I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11657

Has the College Wage Premium Continued to Rise? Evidence from Multiple U.S. Surveys

Jared Ashworth
Tyler Ransom

JULY 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11657

Has the College Wage Premium Continued to Rise? Evidence from Multiple U.S. Surveys

Jared Ashworth
Pepperdine University

Tyler Ransom
University of Oklahoma and IZA

JULY 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Has the College Wage Premium Continued to Rise? Evidence from Multiple U.S. Surveys*

This paper examines trends in the college wage premium (CWP) by birth cohort across the five major household surveys in the United States: the Census/ACS, CPS, NLSY, PSID, and SIPP. We document a flattening in the CWP for birth cohorts 1978 and onward in each survey and even a decline for birth cohorts 1980–1985 in the NLSY and SIPP. We discuss potential reasons for this finding and show that the empirical discrepancy is not a function of differences in composition across surveys. Our results provide crucial context for the vast economic literatures that use these surveys to measure returns to skill, and intertemporal changes in those returns.

JEL Classification: I26, J30
Keywords: college wage premium, returns to education

Corresponding author:
Jared Ashworth
Graziadio School of Business and Management
Pepperdine University
Malibu, CA 90263
USA
E-mail: jared.ashworth@pepperdine.edu

* We are grateful to Michael Böhm, Patrick Coate, and Jamin Speer for sharing helpful code with us. We also thank Dan Black, Timothy Bond, V. Joseph Hotz, Nick Huntington-Klein, and Arnaud Maurel for helpful comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: none. All errors are our own.

1 Introduction

The college wage premium (CWP) measures the wage differential between college graduates and high school graduates and is commonly understood to measure an economy's demand for skilled labor. A well-documented and seminal point in the economic history of the United States is when the CWP suddenly rose in the 1980s and continued to rise throughout the 1990s and into the early 2000s. We ask whether this trend has continued to hold more recently and how consistently the trend holds across commonly used surveys.

Using the five major U.S. household surveys, we document a substantial rise in the CWP in each of the surveys for birth cohorts 1950–1970. However, this was followed by a flattening thereafter. This finding corroborates recent studies that have documented declining employment and income prospects and declining returns to skill among recent birth cohorts (see Beaudry, Green, and Sand (2014), Guvenen et al. (2017), Valletta (Forthcoming), and Gallipoli and Makridis (2018)). Surprisingly, we document a *decline* in the CWP in the SIPP and NLSY for birth cohorts 1978–1984. The decline is more pronounced among men than women.

The five major household surveys are the Decennial Census 5% Public Use Micro Sample (hereafter Census) and the American Community Survey (ACS); the Current Population Survey Outgoing Rotation Groups (CPS); the 1979 and 1997 National Longitudinal Surveys of Youth (NLSY79 and NLSY97); the Panel Study of Income Dynamics (PSID); and the Survey of Income and Program Participation (SIPP). In each survey and for each birth cohort, we estimate unconditional and Mincer (1974)-style log wage regressions to calculate the CWP for full-time/full-year workers aged 25–34.[1] Ours is the first study to compare trends in the CWP across these five commonly-used household surveys.

We investigate whether our findings can be explained by differences across surveys in the levels of observed characteristics such as demographic, education, or employment variables. We find no major discrepancies. We conclude that the differences are likely due to differences in survey architecture (i.e. sample size and collection methods, or whether the survey is repeated cross-section versus longitudinal).

Our results have implications for the long and growing list of studies that examine cross-cohort changes in the returns to skill. Many studies use the CPS or decennial Censuses for this type of research (see Goldin and Katz, 2007, and many others), but there are a growing number of studies using the NLSY (see, e.g. Altonji, Bharadwaj, and Lange, 2012; Ashworth et al., 2017; Bacolod and Hotz, 2006; Böhm, 2017; Castex and Dechter, 2014; Lee, Shin, and Lee, 2015; Deming, 2017), as well as the PSID (see Cortes, 2016; Yamaguchi, 2018, and

[1] Analysis for other age groups is available in the online appendix.

others). Our findings suggest that researchers should not necessarily expect to see the same trends in each major survey. Furthermore, to the extent that the CWP does in fact measure demand for skill, we document that this demand is leveling off and may even be declining. This leveling off is correlated with a stark decline in the labor force participation rate of men in birth cohorts 1978–1984.

The remainder of the paper is organized as follows: the next section briefly discusses in more detail the data sets and key variables we use; Section 3 discusses our key results; and Section 4 offers discussions and conclusions.

2 Data

In this section we briefly describe the data sets used in our analysis. As mentioned previously, we use the five major US household surveys spanning birth cohorts 1950–1985: the 1980, 1990, and 2000 Census 5% Public Use Micro Samples and the 2001-2016 ACS (Ruggles et al., 2017); the CPS-ORG; the NLSY79 and NLSY97; the PSID; and the SIPP. In the interest of brevity and due to the well-known nature of each of these surveys, we refer the reader to the online appendix for additional details regarding the structure and mechanics of each survey.

2.1 Key variables

Here we briefly discuss our construction of the three main variables that enter our analysis: wages, educational attainment, and employment status. We restrict our attention to full-time, full-year workers in each of our analyses that follow.

We define wages as hourly earnings, which are constructed in various ways depending on the survey. In the NLSY, workers report hourly earnings even if they work at a salaried job. In the CPS and SIPP, workers who are paid by the hour report hourly earnings. For the Census/ACS and the PSID, and for salaried workers in the CPS and SIPP, we compute hourly earnings as the annual, monthly, or weekly wage income divided by the hours worked in the corresponding year, month, or week. We express all wage or income variables in \$1982-84 using the CPI-U.

Educational attainment is taken from respondent reports in each survey. We define high school graduates as those who completed at least 12 years of schooling, who hold at least a high school diploma, or who hold a GED. We define college graduates as those who completed at least 16 years of schooling or who hold at least a bachelor's degree.

Employment status is defined as full-time, part-time, or not employed. To the extent possible, we attempt to focus on full-time full-year workers. This classification slightly

differs by dataset. In the CPS, workers report working full-time but not full-year because they are surveyed about only a recent workweek. In the PSID, full-time workers work more than 1500 hours during the year. In the Census/ACS and NLSY, full-time workers work at least 35 hours per week and at least 40 weeks in the past year. In the SIPP they work at least 30 hours per week in at least 90% of the observed non-school months.

Additional details on each of our three main variables are available in the appendix.

3 Methodology & Results

This section briefly introduces our methodology and reports and discusses our main findings.

3.1 Methodology

To estimate the unconditional CWP, we estimate weighted regression models of the following form for individuals aged 25–34, separately for each birth cohort c and for each survey s:[2]

$$\ln w_{isc} = \alpha_{0sc} + \alpha_{1sc} gradHS_{isc} + \alpha_{2sc} grad4yr_{isc} + \varepsilon_{isc} \tag{3.1}$$

where w_{isc} is the log hourly wage, $gradHS_{isc}$ is an indicator for if individual i in birth cohort c in survey s holds at least a high school diploma (or GED) and where $grad4yr_{isc}$ is an indicator for if the individual has completed at least a bachelor's degree. Thus, α_{0sc} measures the average log wage of high school dropouts, α_{1sc} the wage premium for holding a high school diploma (relative to not completing high school), and α_{2sc} measures the wage premium for holding a bachelor's degree (relative to completing high school), i.e. the CWP.[3]

We also estimate the CWP corrected for observable differences across individuals. Our main specification is a variant of the Mincer (1974) model:

$$\ln w_{isc} = \beta_{0sc} + \beta_{1sc} gradHS_{isc} + \beta_{2sc} grad4yr_{isc} + \beta_{3sc} S_{isc} + \beta_{4sc} X_{isc} + \beta_{5sc} X_{isc}^2 + \eta_{isc} \tag{3.2}$$

where S_{isc} measures the individual's years of completed schooling, and X_{isc} the individual's years of potential work experience, measured as age (in years) minus years of completed schooling minus six.

