Online Resource 1

“Growing Wealth Gaps in Education”

1 Sample Attrition

Here, I describe patterns and potential consequences of sample attrition. In particular, I focus on how the cohorts studied here were affected by the PSID sample reduction in 1997 and assess the sensitivity of the trends described in the paper to the differential impact of this sample reduction.

Table S1: Sample Structure

		Baseline		Outcome (age 20)		Outcome (age 25)	
Cohort	Birth Yr	Age	Survey Yr	Year	Survey Yr	Year	Survey Yr
1970s	1970	14	1984	1990	1990	1995	1995
	1971	13	1984	1991	1991	1996	1996
	1972	12	1984	1992	1992	1997	1997
	1973	11	1984	1993	1993	1998	1999
	1974	10	1984	1994	1994	1999	1999
	1975	14	1989	1995	1995	2000	2001
	1976	13	1989	1996	1996	2001	2001
	1977	12	1989	1997	1997	2002	2003
	1978	11	1989	1998	1999	2003	2003
	1979	10	1989	1999	1999	2004	2005
1980s	1980	14	1994	2000	2001	2005	2005
	1981	13	1994	2001	2001	2006	2007
	1982	12	1994	2002	2003	2007	2007
	1983	11	1994	2003	2003	2008	2009
	1984	10	1994	2004	2005	2009	2009
	1985	14	1999	2005	2005	2010	2011
	1986	13	1999	2006	2007	2011	2011
	1987	12	1999	2007	2007	2012	2013
	1988	11	1999	2008	2009	2013	2013
	1989	10	1999	2009	2009	2014	2015

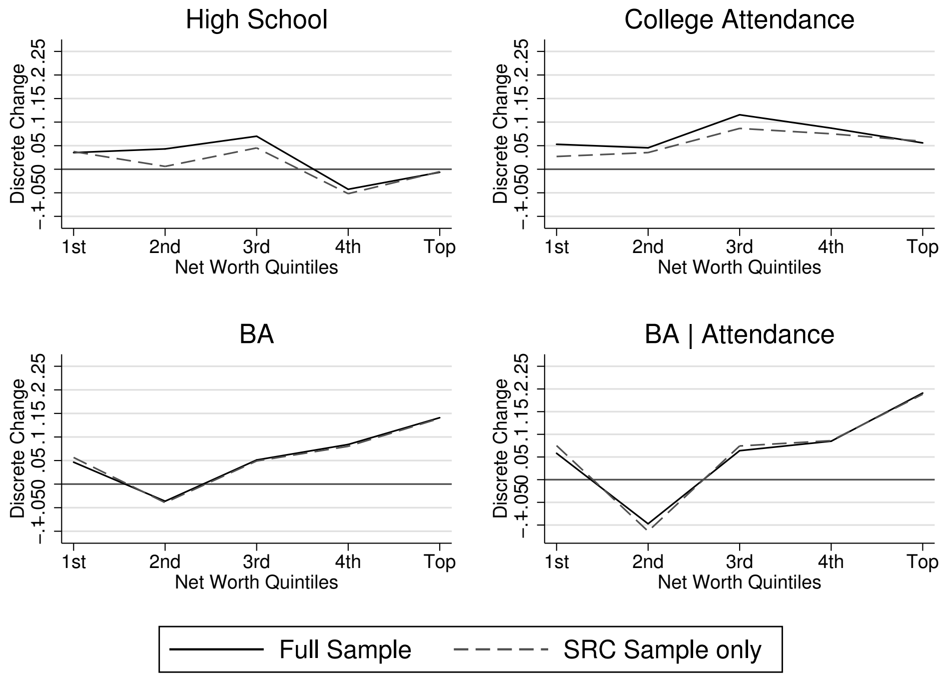
Due to funding shortfalls in the late 1990s, the PSID was forced to reduce its sample by more than 2,000 families between the 1996 and the 1997 survey wave. It did so chiefly by dropping family lineages with low sampling weights, resulting in much more substantial cuts to its initial oversample of low-income African-American families (the Survey of Economic Opportunity, or SEO, sample) than its population-representative sample (the Survey Research Center, or SRC, sample). The two cohorts analyzed here are affected by this drop in different ways, as shown in Table S1: The great majority of the 1970 birth cohort is unaffected at baseline (when their parental wealth is measured) as well as at the age 20 (when their high school attainment and college entry is assessed). In contrast, the 1980 birth cohort is affected by the sample drop at both measurement points: Educational outcomes at age 20 cannot be observed for any cohort members from families affected by the sample drop (marked in gray) and, in fact, among those born born in the second half of the 1980s, those cases are already lost at baseline (since their parental wealth is measured in 1999, i.e. after the sample drop; marked in dark gray). As a result – and as displayed in Table S2 – the unweighted attrition rate between baseline and age 20 is slightly higher among the 1980s birth cohort, largely driven by the higher attrition of the SEO sample through the sample reduction. The attrition pattern is different at age 25, when most of the 1970s cohort is also affected by the sample reduction, reflected in the higher loss of SEO sample members.[1]

