DECOMPOSING TRENDS IN THE GENDER GAP FOR HIGHLY EDUCATED WORKERS

By

Joseph G. Altonji, John Eric Humphries,

Yagmur Yuksel and Ling Zhong

August 2025

COWLES FOUNDATION DISCUSSION PAPER NO. 2457

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

NBER WORKING PAPER SERIES

DECOMPOSING TRENDS IN THE GENDER GAP FOR HIGHLY EDUCATED WORKERS

Joseph G. Altonji
John Eric Humphries
Yagmur Yuksel
Ling Zhong

Working Paper 34133
http://www.nber.org/papers/w34133

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2025

We are grateful to Lauren Harris, Lucas Marron, Hoon Pyo Jeon and Owen Rask for outstanding research assistance. We thank Edwin Leuven, Petra Moser, participants in seminars at ANU, CMU, Deakin, Essex, ITAM, Monash, NYU, Northwestern, Seoul National University, Singapore Management University, Sussex, SWUFE, the University of Hong Kong, the University of Melbourne, the University of Oslo, Warwick, Zhejiang University, and participants in conferences organized by the AASLE, AEA, CESifo, CReAM-RFBerlin, NCER-CCER, and SOLE for helpful comments. This research was supported by the Yale Tobin Center for Economic Policy, the Cowles Foundation, and the National Science Foundation (award 2117362). It uses restricted-use data under a license with the National Center for Science and Engineering Statistics, National Science Foundation. All statistics in the paper that are based on that data went through disclosure review. All errors and shortcomings are our responsibility. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Decomposing Trends in the Gender Gap for Highly Educated Workers
Joseph G. Altonji, John Eric Humphries, Yagmur Yuksel, and Ling Zhong
NBER Working Paper No. 34133
August 2025
JEL No. I24, I26, J16, J31, J7, N32

ABSTRACT

This paper examines the gender gap in log earnings among full-time, college-educated workers born between 1931 and 1984. Using data from the National Survey of College Graduates and other sources, we decompose the gender earnings gap across birth cohorts into three components: (i) gender differences in the relative returns to undergraduate and graduate fields, (ii) gender-specific trends in undergraduate field, graduate degree attainment, and graduate field, and (iii) a cohort-specific "residual component" that shifts the gender gap uniformly across all college graduates. We have three main findings. First, when holding the relative returns to fields constant, changes in fields of study contribute 0.128 to the decline in the gender gap. However, this decline is partially offset by cohort trends in the relative returns to specific fields that favored men over women, reducing the contribution of field-of-study changes to the decline to 0.055. Second, gender differences in the relative returns to undergraduate and graduate fields of study contribute to the earnings gap, but they play a limited role in explaining its decline over time. Third, much of the convergence in earnings between the 1931 and 1950 cohorts is due to a declining "residual component." The residual component remains stable for cohorts born between 1951 and the late 1970s, after which it resumes its decline.

Joseph G. Altonji
Yale University
Department of Economics
and NBER
joseph.altonji@yale.edu

John Eric Humphries
Yale University
Department of Economics
and NBER
johneric.humphries@yale.edu

Yagmur Yuksel
Northwestern University
Weinberg College of Arts and Sciences
Department of Economics
cidamyagmur@gmail.com

Ling Zhong
Yale University
ling.zhong@yale.edu

1 Introduction

Since the 1930s, the gender gap for full time workers has decreased but not fully closed. Goldin (2006), Goldin (2021), Ruggles (2015), and many others have documented this fact and studied the factors that lie behind the changes. They include technical change that has raised market productivity of women relative to home productivity, changes in gender norms, reduced discrimination in education and the labor market, and changes in marriage and fertility. These factors operated in part by leading to higher levels of labor market experience and job seniority, fewer labor force interruptions, and choice of better paying occupations. Part of the change in the gender gap, particularly the early convergence, was driven by the closing of the educational attainment gap. Yet, large gender gaps still exist among highly educated workers. While many previous papers have studied the importance of educational attainment in explaining the gender gap over time (e.g., Blau and Kahn, 2017), only a few papers study the role of differential specialization in college and even fewer consider graduate education.

This paper studies the gender gap among full time college educated workers born between 1931 and 1984, focusing on the differential contribution of undergraduate field, graduate degree attainment and field, and field-specific returns. Using rich data from the NSF on college graduates and their labor market outcomes, we decompose trends in the gender wage gap across birth cohorts into: (a) trends in the gender gap in earnings for a given undergraduate and graduate degree field combination, (b) trends in gender differences in the composition of undergraduate and graduate degrees, and (c) trends in a cohort and gender specific "residual" that shifts the gender gap in earnings by the same amount for all college graduates. We start by decomposing the earnings gap across birth cohorts and then distinguish effects that operate through occupation and effects that operate within occupations. We decompose the effect of education into changes in the gender gap in college major choice, changes in the gender gap in graduate degree attainment conditional on college major, and changes in graduate field conditional on college major and having a graduate degree.