[2] We also explore other age ranges (reported in the online appendix). The trends are similar, although as we consider higher age ranges, we lose the ability to measure wages for later birth cohorts.

[3] In results not reported, but available upon request, we repeat this analysis for those with *exactly* a high school diploma and *exactly* a bachelor's degree. We find similar trends in the CWP, although the magnitudes are different.

We present and discuss estimates of (3.1) and (3.2) in the following subsection.

3.2 Results

Our main findings are graphically reported in Figure 1. This figure plots a smoothed version of the α_2 vector in (3.1) across birth cohorts (on the x-axis) and surveys (separate lines).[4] Smoothing is done using local linear regression (LOWESS).[5] The main finding is that, while all five surveys show a steep increase in the CWP for birth cohorts 1950 through about 1965, there is a distinct flattening beginning around birth cohort 1970. We even observe a decline in the CWP in the NLSY and SIPP for those born after 1977. This decline is more pronounced among men than women. To visualize the amount of uncertainty in our estimates, we include a 95% confidence band around the NLSY estimates. These do not intersect with the ACS or CPS lines for the later birth cohorts in question.

We further explore trends in the CWP by considering a measure of the wage premium that is purged of some forms of selection. In Figure 2 we present smoothed estimates of the β_2 vector in (3.2).[6] This graph again shows a flattening for birth cohorts after the early 1970s, with the NLSY and SIPP each having a lower measured CWP for the youngest cohorts. For each survey, the CWP from the Mincer model is roughly half the amount of the raw CWP. Examining the 95% confidence bands shows that the NLSY and SIPP are not significantly different from the ACS and CPS for the youngest cohorts, with the exception of the 1982 and 1983 cohorts for men and the 1980 and 1981 cohorts for women.

Overall, our results of a flattening CWP are consistent with the findings of Beaudry, Green, and Sand (2014), Guvenen et al. (2017), and Valletta (Forthcoming) who respectively document declining probability of obtaining cognitive jobs early in their careers for college graduates in more recent birth cohorts, declining lifetime income for more recent birth cohorts, and a recent flattening of the CWP. Our study is the first to document the apparent decline in the CWP for recent cohorts in both the NLSY and SIPP, though (Ashworth et al., 2017) indirectly document the decline in the NLSY.

One remaining question is whether these surveys consistently measure education, wages, employment, and demographics. We present graphical evidence that they do, in fact, consistently measure these outcomes among the population of full-time, full-year workers. Figures 3, 4, and 5 respectively show cohort-specific averages of college graduates, high school graduates, and full-time workers. Similar figures for demographics can be found in the online

[4]The PSID line disappears after the 1960s due to sample sizes by birth cohort that become unreliably small.

[5]The unsmoothed version of Figure 1 is reported in the online appendix.

[6]The unsmoothed version of Figure 2 is reported in the online appendix.

appendix.

We assess the robustness of our findings by examining alternate age ranges, dropping imputed earnings in the CPS (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006, 2013), and using log earnings instead of log wages for the ACS (Baum-Snow and Neal, 2009). These results are reported in the online appendix or available from the authors upon request. None of our findings is meaningfully affected.

4 Discussion & Conclusions

The most plausible explanation for our finding that the NLSY and SIPP differ from the CPS and ACS with respect to measuring the CWP has to do with survey architecture. The NLSY and SIPP are longitudinal studies, whereas the ACS and CPS are repeated cross sections. The goals of each survey are sufficiently different that the surveys might end up with different measures of wages and hence different measures of the CWP. Furthermore, longitudinal surveys are subject to non-random attrition.[7] This could explain some of the discrepancies, although we argue that if non-random attrition were problematic, it would show up in significant differences of key observable variables. Furthermore, attrition tends to be negatively selected, which would imply—if anything—an *upward* bias in the CWP.

Another potential, though less plausible, explanation is the Great Recession. This recession impacted post-1977 birth cohorts most strongly, which can be seen in Figure 5 as a steep decline in male full-time employment rates for those cohorts. What is puzzling, and what makes this explanation less plausible, is that there does not seem to be any explanation for why the Great Recession would affect the NLSY or SIPP any differently than the ACS or CPS.

A primary implication of our findings is that the demand for skill is flattening and may even be falling, to the extent that the CWP actually measures skill demand. This interpretation is consistent with recent literature cited above that has documented declining income and employment prospects for younger birth cohorts. A secondary implication is that researchers should not necessarily expect the NLSY and SIPP to look the same as the CPS in terms of CWP dynamics. Thus, whether the "correct" CWP is the one measured by the ACS, the CPS, or some other survey, is an open question. It behooves researchers to take note of the differences across surveys.

[7] See the online appendix for a comparison of attrition rates in the NLSY79 and NLSY97.

References

Altonji, Joseph G., Prashant Bharadwaj, and Fabian Lange. 2012. "Changes in the Characteristics of American Youth: Implications for Adult Outcomes." *Journal of Labor Economics* 30 (4):783–828.

Ashworth, Jared, V. Joseph Hotz, Arnaud Maurel, and Tyler Ransom. 2017. "Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences." Working Paper 24160, National Bureau of Economic Research.

Bacolod, Marigee and V. Joseph Hotz. 2006. "Cohort Changes in the Transition from School to Work: Evidence from three NLS surveys." *Economics of Education Review* 25 (4):351–373.

Baum-Snow, Nathaniel and Derek Neal. 2009. "Mismeasurement of Usual Hours Worked in the Census and ACS." *Economics Letters* 102 (1):39–41.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2014. "The Declining Fortunes of the Young Since 2000." *American Economic Review* 104 (5):381–386.

Böhm, Michael Johannes. 2017. "The Price of Polarization: Estimating Task Prices under Routine-Biased Technical Change." Discussion Paper 11220, IZA.

Bollinger, Christopher R. and Barry T. Hirsch. 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." *Journal of Labor Economics* 24 (3):483–519.

———. 2013. "Is Earnings Nonresponse Ignorable?" *Review of Economics and Statistics* 95 (2):407–416.

Castex, Gonzalo and Evgenia Kogan Dechter. 2014. "The Changing Roles of Education and Ability in Wage Determination." *Journal of Labor Economics* 32 (4):685–710.

Cortes, Guido Matias. 2016. "Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data." *Journal of Labor Economics* 34 (1):63–105.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* 132 (4):1593–1640.

Gallipoli, Giovanni and Christos A. Makridis. 2018. "Structural Transformation and the Rise of Information Technology." *Journal of Monetary Economics* .

Goldin, Claudia and Lawrence F. Katz. 2007. "Long-Run Changes in the US Wage Structure: Narrowing, Widening, Polarizing." *Brookings Papers on Economic Activity* 2007 (2):135–165.

Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner. 2017. "Lifetime Incomes in the United States over Six Decades." Working Paper 23371, National Bureau of Economic Research.

Hirsch, Barry T. and Edward J. Schumacher. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics* 22 (3):689–722.

Lee, Sang Yoon, Yongseok Shin, and Donghoon Lee. 2015. "The Option Value of Human Capital: Higher Education and Wage Inequality." Working Paper 21724, National Bureau of Economic Research.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: Columbia University Press for National Bureau of Economic Research.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. *Integrated Public Use Microdata Series: Version 7.0 [Machine-readable database]*. Minneapolis: University of Minnesota.

Valletta, Robert G. Forthcoming. "Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both?" In *Education, Skills, and Technical Change: Implications for Future U.S. GDP Growth*, edited by Charles Hulten and Valerie Ramey. University of Chicago Press. URL http://www.nber.org/chapters/c13705.