Table S2: Sample Attrition (Unweighted)

	Analytic sample at			Percent of sample lost at	
	Baseline (1)	Age 20 (2)	Age 25 (3)	Age 20 (2) vs. (1)	Age 25 (3) vs. (1)
Born 1970s	**3,283**	**2521**	**1824**	**23.2%**	**44.4%**
Main/SRC	1,617	1,345	1,163	16.8%	28.1%
Main/Census	1,666	1,176	661	29.4%	60.3%
Born 1980s	**3,806**	**2832**	**2582**	**25.6%**	**32.2%**
Main/SRC	1,974	1,706	1,547	13.6%	21.6%
Main/Census	1,832	1,126	1,035	38.5%	43.5%

[1] The attrition rate of the SEO sample between baseline and age 25 among the 1970s cohort appears to not only catch up to but surpass that of the 1980s cohort, which should be understood as a result of the latter already being affected by the sample drop at baseline.

Figure S1: Trends in Wealth Gaps by Sample



In principle, the sampling weights provided by the PSID should compensate for the differential representation of the SEO sample among the two cohorts compared here, much like they do for any other analysis of PSID data that includes the SEO sample. Nevertheless, I also replicate one of the main analyses of this paper – the assessment of cohort trends in education gaps (see Figure 2 in paper) – based on the SRC sample only. Figure S1 directly compares the trend in the wealth gap for each educational level when based on the full sample (black line) and when based on only the SRC sample (gray dashed line). The differences are negligible for the analysis of educational outcome at age 25, where, as described above, both cohorts are affected by the 1997 sample reduction. The deviation is larger for the analysis of educational outcomes at age 20, when only the 1980s birth cohort suffered attrition through the sample reduction. Overall, however, the trends are quite similar and in line with the conclusions based on the full sample (with higher statistical power).

2 Unweighted Analyses

Though I recommend against inferential analyses of PSID data without the use of survey weights (including for the reasons described in the preceding section), results from unweighted analyses are presented below. Figure S2 compares cohort trends derived from weighted regression models (see also Figure 2 in paper) with those derived from unweighted regression models. Figure S3 compares cohort trends based on quintiles drawn from the weighted distribution (see also Figure 2 in paper) with those based on quintiles drawn from the unweighted distribution. .