Using our decompositions, we present three main results. First, changes in fields of study significantly contribute to the decline in the gender gap when relative returns are held constant across cohorts. However, this decline is offset by changing relative returns to specific fields that favored men over women, reducing the contribution of field-of-study changes to the decline in the gap. Second, while gender differences in relative returns to undergraduate and graduate fields of study contribute to the overall earnings gap, these differences explain very little of its decline over time. Third, much of the convergence in earnings between the

1931 and 1950 birth cohorts is from a declining "residual component" that applies to all degree combinations. The residual component remains stable for cohorts born from 1951 to the late 1970s, after which it resumes its decline.

Our research question requires information on earnings, undergraduate major, graduate degree attainment, graduate field, and occupation covering a long time span. We use multiple waves of the National Survey of College Graduates (NSCG) as this is the only US data source that meets all of these requirements to the best of our knowledge.[1] The NSCG data begins in 1993 and contains earnings observations back to 1990. Thus, those born in the 1930s and early 40s are only observed in the labor market late in their careers. This limitation implies we cannot estimate cohort-specific earnings trajectories without strong assumptions. We address this limitation by supplementing the NSCG data with information about age and cohort effects based on the 1960-2000 decennial Census and 2001-2018 American Community Survey (ACS). As explained below, we use the estimates from the Census/ACS to constrain the interactions between birth cohort and age when we estimate earnings functions in the NSCG. We also use the Census/ACS data to estimate birth and age specific occupational premiums, which we use as dependent variables in our analysis based on the NSCG.[2]

Section 3 provides basic facts about cohort trends in education choice and in the earnings gap. We show that undergraduate field of study changed dramatically across birth cohorts, especially for women. More than 50% of female college graduates from the 1930s and early 40s cohorts majored in Education, English, other Humanities, or Nursing. The share in Education was a remarkable 31.1% for the 1935 birth cohort and still 26.7% in 1947 (throughout the paper, years refer to birth cohorts unless stated otherwise). For men, Business and Engineering accounted for more than 44% of college graduates for the 1935 cohort, dropping to 32.3% in 1949. During this period men moved into lower-paying majors such as Humanities, Political Science, the Social Sciences, and the Arts.[3]

Women born in the 1950s and 1960s shifted toward business related and health related fields. Education fell to 12% by 1962, while Business related degrees grew from less than

[1]The Decennial Census and the Current Population Survey (CPS), primary datasets used for studying long term trends in the gender gap in the US, lack information on field of study. The American Community Survey (ACS) has information on college majors beginning in 2009, but this is too late for the early cohorts. It also lacks information on graduate field.

[2]We investigated two other datasets: the 1972 Survey of Natural and Social Scientists and Engineers and the 1982 Survey of Natural and Social Scientists and Engineers. Unfortunately, the only files that we were able to locate are largely limited to individuals in STEM occupations in the first wave of the survey, even though the base year survey included a subsample from a broader set of occupations.

[3]We start with the 1931 birth cohort because of limits on the availability of earnings in the NSCG, but Gemici and Wiswall (2014) use the 1993 NSCG to show that for the 1920 birth cohort, 84% of female college graduates obtained degrees in the humanities, social sciences, or education, while only 11% obtained degrees in science, mathematics, or engineering and only 5% obtained business degrees.

10% in 1950 to 22% in 1962. These changes reduced the gender gap, but they were partially offset by men shifting from Education and Humanities toward Engineering, Business, and Computer Science and Math.

From the late 1960s through 1984 birth cohorts, the fraction of women majoring in Business declined to 12.2%, while the percentage of women in Psychology, Biology and other Social Sciences (relatively low paying majors) grew to 25%. The percentage of men majoring in Computer Science and Math rose from 6.3% to 10.7% over the same period.

Overall, the trends in college majors are more complicated than a monotonic shift of women toward higher-paying fields previously dominated by men. Fields shifted significantly for both genders over time, but overall, women shifted toward higher-paying majors relative to men, with most of the change occurring before the early 1960s birth cohorts.

Graduate degree attainment also increased dramatically for women compared to men. The share of women with graduate degrees increased from 10% for the 1932 cohort to nearly 40% for the 1984 cohort. Starting 15 percentage points behind men, women ended up 6 points ahead. Graduate field composition also changed. From the 1950s to 70s cohorts, women's shares in Medical and Professional Degrees (MD and Law) caught up with men. Both men and women shifted towards MBAs and engineering degrees.