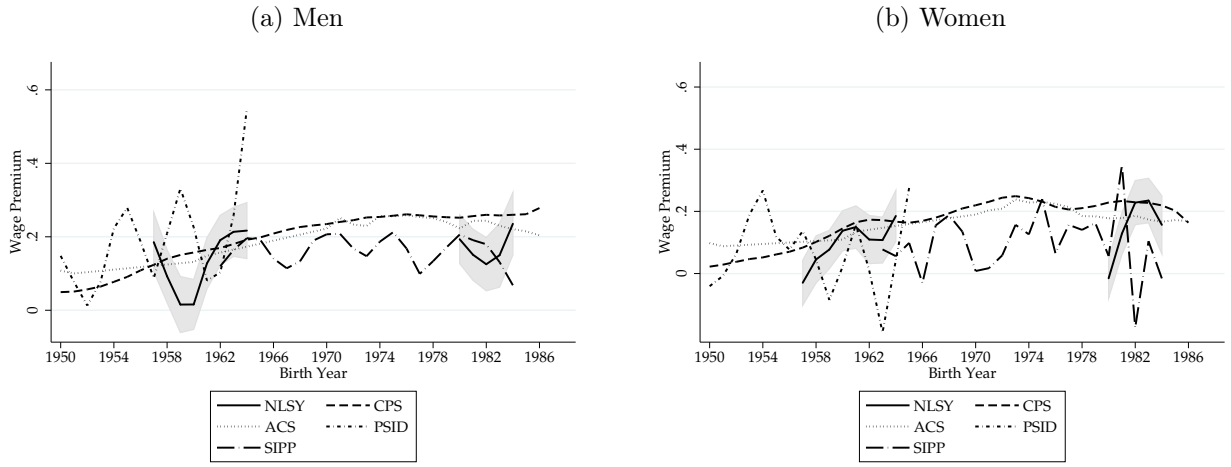
Yamaguchi, Shintaro. 2018. "Changes in Returns to Task-Specific Skills and Gender Wage Gap." *Journal of Human Resources* 53 (1):32–70.

Figures and Tables

Figure 1: Raw (smoothed) college wage premium (25–34 year olds) by birth cohort across five U.S. surveys

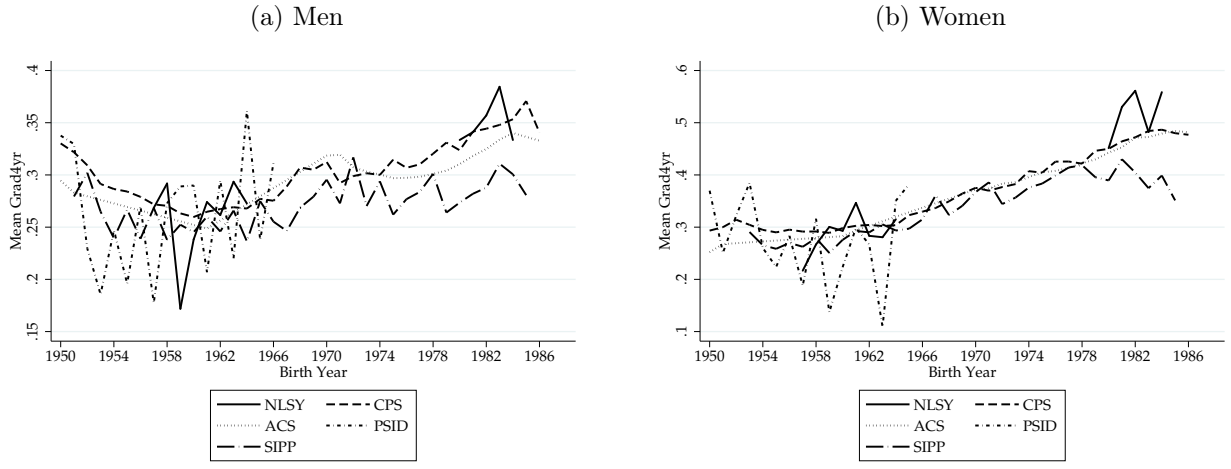
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort, smoothed using local linear regression (LOWESS). Sample includes only those who are working full-time, full-year and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census. For additional details regarding construction of the data, see the online appendix.

Figure 2: Mincer college wage premium (25–34 year olds) by birth cohort across five U.S. surveys



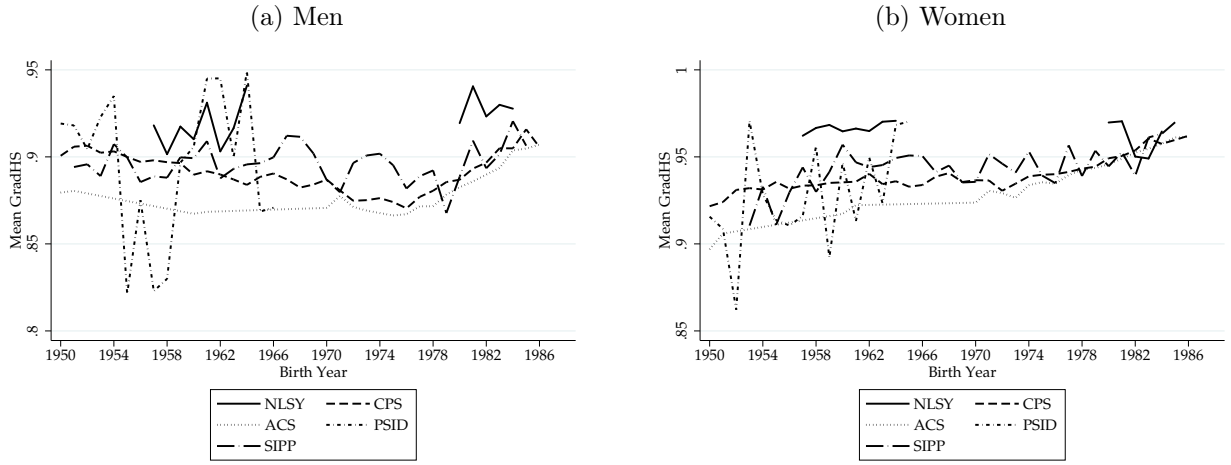
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort, adjusted for observable skills using the classic Mincer (1974) model. See note to Figure 1.

Figure 3: Raw college graduation rates (25–34 year olds) by birth cohort across five U.S. surveys



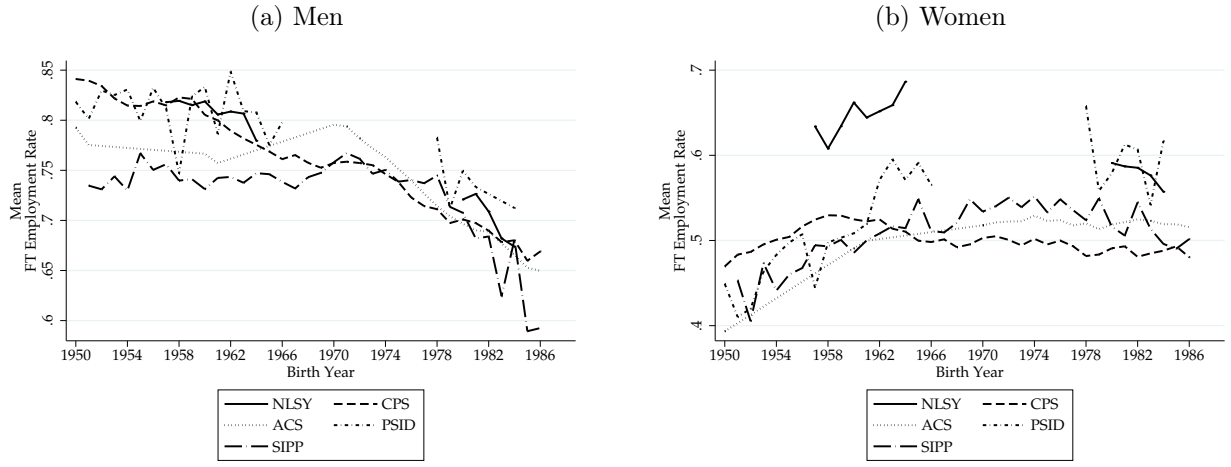
Notes: The above figures plot the proportion of the population that are college graduates by birth cohort. See note to Figure 1.

