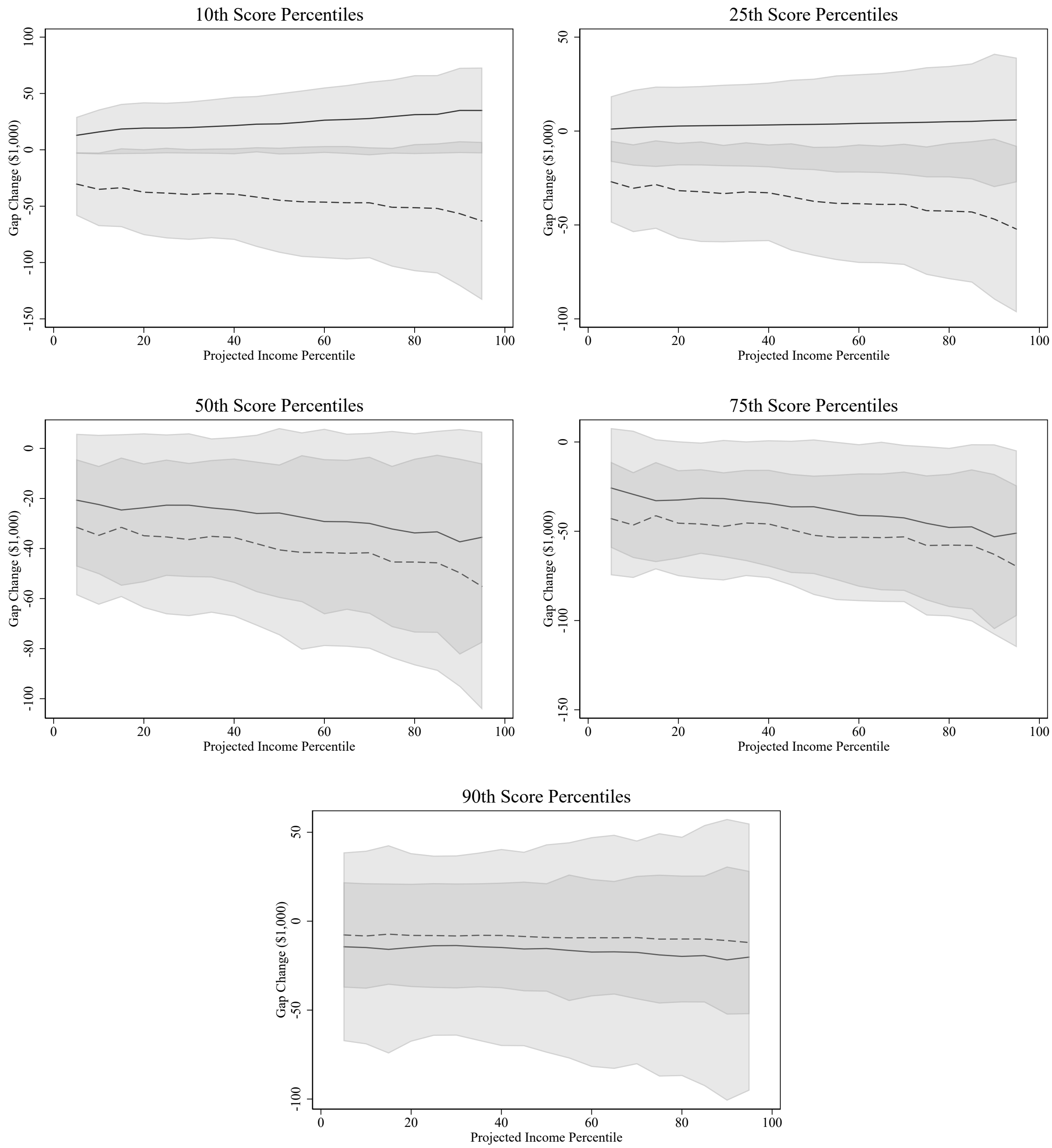


Figure 2

Black-only Percentile-Percentile Curves, Alternative Income Definitions and Sample Weights