Figure S2: Unweighted Regressions

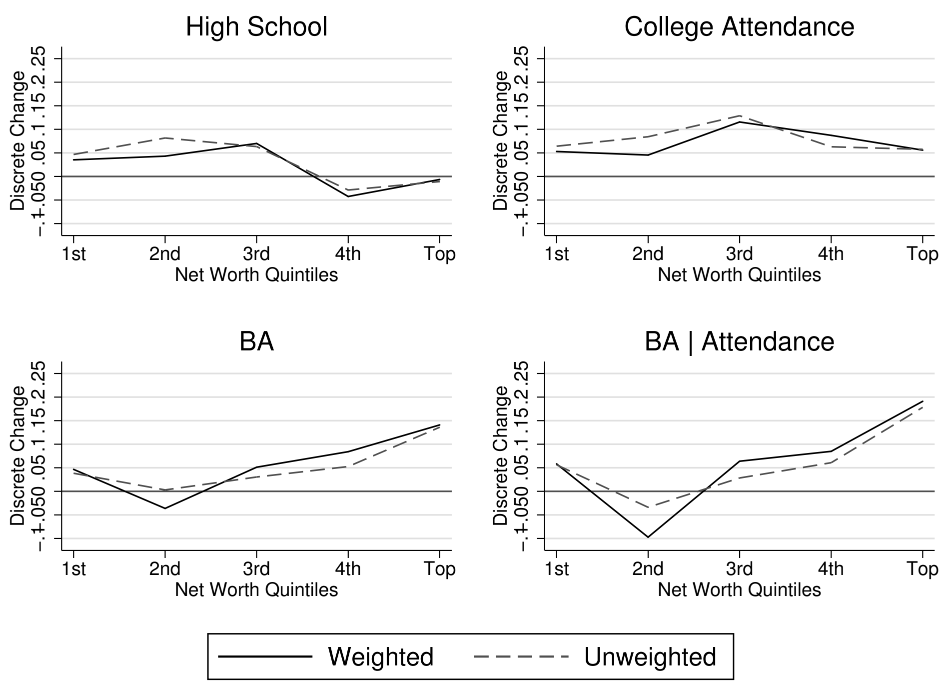
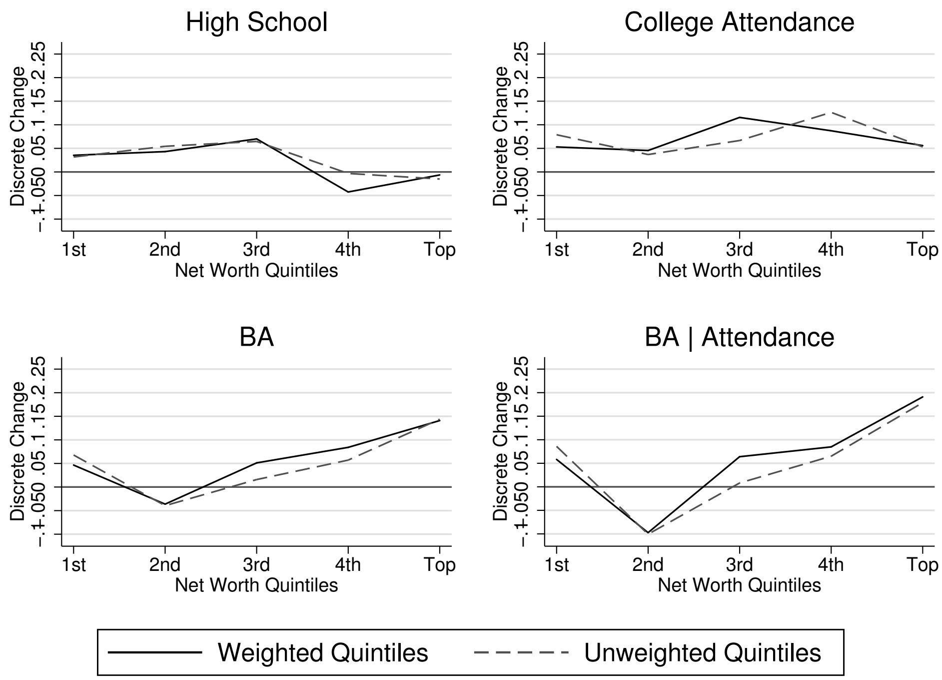