To get an initial sense of the role of college major and graduate degree attainment in changes in the gender gap over time, we estimate a regression model with a dummy for male, interacted with indicators for the 1931-39, 1940-47, 1948-63, and 1964-94 birth cohorts using the NSCG data. Adding controls for college major reduces the gap by about one third for the early cohorts and one quarter for the later cohorts. Adding controls for graduate field matters less, while controls for occupation further reduce the gap, consistent with the findings of Sloane et al. (2021) for undergraduate degrees.[4]

Section 4 presents the methodology that we use to decompose cohort trends in the gender gap for full time workers. We first use a regression model with gender specific intercepts for each combination of 19 college majors and 21 graduate outcomes (no graduate degree plus 20 graduate fields). The model allows for earnings profiles to flexibly differ by gender, major, and birth cohort. We then use this model to estimate the gender gap in average earnings between ages 28 and 54. Next we use the model to decompose the gender gap for each birth cohort into a "relative returns" gap which captures the gender differences in the returns to specific degree combinations, an "education gap" which captures the differences in college majors and graduate education between genders, a "residual gap" which captures shifts in the

[4]However, Altonji and Zhong (2021) (Appendix Figure B1) find using the NSCG that the male - female difference in the occupational premium for most graduate degrees accounts for only a modest fraction of the male-female difference in earnings for a given graduate degree.

gap that affect all major and graduate degree combinations equally, and a "demographic gap" coming from differences in demographic controls. We separately decompose log earnings and the occupational component of log earnings. We first estimate a specification that assumes the relative returns to fields of study are constant across cohorts. Assuming constant relative returns simplifies the interpretation of the trends in the gender gap. We then turn to a specification that allows relative returns to vary across cohorts.

Section 5 presents the results from our decomposition. Much of our analysis concerns the "education gap" contributed by gender differences in college major, graduate degree attainment, and graduate field. Our results are nuanced and surprising. When we evaluate differences between men and women in education choices using constant relative returns, the education gap is very large for the early birth cohorts, starting at 0.191 in 1931 and peaking at 0.206 in 1935. The latter value is close to the *total gap* for the 1984 birth cohort. The education gap begins to fall after the 1936 cohort. The decline is particularly rapid between the 1940 and 1952 birth cohorts, averaging 0.005 per year, arriving at 0.124 for the 1952 cohort. The education gap continues to decline between the 1952 and 1972 cohorts, but only by an average of 0.003 per year. The education gap then increases slightly until 1977, consistent with evidence in Altonji et al. (2012) and Sloane et al. (2021) for the more recent cohorts based on the ACS. It then declines at an accelerating rate for the most recent cohorts, ending at 0.067 for the 1984 cohort (27.6% of the total gap). The share of the education gap that is within occupations varies somewhat over time but averages 62%.

The surprise comes when we allow relative returns to specific degrees to vary across cohorts. We find that the education gap is 0.144 in 1931 (21% of the total gap) and peaks at 0.17 in 1938. The gap then declines slowly to 0.089 in 1984 (37% of the total gap). Thus, when we allow relative returns to vary across cohorts, changes in the education gap decrease the gender gap by only 0.055 between the early 1930s and the early 1980s. This relatively modest decline of the education gap contradicts the findings discussed above for the constant relative returns specification. In that case we find that the partial convergence between education choices was an important factor in long term trends in the gender gap.

Digging deeper, we show that the modest decline in the education gap is the net result of two offsetting cohort trends. When we evaluate the education gap using relative returns for our base cohort (1961), we find a large decline in the size of the education gap across cohorts. The pattern is very similar to what we obtain in the constant returns case. However, this decline is offset by cohort trends in the relative returns to specific fields that worked in favor of men. Returns to degree types dominated by men rose across cohorts while those dominated by women declined. Our results are in part a reflection of changes in the occupational pay structure. Using the Census/ACS data between 1960 and 2018, we find that relative earnings fell in occupations typically associated with a degree in Education (such as teachers) and rose in occupations associated with Engineering degrees (such as engineers and managers). These shifts in relative returns to majors in favor of traditionally "male" majors are consistent with Gemici and Wiswall (2014)'s results using a very different estimation strategy.

What specific college majors contribute the most to the narrowing of the education gap? To answer this question, we use the constant returns specification to measure the contributions of each of the 19 undergraduate majors, aggregating over various graduate degree outcomes conditional on major. Between the 1930s and 1960s cohorts, the decline at constant prices is driven by a drop in gender differences in the probability of majoring in Education and the Humanities, with Business and Fine Arts also playing a role. The decline after the late 70s birth cohort is the net effect of partially offsetting changes in several fields, with Business, Biology, Nursing, and Health contributing to the decline and Engineering and Computer Science/Math working in the opposite direction.