Figure 4: Raw high school graduation rates (25–34 year olds) by birth cohort across five U.S. surveys



Notes: The above figures plot the proportion of the population that are high school graduates by birth cohort. See note to Figure 1.

Figure 5: Raw full-time, full-year employment rates (25–34 year olds) by birth cohort across five U.S. surveys



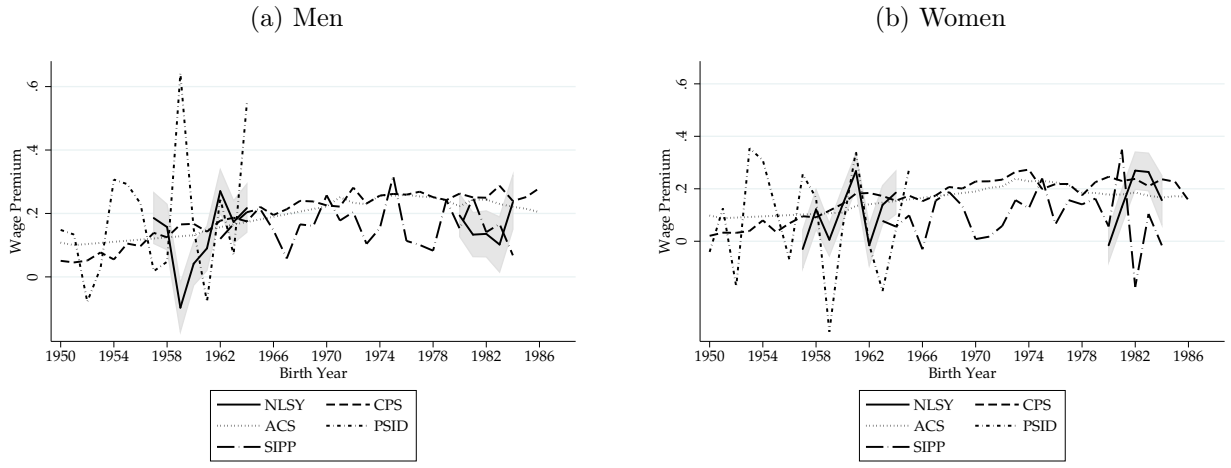
Notes: The above figures plot the proportion of the population that are employed full-time, full-year by birth cohort. See note to Figure 1.

A Online Appendix – Supplementary figures and tables

Figure A1: Raw (unsmoothed) college wage premium (25–34 year olds) by birth cohort across five U.S. surveys

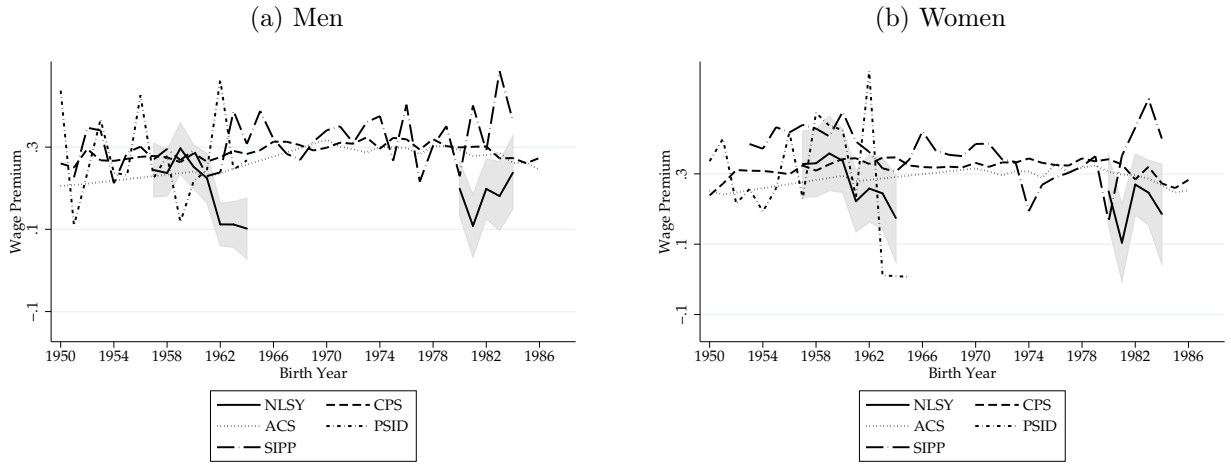
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census.

Figure A2: Mincer (unsmoothed) college wage premium (25–34 year olds) by birth cohort across five U.S. surveys



Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census.

Figure A3: Raw high school wage premium (25–34 year olds) by birth cohort across five U.S. surveys



Notes: The above figures plot the difference in log wages between high school graduates and high school dropouts by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census.

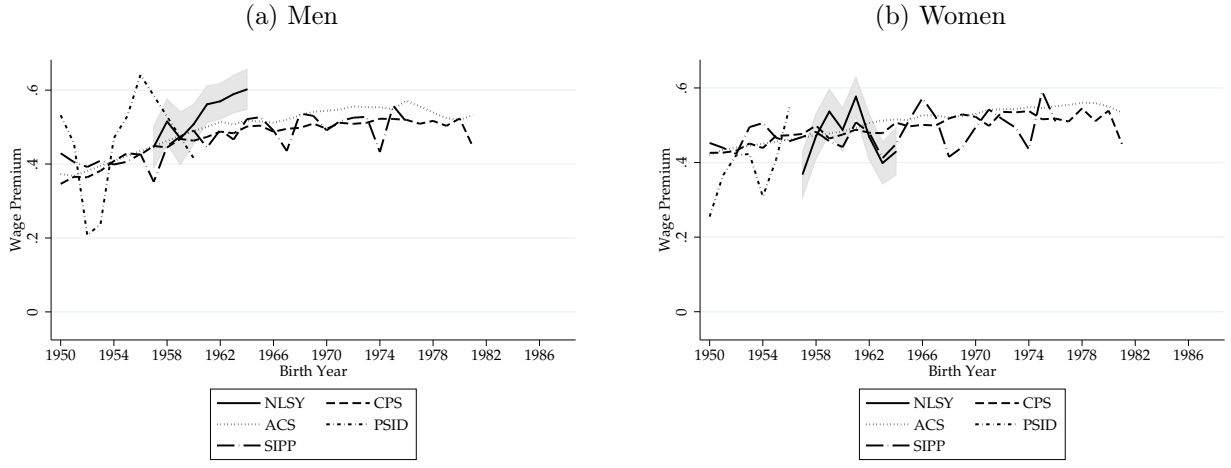
Figure A4: Raw college wage premium (26–30 year olds) by birth cohort across five U.S. surveys

Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 26–30. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1952 for the 1980 Census, 1962 for the 1990 Census, and 1972 for the 2000 Census.