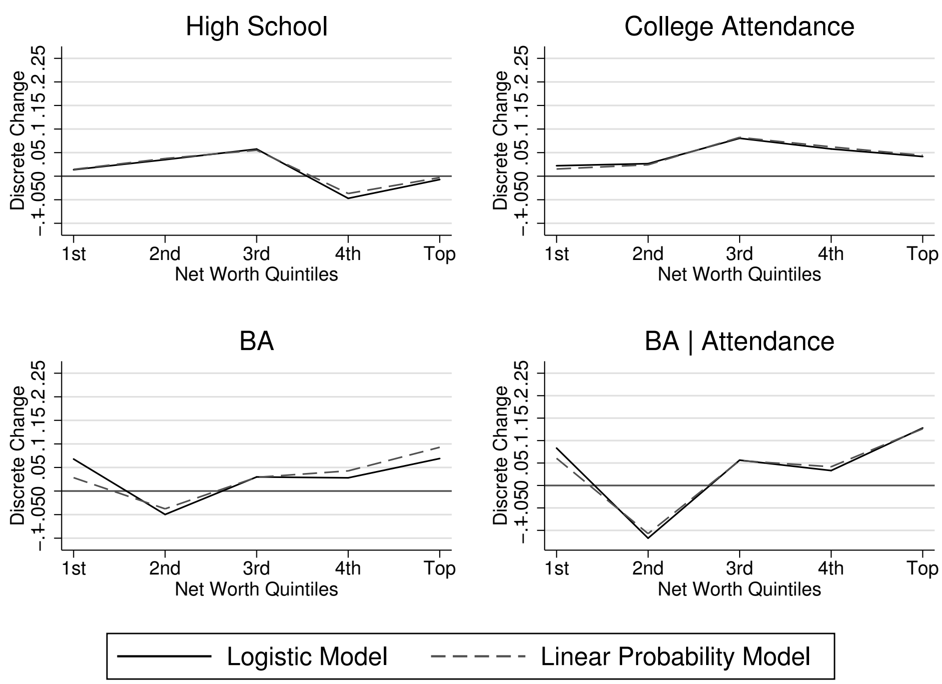
Figure S3: Unweighted Quintiles



3 Linear Probability Models

Unlike odds ratios or logit coefficients, the use of average marginal effect coefficients circumvents many of the methodological challenges entailed in logistic regression models (e.g. Allison 1999; Mood 2009). Still, the relative benefits of Average Marginal Effects (AME) compared to coefficients from linear probability models (LPM) continue to be debated (e.g. Horrace and Oaxaca 2006; Angrist and Pischke 2008). Here, I report results based on LPMs as a stability analysis. Figure S4 compares the AME estimates from the logistic regressions with controls (presented in Figure 3 in the paper) to coefficient estimates for LPM models with controls (the two approaches yield equivalent results in uncontrolled models, such as those presented in Figure 2 in the paper).

Figure S4: Trends in Controlled Wealth Gaps: Logistic Regression vs. LPM



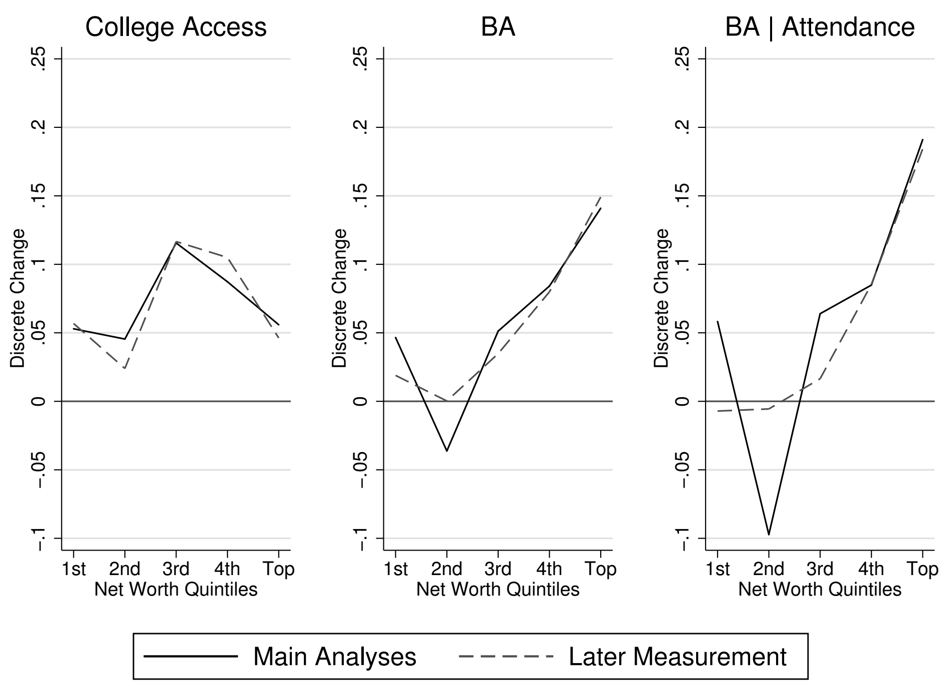
Overall, the differences between the two approaches are negligible. The only deviation – in degree but not structure – relates to the growth of the wealth gap in college graduation: Everything else equal, LPM coefficients indicate less progress between the two cohorts among those in the bottom quintile of the wealth distribution[2] and even more progress among those in the top quintile when compared to the cohort trends based on AMEs. The latter thus appear to be conservative estimates of the growing wealth inequality in college attainment.

