Online Appendix to

Extending the Race between Education and Technology

David Autor
MIT and the NBER

Claudia Goldin
Harvard University and the NBER

Lawrence F. Katz
Harvard University and the NBER

Table A1: Determinants of the College Wage Premium: 1914 to 2017

	(1)	(2)	(3)	(4)
(College/high school) supply	-0.592	-0.619	-0.640	-0.651
	(0.070)	(0.077)	(0.057)	(0.071)
(College/high school) supply × post-1949				0.0111
				(0.0414)
Time	0.00472	0.0102	0.0106	0.0111
	(0.00182)	(0.00205)	(0.0015)	(0.0026)
Time × post-1949	0.0197			
	(0.0011)			
Time × post-1959		0.0161	0.0160	0.0154
		(0.0010)	(0.0008)	(0.0022)
Time × post-1992	-0.00769	-0.00971	-0.00938	-0.00940
	(0.00135)	(0.00156)	(0.00117)	(0.00118)
1949 Dummy			-0.136	-0.143
			(0.021)	(0.035)
Constant	-0.592	-0.694	-0.717	-0.742
	(0.148)	(0.163)	(0.122)	(0.156)
R^2	0.953	0.945	0.970	0.970
Number of observations	59	59	59	59

Sources and Notes: Each column is an OLS regression of the college wage premium on the indicated variables using a sample covering the years 1914, 1939, 1949, 1959, and 1963 to 2017. Standard errors are given in parentheses below the coefficients. The college wage premium is a fixed weighted average of the estimated college (exactly 16 years of schooling) and post-college (17+ years of schooling) wage differential relative to high school graduates (those with exactly 12 years of schooling). (College/high school) supply is the log supply of college equivalents to high school equivalents both measured in efficiency units. "Time" is measured as years since 1914. The samples used include workers from 16 to 64 years old. The data for 1963 to 2017 are from the 1964 to 2018 March CPS samples. The college wage premium and relative supplies in efficiency units for 1963 to 2017 use the same data processing steps and sample selection rules as those described in the data appendix to Autor, Katz, and Kearney (2008). The college wage premium for 1963 to 2017 uses the log weekly earnings of full-time, full-year workers. The college wage premium series is the same as plotted in Figure 1. The observations for 1914, 1939, 1949, and 1959 append the changes in the college wage premium series from 1915 to 1970 (actually 1914 to 1969) plotted in Figure 8.1 of Goldin and Katz (2008) to the 1969 data point from the March Current Population Survey (CPS) series. The log relative supply observations for 1914 to 1959 similarly append changes in the relative supply of college equivalents from 1914 to 1939 for Iowa and for the United States from 1939 to 1949, 1949 to 1959, and 1959 to 1969 from the Census Integrated Public Use Micro-data samples (IPUMS) using the efficiency-units measurement approach of Tables 8.5 and 8.6 of Goldin and Katz (2008).

Table A2: Changes in the College Wage Premium and the Supply and Demand for College Educated Workers: 1914 to 2017 (100 × Annual Log Changes)

	Changes in the Relative Wage	Changes in Relative Supply	Changes in Relative Demand (σ_{SU} = 1.62)
	(1)	(2)	(3)
1914-1939	-0.56	2.57	1.66
1939-1959	-0.51	2.63	1.80
1959-1979	0.07	3.51	3.63
1979-1999	1.19	2.28	4.21
1999-2017	0.20	1.96	2.28
1939-1979	-0.22	3.07	2.72
1979-2017	0.72	2.13	3.30
1914-2017	0.04	2.60	2.68

Sources: The underlying data are from the 1915 Iowa State Census, 1940 to 1970 Census IPUMS, and 1963 to 2018 CPS Merged Outgoing Rotation Group (MORG) samples.

Notes: The "relative wage" is the log (college/high school) wage differential, which is the college wage premium. The underlying college wage premium series is plotted in Figure 1. The relative supply and demand measures are for college "equivalents" (college graduates plus half of those with some college) relative to high school "equivalents" (those with 12 or fewer years of schooling and half of those with some college). Relative skill supplies are measured in efficiency units and are the same series using for the regressions in Table A1. The log relative demand measure (D_t) is based on equation (2) in the text and is given by $D_t = \ln(L_{S_t}/L_{U_t}) + \sigma_{SU} \ln(w_{S_t}/w_{U_t})$, under the assumption that σ_{SU} = 1.62 based on the estimate from col. (2) of Table A1.