Figure A5: Raw college wage premium (30–39 year olds) by birth cohort across five U.S. surveys

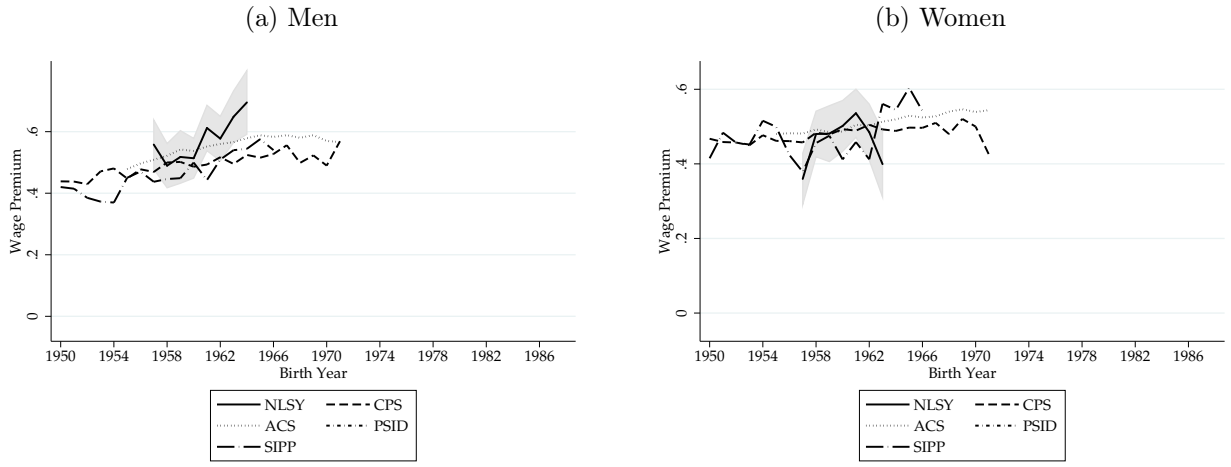
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 30–39. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1945 and 1946 for the 1980 Census, 1955 and 1956 for the 1990 Census, and 1965 and 1966 for the 2000 Census.

Figure A6: Raw college wage premium (35–44 year olds) by birth cohort across five U.S. surveys



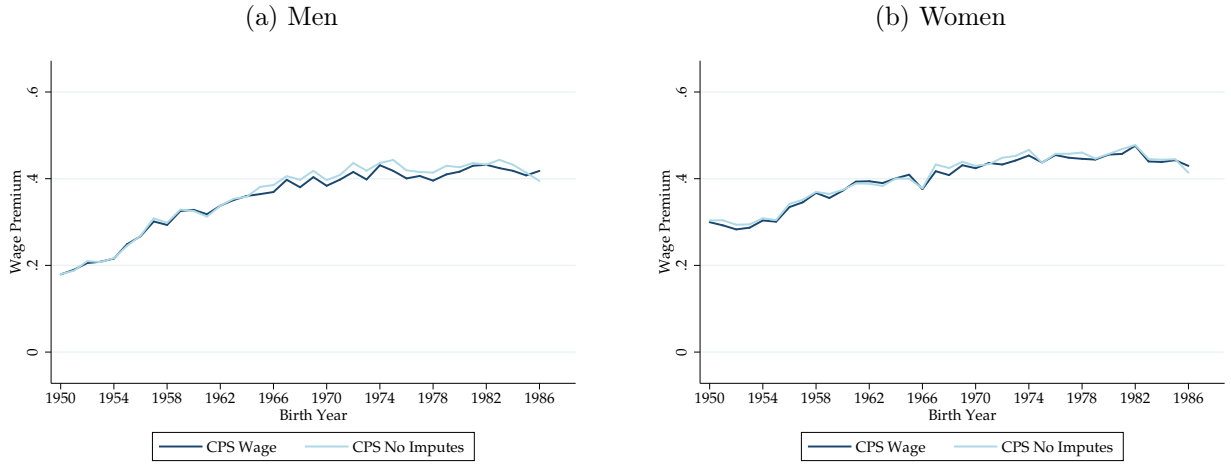
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 35–44. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1990 Census and 1960 and 1961 for the 2000 Census.

Figure A7: Raw college wage premium (45–54 year olds) by birth cohort across five U.S. surveys



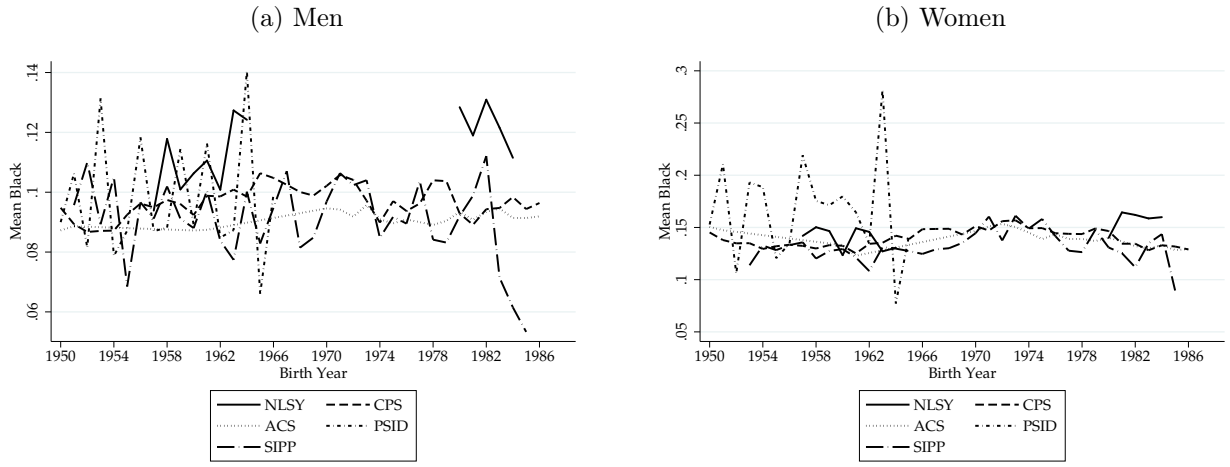
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 45–54. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 2000 Census.

Figure A8: Raw college wage premium in CPS, with and without imputed earnings



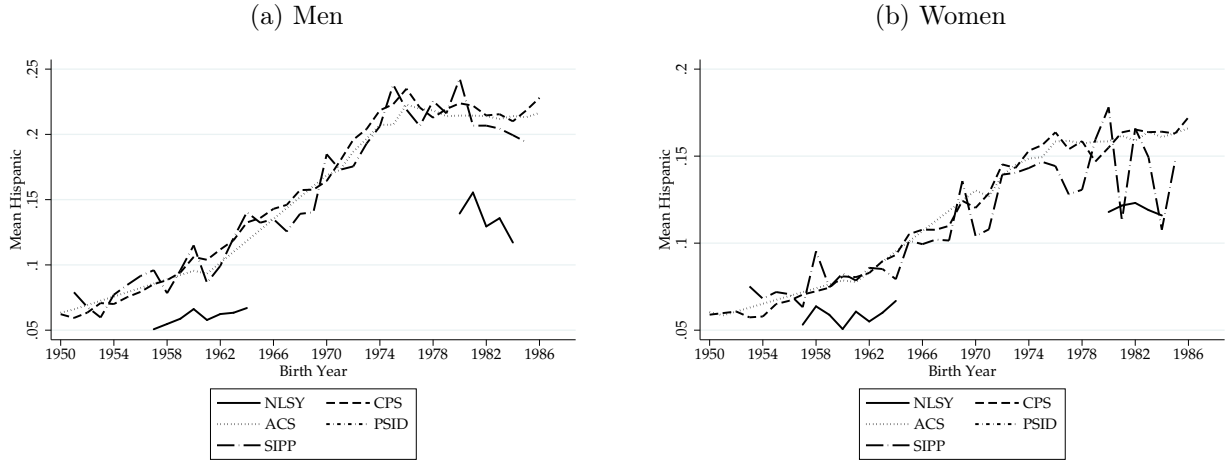
Notes: The above figures plot the difference in log wages between college graduates and high school graduates by birth cohort. Sample includes only those in the CPS who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by the CPS.