[2]However, it should be noted that the predicted probabilities in the LPM are negative for 13 percent of cases.

4 Measuring Family Wealth at Later Ages

The main analyses reported in the paper rely on measures of family wealth taken when children are between 10 and 14 years of age. Stability analyses based on measures taken at later ages, 15 to 18, which may be of particular interest to assess the stability of findings as they relate to wealth fluctuations during the Great Recession (see Appendix B), are reported in Figure S5. Substantively, conclusions about trends in wealth gaps are unaltered. The only notable deviation between trends based on the two different measurement approaches applies to the college persistence rates among children from the bottom three wealth quintiles. While the confidence intervals of the corresponding estimates in the main analyses already covered zero (no change; see Figure 2 in paper), the point estimates based on the later measurement points are also zero, providing an even more uniform picture of lacking progress among the bottom 60% of the wealth distribution in terms of college persistence.

Figure S5: Cohort Trends in Wealth Gaps by Timing of Wealth Measurement



5 Net Worth Excluding Home Equity

Table S3 reports median wealth excluding home equity for three cohorts of children to allow a comparison of wealth inequality trends with and without the inclusion of housing wealth (compare to Table 2 in paper).

Table S3: Trends in Wealth Inequality Among Children - Excluding Home Equity

	Cohort		
	Earlier 10-14 in 80s	Later 10-14 in 90s	Current 10-14 in 2015
Median net worth			
Top 10%	520,118	682,944	840,000
Next 10%	175,729	212,680	244,600
Bottom 80%	13,693	13,912	8,000
Ratios			
Top 10% / next 10%	3.0	3.2	3.4
Top 10% / bottom 80%	38.0	49.1	105.0
Next 10% / bottom 80%	12.8	15.3	30.6
Share with zero/negative net worth	0.139	0.168	0.286
Gini coefficient	0.839	0.878	0.979
Gini coefficient (positive wealth)	0.787	0.816	0.826

6 Decomposition Model

The decomposition analyses in the paper are geared at further elucidating one particular finding, namely the growth of the gap in college attainment between children from the top quintile and children from the bottom four quintiles of the parental wealth distribution. While off-the-shelf decomposition techniques for non-linear models are available (Fairlie 2005; Sinning et al. 2008), this specific explanatory focus profits from a more flexible approach, such as the spline regression model applied here. The choice of the number and location of spline knots is ultimately arbitrary, so the guiding aim is to closely capture the observed wealth gap in college attainment in the baseline cohort (born in 1970s) while retaining parsimony. As shown in the first two columns of Table S4 (see also Table 3 in the paper), the spline regression model (1) with two knots at wealth values representing the 80th and 90th percentile of the unweighted distribution[3] meets that aim very well: deviations between predicted and observed probabilities are negligible or null. Other specifications are less effective in reproducing the observed gaps, including (2) a spline model with just one knot at the 80th percentile, for which the deviation between observed and predicted probabilities are much larger. Predicted probabilities from this model misrepresent the college graduation gap between those from the top quintile of the wealth distribution and everyone else by about five percentage points. A less parsimonious specification using (3) a spline model with knots at the 20th, 40th, 60th, and 80th percentiles does better than that (with deviations between 1 and 2 percentage points) but not as well as the main model (1). The same applies to a specification that (4) separately estimates regressions for the top quintile and the bottom four quintiles.