Table A3: Contribution of Changes in Returns to Schooling to Increased Hourly Wage Inequality, 1980 to 2017

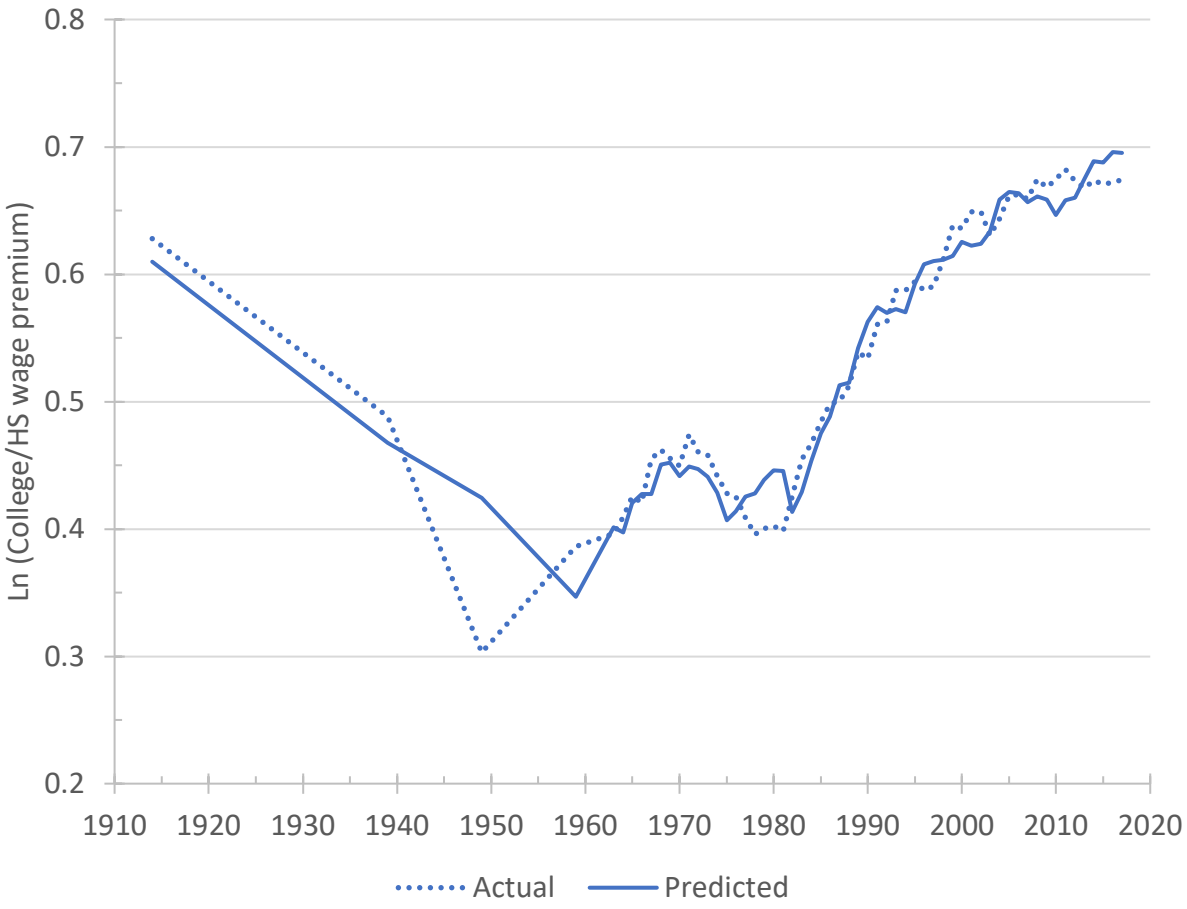
Panel A:	Var ln(w)	90-10	K to 12	College	Post-College
1980	0.250	1.247	0.063	0.077	0.067
2000	0.315	1.436	0.075	0.126	0.131
2017	0.374	1.553	0.062	0.141	0.176

Panel B:	Change in	Change in	Education Return Contribution	
	Var ln(w)	90-10	Var ln(w)	90-10
1980 to 2000	0.065	0.189	0.746	0.649
2000 to 2017	0.058	0.116	0.384	0.483
1980 to 2017	0.123	0.305	0.567	0.572

Sources and Notes: CPS MORG files for 1979 to 1981, 1999 to 2001, and 2016 to 2018. The samples include wage and salary workers aged 18 to 64 years with 0 to 39 years of potential experience using the data processing methods for the CPS May/MORG samples described in the appendix to Autor, Katz, and Kearney (2008). 1980 pools the 1979 to 1981 samples; 2000 pools the 1999 to 2001 samples; 2017 pools the 2016 to 2018 samples. Var ln(w) is the variance of the log hourly wage. 90-10 is the log 90-10 wage ratio. Education returns are estimated for each period from human capital earnings regressions with the log hourly wage as the dependent variable run on a linear spline in years of schooling with break points after 12 and 16 years of schooling; a quartic in experience; race, region, and gender dummies; year dummies; and interactions of gender and the experience quartic.