Figure A9: Composition: Percent black by birth cohort across five U.S. surveys



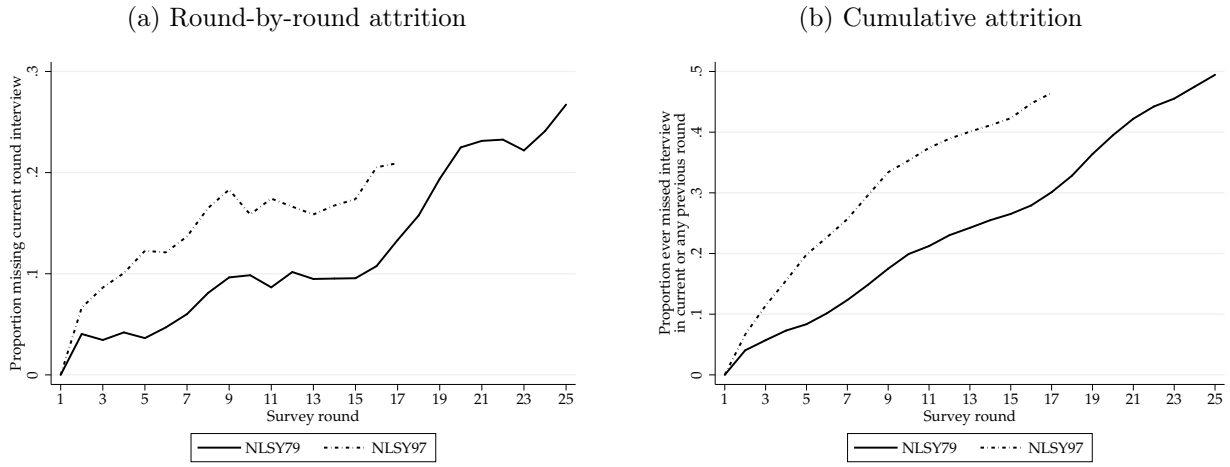
Notes: The above figures plot the proportion of the population that are black by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census.

Figure A10: Composition: Percent Hispanic by birth cohort across five U.S. surveys



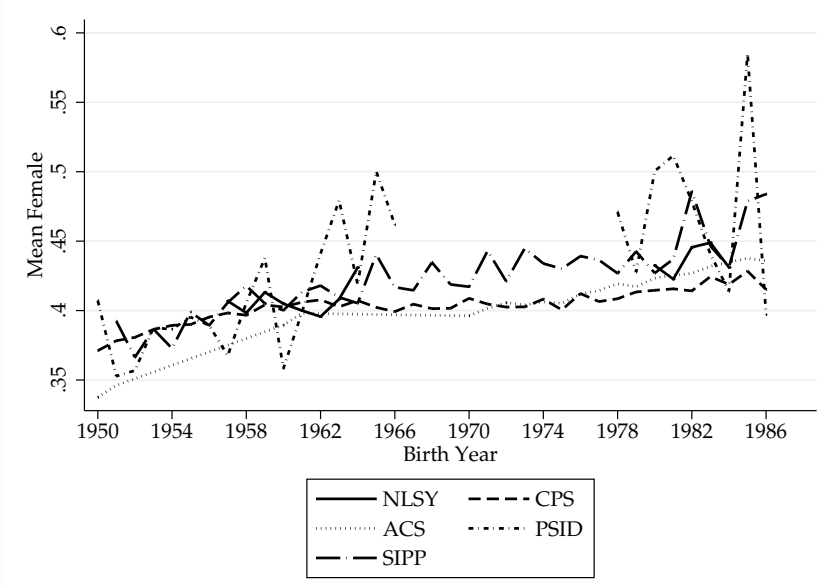
Notes: The above figures plot the proportion of the population that are Hispanic by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census.

Figure A11: Attrition in the NLSY surveys



Notes: The above figures respectively plot the round-by-round and cumulative attrition rates in each of the NLSY surveys.

Figure A12: Composition: Percent female by birth cohort across five U.S. surveys



Notes: The above figures plot the proportion of the population that are female by birth cohort. Sample includes only those who are working full-time, full-year, and who are between the ages of 25–34. Each point on each line requires an underlying sample of $N \geq 400$. All statistics are computed using the sampling weights provided by each survey. The ACS series is restricted to birth cohorts 1950 and 1951 for the 1980 Census, 1960 and 1961 for the 1990 Census, and 1970 and 1971 for the 2000 Census.

Table A1: Number of observations (25–34 year old men) by birth cohort and survey

Birth year	ACS	CPS	NLSY	PSID	SIPP	Total
1950	91,646	23,647		837	304	29,108
1951	92,773	27,524		864	564	30,431
1952		31,252		1,100	742	11,031
1953		35,219		1,024	1,251	12,498
1954		37,812		1,190	1,640	13,547
1955		37,449		1,162	1,556	13,389
1956		36,947		1,197	2,099	13,414
1957		37,270	4,054	965	2,693	11,246
1958		36,138	4,484	1,175	3,272	11,267
1959		35,174	4,609	1,092	4,094	11,242
1960	110,358	35,620	5,404	1,198	4,461	31,408
1961	105,433	33,976	5,370	1,162	3,881	29,964
1962		32,462	5,162	1,009	4,285	10,730
1963		31,421	4,917	885	4,649	10,468
1964		30,091	3,650	839	4,810	9,848
1965		27,968		692	4,728	11,129
1966		26,684		645	4,055	10,461
1967		25,197		603	3,946	9,915
1968		25,427		509	3,836	9,924
1969		26,045		401	3,804	10,083
1970	131,305	26,465		464	3,839	40,518
1971	147,774	25,514		335	4,243	44,466
1972	156,042	23,948		389	4,126	46,126
1973	165,239	22,669		315	3,594	47,954
1974	174,451	23,317		413	3,509	50,422
1975	194,287	23,828		393	3,700	55,552
1976	122,506	23,776		402	3,852	37,634
1977	134,475	24,128		442	3,915	40,740
1978	146,456	24,079		488	3,727	43,688
1979	158,486	24,844		568	3,685	46,896
1980	175,461	25,549	5,246	568	3,153	41,995
1981	179,464	25,190	5,241	518	2,577	42,598
1982	180,059	23,695	4,501	486	2,108	42,170
1983	162,565	20,543	3,746	457	2,011	37,864
1984	142,781	18,403	2,933	556	1,436	33,222
1985	127,479	15,902		366	961	36,177

Table A2: Number of observations (25–34 year old women) by birth cohort and survey

Birth year	ACS	CPS	NLSY	PSID	SIPP	Total
1950	94,059	25,695		1,035	310	30,275
1951	94,502	29,976		1,107	614	31,550
1952		34,479		1,052	807	12,113
1953		38,432		1,142	1,339	13,638
1954		41,005		1,400	1,787	14,731
1955		40,726		1,272	1,845	14,614
1956		40,706		1,397	2,375	14,826
1957		40,726	5,069	1,131	3,034	12,490
1958		39,407	5,076	1,245	3,820	12,387
1959		39,757	5,438	1,217	4,520	12,733
1960	113,339	39,314	5,379	1,192	4,995	32,844
1961	108,970	37,751	5,687	1,185	4,594	31,637
1962		35,943	5,138	1,292	4,912	11,821
1963		34,306	4,787	1,083	5,036	11,303
1964		33,125	3,379	968	5,167	10,660
1965		30,864		878	5,565	12,436
1966		28,619		759	4,719	11,366
1967		27,768		505	4,398	10,890
1968		27,715		516	4,529	10,920
1969		27,956		455	3,942	10,784
1970	132,612	29,219		500	4,055	41,596
1971	153,882	27,593		482	5,066	46,756
1972	161,721	25,819		389	4,701	48,158
1973	172,583	24,833		383	4,314	50,528
1974	178,930	25,207		376	4,085	52,150
1975	197,513	25,681		527	4,221	56,986
1976	128,542	26,058		475	4,456	39,883
1977	140,810	26,019		501	4,576	42,976
1978	153,031	26,113		571	4,438	46,038
1979	163,748	26,646		552	4,086	48,758
1980	180,352	27,427	5,430	717	3,613	43,508
1981	181,209	26,768	5,243	616	3,077	43,383
1982	181,922	25,805	4,557	570	2,699	43,111
1983	163,803	22,709	3,598	509	2,097	38,543
1984	143,293	19,783	2,859	502	1,575	33,602
1985	126,687	17,009		437	1,099	36,308