[3]The values representing the same percentiles of the weighted distribution provided a lower fit.

Table S4: Decomposition: Alternative Predictive Models

	Observed	Predicted			
		(1)	(2)	(3)	(4)
Probability of BA					
(1.1) Lowest four wealth quintiles	18.8%	17.6%	18.6%	17.3%	16.0%
(1.2) Highest wealth quintile	46.0%	45.2%	40.5%	43.2%	44.4%
(1.3) Gap [1.2-1.1]	27.2%	27.5%	21.9%	25.9%	28.4%
Deviation from observed (perc. points)		-0.4	5.2	1.3	-1.2
Probability of BA | Attendance					
(1.1) Lowest four wealth quintiles	34.2%	33.5%	34.7%	33.5%	31.7%
(1.2) Highest wealth quintile	53.7%	53.0%	49.6%	51.5%	53.1%
(1.3) Gap [1.2-1.1]	19.5%	19.5%	14.9%	17.9%	21.4%
Deviation from observed (perc. points)		0.0	4.6	1.6	-1.9

Note: Predictions derived from models specified as:
(1) Splines at P80 and P90 (drawn from full, unweighted sample at age 25) = Main model
(2) Spline at P80 (drawn from full, unweighted sample at age 25)
(3) Splines at quintiles (drawn from full, unweighted sample at age 25)
(4) Separate regressions for highest quintile and bottom four quintiles

7 Decomposition of Growth in Controlled Wealth Gaps

Table S5 reports results for the decomposition of the growing wealth gap in college attainment controlled for other observed characteristics. 35 % of the growth in the controlled wealth gap in college attainment among all and 29 % among those accessing college can be ascribed to growing wealth inequality.

Table S5: Decomposition of Growth in Controlled Wealth Gaps in Education

	Probability of BA		Prob. of BA | Attendance	
	Predicted	Simulated	Predicted	Simulated
(1) Cohort born in 1970s				
(1.1) Lowest four wealth quintiles	16.0%		31.7%	
(1.2) Highest wealth quintile	44.9%		52.9%	
(1.3) Gap [1.2-1.1]	28.9%		21.2%	
(2) Cohort born in 1980s				
(2.1) Lowest four wealth quintiles	18.5%	17.5%	33.4%	32.6%
(2.2) Highest wealth quintile	58.9%	50.4%	71.5%	58.7%
(2.3) Gap [2.2-2.1]	40.4%	32.9%	38.1%	26.1%
(3) Cohort difference in gap [2.3-1.3]	11.5%	4.0%	16.9%	4.9%
(4) Growth in gap accounted for		34.6%		28.7%

References

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods and Research* 28:186–208.

Angrist, Joshua D. and Joern-Steffen Pischke. 2008. *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton: Princeton University Press.

Fairlie, Robert W. 2005. "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models." *Journal of Economic and Social Measurement* 30:305–316.

Horrace, William C. and Ronald L. Oaxaca. 2006. "Results on the bias and inconsistency of ordinary least squares for the linear probability model." *Economics Letters* 90:321–327.

Mood, Carina C. 2009. "Logistic Regression. Why we cannot do what we think we can do, and what we can do about it." *European Sociological Review* 26:67–82.

Sinning, Mathias, Markus Hahn, Thomas K. Bauer, and others. 2008. "The Blinder-Oaxaca decomposition for nonlinear regression models." *The Stata Journal* 8:480–492.