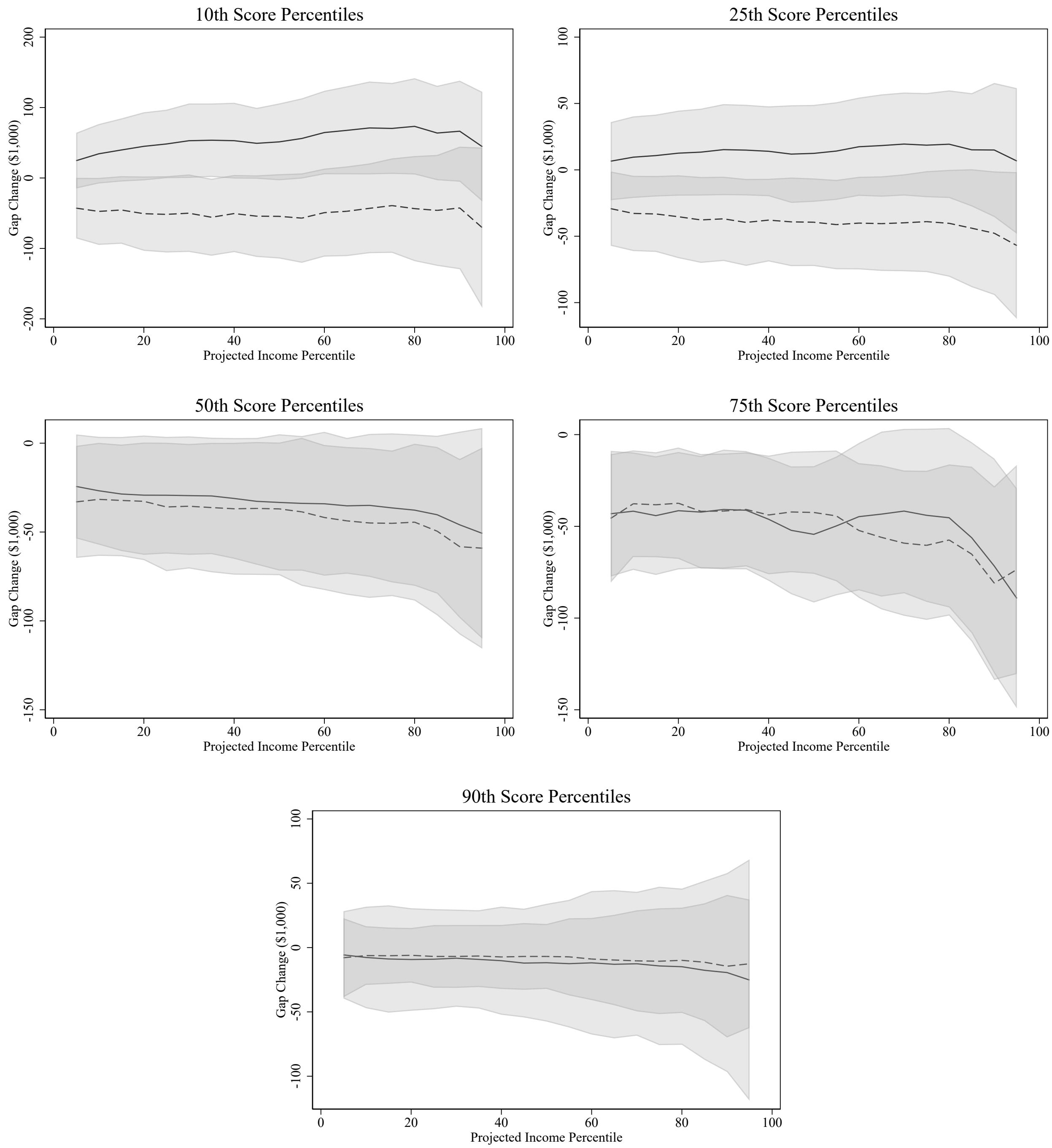
Note: Figures compare low- (high-) income youth in the NLSY79 to low- (high-) income youth in the NLSY97. Dotted 45-degree line of equality plotted for reference. Curves estimated using the empirical top income quintile (high) and bottom income quintile (low) score cdfs.

Figure 3: Math and Reading Income-Anchored Changes, Various Score Percentiles, Non-black, Non-Hispanic Men



Note: 95% bootstrapped confidence intervals based on 1,000 bootstrap iterations shown. Estimates based on quantile regressions estimated on non-black, non-Hispanic men of the form $\log(\text{pdv\_labor})^{(\tau)} = \alpha^{(\tau)} + \beta^{(\tau)} s + \gamma_2^{(\tau)} \text{age dummies} + \varepsilon^{(\tau)}$ for $\tau \in \{0.05, 0.10, \ldots, 0.95\}$.

Figure 4: Distributional Math and Reading Income-Anchored Changes, Cubic Models



Note: 95% bootstrapped confidence intervals based on 1,000 bootstrap iterations shown. Estimates based on quantile regressions of the form $\log(\text{pdv\_labor})^{(\tau)} = \alpha^{(\tau)} + \beta_1^{(\tau)} s + \beta_2^{(\tau)} s^2 + \beta_3^{(\tau)} s^3 + \gamma_2^{(\tau)} \text{age dummies} + \gamma_2^{(\tau)} \text{female} + \gamma_3^{(\tau)} \text{black} + \gamma_4^{(\tau)} (\text{female} \times \text{black}) + \varepsilon^{(\tau)}$ for $\tau \in \{0.05, 0.10, \ldots, 0.95\}$.

References

Barrett, G. and Donald, S. (2003). Consistent Tests for Stochastic Dominance. *Econometrica*, 71:71–104.

Reardon, S. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances.*