Table A3: Number of full-time, full-year, wage observations (25–34 year old men) by birth cohort and survey

Birth year	ACS	CPS	NLSY	PSID	SIPP	Total
1950	66,625	17,214		561	203	21,151
1951	66,253	20,110		613	401	21,844
1952		22,662		777	538	7,992
1953		25,383		754	924	9,020
1954		27,086		814	1,187	9,696
1955		27,001		792	1,178	9,657
1956		26,794		800	1,547	9,714
1957		26,988	2,528	623	2,002	8,035
1958		26,488	2,893	718	2,374	8,118
1959		25,712	2,837	645	2,999	8,048
1960	79,975	25,733	3,406	708	3,245	22,613
1961	75,920	24,242	3,354	615	2,852	21,397
1962		23,105	3,178	617	3,125	7,506
1963		22,160	3,062	492	3,376	7,272
1964		20,866	2,188	478	3,577	6,777
1965		19,189		348	3,523	7,687
1966		18,187		348	3,010	7,182
1967		17,120		343	2,897	6,787
1968		17,250		285	2,869	6,801
1969		17,375		268	2,840	6,828
1970	96,385	17,836		298	2,838	29,339
1971	108,752	17,297		221	3,177	32,362
1972	112,965	16,259		264	3,065	33,138
1973	117,819	15,279		223	2,652	33,993
1974	122,604	15,573		246	2,581	35,251
1975	134,186	15,717		268	2,646	38,204
1976	86,364	15,423		266	2,767	26,205
1977	92,406	15,584		316	2,766	27,768
1978	98,789	15,547		307	2,650	29,323
1979	105,338	15,773		340	2,512	30,991
1980	114,732	16,467	3,196	353	2,155	27,381
1981	116,030	16,078	3,090	320	1,691	27,442
1982	116,156	14,960	2,596	281	1,399	27,078
1983	102,912	12,870	2,103	258	1,236	23,876
1984	88,932	11,590	1,618	287	912	20,668
1985	77,750	9,862		182	555	22,087

Table A4: Number of full-time, full-year, wage observations (25–34 year old women) by birth cohort and survey

Birth year	ACS	CPS	NLSY	PSID	SIPP	Total
1950	35,051	11,182		445	127	11,701
1951	36,125	13,383		401	265	12,544
1952		15,548		406	316	5,423
1953		17,731		512	614	6,286
1954		19,092		635	759	6,829
1955		19,128		573	823	6,841
1956		19,599		622	1,072	7,098
1957		19,913	2,050	441	1,466	5,968
1958		19,366	2,027	548	1,850	5,948
1959		19,675	2,241	508	2,209	6,158
1960	52,050	19,440	2,245	474	2,330	15,308
1961	51,151	18,496	2,564	488	2,254	14,991
1962		17,720	2,191	525	2,433	5,717
1963		16,639	2,048	440	2,518	5,411
1964		15,879	1,591	388	2,610	5,117
1965		14,470		424	2,988	5,961
1966		13,285		300	2,329	5,305
1967		13,015		226	2,175	5,139
1968		12,619		227	2,246	5,031
1969		12,853		219	2,090	5,054
1970	66,768	13,484		263	2,105	20,655
1971	77,912	12,905		258	2,648	23,431
1972	81,530	12,068		196	2,498	24,073
1973	86,323	11,431		203	2,229	25,046
1974	90,034	11,848		207	2,143	26,058
1975	98,920	11,897		315	2,134	28,316
1976	64,832	12,249		232	2,342	19,914
1977	70,358	12,028		285	2,291	21,240
1978	76,809	11,919		317	2,214	22,815
1979	81,788	12,229		267	2,123	24,102
1980	90,123	12,642	2,596	374	1,790	21,505
1981	91,134	12,503	2,508	341	1,528	21,603
1982	92,248	11,847	2,175	307	1,379	21,591
1983	82,657	10,526	1,671	249	1,007	19,222
1984	71,581	9,126	1,297	275	760	16,608
1985	63,269	8,027		240	515	18,013

B Online Appendix – Data details

In this appendix, we introduce and detail our construction of each of the main data sets used in the analysis. We compare the coverage of each, as well as explaining how wages, education, and employment are measured in each.

B.1 Overview of the data sets

We use the following five data sets in our analysis:

1. Decennial Census (1980, 1990, 2000) and American Community Survey (ACS; 2001–2016)
2. Current Population Survey (CPS; 1979–2016)
3. National Longitudinal Survey of Youth (NLSY), panels 1979 and 1997
4. Panel Study of Income Dynamics (PSID; 1968–2015)
5. Survey of Income and Program Participation (SIPP; panels 1984–2008)

Below we present further detail regarding the nature of each of these five commonly used data sets. Each of the data sets is a household survey that collects a common set of information of interest to researchers. An overview of these data sets is listed in Table B1.

B.1.1 Census/ACS

We make use of 5% population samples from the Decennial Censuses of 1980, 1990, and 2000, which are collected by the US Census Bureau. The ACS—also collected by the Census Bureau—is an annual 1% sample of the US population. Each survey contains information on all members of the sampled household.

B.1.2 CPS

The CPS is a repeated cross-sectional survey conducted by the Bureau of Labor Statistics (BLS) that samples between 50,000 and 60,000 housing units each month. As with the Census and ACS, information is collected on all members of the household which reside in the sampled dwelling. We make use of the Merged Outgoing Rotation Groups (ORG), which are households in their 4th or 8th month of participation in the CPS.

B.1.3 NLSY

The NLSY panels are longitudinal surveys collected by the BLS that follow specific cohorts of youth from adolescence throughout adulthood. The NLSY79 follows individuals in birth cohorts 1957–1964, while the NLSY97 follows youth born in years 1980–1984. The cross-sectional sample size of the panel is about 13,000 for the NLSY79 and about 9,000 for the NLSY97. Surveys were conducted annually for each panel for approximately the first 14 rounds of data collection and then biennially thereafter.

B.1.4 PSID

The PSID is a longitudinal survey collected by the University of Michigan that sampled individuals in 1968, and then followed them and each of their descendants for an extended period of time. The initial sample consisted of about 5,000 families (18,000 individuals). Surveys were conducted annually from 1968–1997 and biennially thereafter. The PSID is the longest running longitudinal household survey in the world.

B.1.5 SIPP

The SIPP is a longitudinal household survey conducted by the Census Bureau. It is similar to the CPS in terms of cross-sectional sample size, though there is substantial variation in sample size across panels. The main difference between the SIPP and the CPS is that the SIPP follows households for a short period of time (typically two to four years). The first panel of the SIPP was conducted in 1984. Panels have since continued to be collected nearly continuously until 2018.

Table B1: Overview of Data Sets

Data set	Calendar years	Birth cohorts	Wage measure
Census/ACS	1980, 1990, 2000–2016	1945–1986	Hourly wage
CPS	1979–2016	1945–1986	Hourly wage
NLSY	1979–1994, 1996–2012* 1997–2011, 2013–2015*	1957–1964, 1980–1984	Hourly wage
PSID	1968-1997, 1997-2015*	1945–1986	Hourly wage
SIPP	1984–2012, except 2000	1945–1986	Hourly wage

Notes: * indicates biennial coverage. NLSY79 excludes the disadvantaged white and military oversamples.

B.2 Construction of wages

Workers not paid by the hour report weekly earnings in the CPS or monthly earnings in the SIPP. For these individuals we compute their hourly wage by dividing their income by the product of usual hours worked per week and (for the SIPP) weeks worked in the month. For the Census/ACS, the annual hours are determined by the product of usual hours worked per week and weeks worked in the year. For the PSID, annual hours worked are reported by the respondent.

Wages are expressed as hourly rates and in real terms using the CPI-U with 1982-84 as the base year. We drop all wage observations that lie outside the interval [\$2, \$100] in 1982-84 dollars. We construct wages in the following way for each survey:

B.2.1 Census/ACS

In the decennial Census and ACS, we use hourly wages. Nonetheless, we also use annual earnings as a robustness check, and because Baum-Snow and Neal (2009) find that measurement errors in hours worked are most drastic among those who report part-time work (not full-time, full-year work, which is the sample we focus on).