We examine the role of changing education returns from 1980 to 2017 by first imposing the 2017 returns to schooling on 1980, thereby adjusting individual wages in 1980. We then compare the distributions of actual and adjusted wages in 1980 to determine what wage inequality would have been with education returns at 2017 levels. We adjust 1980 wages to incorporate 2017 education returns by adding to each individual's wage in 1980 the sum of the product of that individual's years of schooling in each category (K-12, college, and post-college) and the difference between the estimated returns to schooling in 2017 and 1980 for that schooling category Wages in 2017 are analogously adjusted by imposing the 1980 education returns. The average of the results of the two simulations is then used. We repeat this approach for 1980 to 2000 and for 2000 to 2017.

Figure A1: Actual vs. Predicted College Wage Premium, 1914 to 2017



Source and Notes: The actual college wage premium is the series plotted in Figure 1. The predicted college wage premium series plots the predicted values for the college wage premium from the regression in col. (2) of Appendix Table A1.

Extended Figure Notes

Figure 1: The clerical/production worker wage ratio series for 1825 to 1875 is based on the mean daily wage of civilian white collar (clerical) and production workers (artisans and common laborers) hired by army forts for each decade from the 1820s to the 1880s using the series in Katz and Margo (2014, Table 1.5). Thus, the 1825 wage ratio is the mean wage ratio for 1821 to 1830, the 1835 ratio is the mean wage ratio for 1831 to 1840, etc. The mean production worker wage for 1825 to 1875 is given by 1/3 times the mean wage of artisans plus 2/3 times the mean wage of common laborers for each decade. The clerical/production worker series for 1890 to 1959 uses the ratio of clerical workers earnings to those of production workers for males in col. 2 of Table 2.2 in Goldin and Katz (2008). The 1939 to 1959 estimates directly use the wage ratios from the 1940 to 1960 Census IPUMS in the bottom panel of Table 2.2. The 1895 to 1939 ratios in the top panel of Table 2.2 are rescaled (multiplied by 1.032) to use the 1939 Census IPUMS ratio as the baseline. The 1890 clerical/production worker wage ratio is backcasted from the 1895 male wage ratio using the proportional change in female wage ratio for 1890 to 1895 in col (1) of Table 2.2. The 1825 to 1875 ratios are rescaled under the assumption of no change in the wage ratio from 1875 to 1890.

The high school wage premium series is the ratio of earnings of high school graduates to those with 8 years of schooling from Table D.1 of Goldin and Katz (2008).

The college wage premium series is a fixed weighted average of the estimated college (16 years of schooling) and post-college (17 or more years of schooling) log wage differential relative to high school graduates (12 years of schooling). The college wage premium series through 2005 is equivalent to the series used in Table 8.2 of Goldin and Katz (2008) and is updated through 2017 using the 2007 to 2018 March CPS. The college wage premium series uses data from the 1915 Iowa State Census, 1940 to 1970 Census IPUMS, the 1964 to 2018 March CPS samples. The data processing and sample selection procedures for March CPS samples follow those in Autor, Katz, and Kearney (2008) updated through the March 2018 CPS to cover earnings data through 2017.

Figure 2: The source data come from the 1940 to 2000 U.S. Census IPUMS and the 2005 to 2018 CPS MORG samples. The figure updates Figure 7 of Goldin and Katz (2007) adding the 1976 to 1987 birth cohorts using the CPS MORG samples for 2005 to 2018. Mean schooling estimates by cohort are adjusted to thirty years of age for the 1876 to 1975 birth cohorts using results of a regression of the log of mean years of schooling by birth cohort–year cell on a full set of birth cohort dummies and a quartic in age, using IPUMS data for 1940–2000. The samples include all native-born residents aged twenty-five to sixty-four.

For further details on the method and data processing, see Goldin and Katz (2008, figure 1.4). Mean years of schooling at age 30 for the 1976 to 1987 birth cohorts are the average of ages 29 to 31 years old for each birth cohort in the CPS MORG files for 2005 to 2018.