- In all years prior to 2008:
  - hourly earnings are computed as annual earnings divided by the product of usual hours worked and weeks worked
- In years 2008 and beyond:
  - hourly earnings are computed as annual earnings divided by the product of usual hours worked and an imputed value of weeks worked. In 2008 and beyond, exact weeks worked are not reported—only an intervalled version. We impute weeks worked as the midpoint of each of the following intervals: (*i*) 1-13 weeks (impute 7); (*ii*) 14-26 weeks (impute 20); (*iii*) 27-39 weeks (impute 33); (*iv*) 40-47 weeks (impute 44); (*v*) 48-49 weeks (impute 48); and (*vi*) 50-52 weeks (impute 51).

B.2.2 CPS-ORG

In the CPS, we follow the approach outlined by the NBER:[1]

- If worker is paid by the hour:
  - reported hourly wage rate

[1] See http://www.nber.org/morg/docs/cpsx.pdf, p. 32.

- If worker is paid by another unit of time:
  - earnings are reported on a weekly basis and the hourly wage rate is computed as the ratio of weekly earnings to usual weekly hours worked

The CPS imputes a sizable fraction of earnings, which has been shown to bias some important earnings estimates (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006, 2013). In our main specification, we include those with imputed earnings. As a robustness check, we compare the CWP in the CPS with and without imputed earnings. The results are unchanged (see Figure A8). This is likely due to the fact that college degree status is used in the CPS's imputation procedure.

B.2.3 NLSY

In the NLSY, we follow the approach by Ashworth et al. (2017) and Arcidiacono et al. (2016):

- For the NLSY79:
  - Hourly pay at the job employed at the time of interview.
- For the NLSY97:
  - Hourly compensation (including bonuses and tips) at the self-reported main job. If missing, use hourly wage at the self-reported main job.

B.2.4 PSID

In the PSID, we use hourly wage rates implied by annual labor income and annual hours worked.[2] This computation is done by the PSID for Heads and Spouses in all years of the survey (except 1993), and for all individuals in years 1999, 2001, and 2003.

B.2.5 SIPP

In the SIPP, we follow the approach by Altonji, Kahn, and Speer (2016):

- If worker is paid by the hour:
  - average hourly wage rate reported in all surveys during the year
- If worker is paid by another unit of time:
  - total earnings (across all surveys during the year) divided by total hours worked (across all surveys during the year)

[2] See variables `V337`, `V338`, and related.

B.3 Construction of employment

Employment is defined as full-time, part-time, or not employed. Table B2 shows how employment is constructed in each data set.

Table B2: Employment definitions

Data set	Full-time	Part-time
Census/ACS	35+ hours/week	10-35 hours/week
CPS	reports working full-time	reports working part-time
NLSY	35+ hours/week & 40+ weeks/year	positive hours worked or positive weeks worked
PSID	annual hours worked > 1500	annual hours worked $\in [500, 1500]$
SIPP	working 30+ hours/week for 90% of observed non-school months	working 30+ hours/week for 25-90% of observed non-school months

Notes: CPS reports are based on the full-time/part-time labor force status question (`ftpt79` for years 1979-88, `ftpt89` for years 1989-1993, and `ftpt94` for years 1994-2016).

B.4 Construction of degree attainment

Degree attainment is defined for high school diplomas (or GEDs) and bachelor's degrees as listed in Table B3. We include GEDs with HS graduates because not all surveys allow for separate (or reliable) identification of GED status in all years. Prior work by Heckman and LaFontaine (2006) has shown that GED recipients appear to be more similar to high school dropouts than to high school graduates. However, this is likely to have little bearing on our results since educational attainment appears to be quite similar across each of the surveys, as shown in Figures 3 and 4.

B.5 Weights

We use the sampling weights provided by each survey to maximize comparability. Using the sampling weights also helps to correct for oversampling of certain demographic groups. For example, the NLSY oversamples racial and ethnic minorities, and the SIPP (in most panels) oversamples low-income households to fulfill its aim to accurately measure participation in government programs.

Table B3: Degree attainment definitions

Data set	HS diploma	Bachelor's degree
Census/ACS	earned at least HS diploma or GED	completed 4 years of college (1980 census); earned at least a Bachelor's degree (all others)
CPS	earned at least HS diploma or GED	earned at least a Bachelor's degree
NLSY	Highest grade completed ≥ 12 or GED	Highest grade completed ≥ 16
PSID	Highest grade completed ≥ 12 or GED	Highest grade completed ≥ 16
SIPP	earned at least HS diploma or GED	earned at least a Bachelor's degree

Notes:

B.6 How to download each data source

CPS-ORG The CPS-ORG data are constructed as follows:

1. Download raw CPS extracts compiled by the National Bureau of Economic Research (NBER) from http://www.nber.org/morg/annual/.

2. Create consistent measures of wages, employment, and completed schooling for each of the 38 CPS years we use (1979-2016).

Census/ACS The decennial Census and ACS data are constructed using resources provided by Ruggles et al. (2017), as follows:

1. Download raw extracts from https://usa.ipums.org/usa/ (requires account registration and login).

2. Create consistent measures of wages, employment, and completed schooling for the 1980, 1990, and 2000 Censuses, as well as the 2001-2016 ACS.

NLSY The NLSY data are constructed as follows:

1. Download raw NLSY extracts from https://www.nlsinfo.org/investigator/pages/login.jsp (requires account registration and login).

2. Create consistent measures of wages, employment, and completed schooling for both the NLSY79 and NLSY97.

PSID The PSID data are constructed as follows:

1. Download raw PSID extracts from `https://simba.isr.umich.edu/Zips/ZipMain.aspx` (requires account registration and login).
2. Make use of Stata package `PSIDTOOLS` (Kohler, 2015)
3. Create measures of wages, employment, and completed schooling comparable with the other surveys.

SIPP We employ the following steps to construct the SIPP data:

1. Download raw SIPP extracts compiled by the National Bureau of Economic Research (NBER) from `http://www.nber.org/sipp`.
2. Create consistent measures of wages, employment, and completed schooling for each of the 14 SIPP waves we use (1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008).

Online Appendix References

Altonji, Joseph G., Lisa B. Kahn, and Jamin D. Speer. 2016. "Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success." *Journal of Labor Economics* 34 (S1):S361–S401.

Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. 2016. "College Attrition and the Dynamics of Information Revelation." Working Paper 22325, National Bureau of Economic Research.

Ashworth, Jared, V. Joseph Hotz, Arnaud Maurel, and Tyler Ransom. 2017. "Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences." Working Paper 24160, National Bureau of Economic Research.

Baum-Snow, Nathaniel and Derek Neal. 2009. "Mismeasurement of Usual Hours Worked in the Census and ACS." *Economics Letters* 102 (1):39–41.

Bollinger, Christopher R. and Barry T. Hirsch. 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." *Journal of Labor Economics* 24 (3):483–519.

———. 2013. "Is Earnings Nonresponse Ignorable?" *Review of Economics and Statistics* 95 (2):407–416.

Heckman, James J. and Paul A. LaFontaine. 2006. "Bias-Corrected Estimates of GED Returns." *Journal of Labor Economics* 24 (3):661–700.

Hirsch, Barry T. and Edward J. Schumacher. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics* 22 (3):689–722.

Kohler, Ulrich. 2015. "PSIDTOOLS: Stata module to facilitate access to Panel Study of Income Dynamics (PSID)." Statistical Software Components, Boston College Department of Economics. URL https://ideas.repec.org/c/boc/bocode/s457951.html.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. *Integrated Public Use Microdata Series: Version 7.0 [Machine-readable database].* Minneapolis: University of Minnesota.