We also decompose the education gap into the contributions of college major, graduate attendance, and graduate field. Using the constant returns specification, we find that changes in gender differences in college major can account for 54% of the decline in the education gap. The large change in graduate attendance rates in favor of women contributes 40%, while changes in gender differences in graduate field probabilities contribute only 6%.[5]

Our second set of results concern how gender differences in the relative returns to undergraduate and graduate degrees contribute to the gender gap. We find that these differences contribute very little to the decline.[6] This holds true despite large observed changes in the popularity of specific degrees. In the constant relative returns case, the gap contributed by gender differences in relative returns is essentially constant at about 0.20, where the 0.20 includes a gender gap in returns that is common to all fields in 1961. About 75% of the relative return gap is within occupation, with little variation across cohorts.

The picture is similar in the variable relative returns specification. With that specification, we find that changes across cohorts in the difference between men and women in returns to specific degree combinations (e.g., education with no graduate degree) leads to an increase in the gender gap of about 0.02 between the early 1930s and the early 1960s. However, this increase slowly reverses between the early 1960s and the 1984 birth cohorts.

Finally, we find that much of the large gap in earnings for the earliest cohorts is due to a common cohort specific component that shifts the gender gap in earnings by the same

[5] Results are similar when we perform the decomposition using relative returns for the 1961 cohort obtained using the earnings model that allows relative returns to vary across cohorts.

[6] We evaluate the contribution of male-female differences in education using male relative returns. We evaluate the contribution of male-female differences in relative returns to specific degree combinations using female education choices. The results are not sensitive to this choice.

amount for all college graduates. We usually refer to the gap in the cohort specific common component as the "residual gap" (which we normalize to be 0 in 1961). Using the constant returns specification, we find that between 1931 and 1950, the residual gap declined by about 0.23, compared to an overall decline of about 0.312. Separate decompositions of the gap in log hourly wage rates suggest that gender differences in annual work hours conditional on full time status are part of this decline. From the mid fifties forward, there is little change in the residual gap.[7] In the variable returns case the cohort trends are similar, but the decline in the gender gap is larger and the residual component is larger, with the residual gap declining an additional 0.07 from the early 1970s to 1984.

This paper relates to a large literature studying the gender earnings gap among highly educated workers and the role of college majors in contributing to this gap. The literature on the role of college major choice in the gender gap is reviewed in Altonji et al. (2012), Altonji et al. (2016), and Patnaik et al. (2020) and includes Brown and Corcoran (1997), Black et al. (2008), Zafar (2013), Gemici and Wiswall (2014), Bronson (2019), and Imberman et al. (2024), among others. Altonji et al. (2012), Gemici and Wiswall (2014), Patnaik et al. (2020), and Sloane et al. (2021) study long term trends in college majors among men and women and demonstrate partial convergence. For example, Patnaik et al. (2020) discuss the change in rates of college attainment among men and women from the 1940 to the 1993 birth cohort. They additionally document differences in degree attainment, showing that from 1940 to 1960 there was a large reduction in the fraction of women majoring in Humanities, Social Science, and Education, though a gap of around ten percentage points persists between men and women after the mid 1960s. They similarly show an increase in Business and Economics degrees among women from 1940 to the mid 1960s, a small increase in STEM majors for women from 1950 to 1960, and an increase between the 1985 to 1993 birth cohorts. These papers and the papers they cite build on a large prior literature on the gender gap between men and women, how it has evolved over time, and how it depends on education (Goldin, 2006, 2014; Blau and Kahn, 2017). We provide gender-specific estimates of trends in college major choice, graduate degree attainment, and graduate field for the 1931-1984 birth cohorts. Thus we consider graduate education, which has grown dramatically, as well as undergraduate field. In this paper we do not provide evidence on the causes behind the changes in major choice, graduate school attendance, and graduate field that we observe, which is the subject of several papers.[8]

[7] In 1931, 86% of the cohort residual gap is within occupation. Both the within and across occupations residual gaps then decline to zero by 1953 (relative to 1961). Thus, much of early steep decline in the birth cohort specific residual gap is within occupation.

[8] See Zafar (2013), Gemici and Wiswall (2014), Bronson (2019) and Abramitzky et al. (2024) for discussion of mechanisms, empirical evidence, and references to the literature. Changes in the financial return to specific

A smaller set of papers have studied the contribution of changes in college major choice to long term trends in the gender gap among college graduates. Using the ACS, Altonji et al. (2012) conduct a Blinder-Oaxaca decomposition by graduation year from the mid-1970s to 2008, showing that differences in coefficients explain approximately half of the male-female gap in hourly wage rates, while differences in college majors can explain the other half. They find that changes in gender composition lead to a narrowing of the gender differential over the 1970s, although the extent is larger when female degree weights are used with male wage coefficients. Gemici and Wiswall (2014) use a structural model to study the generational change in college enrollment, college major, and wage rates for men and women. They find that the female-male wage ratio among college graduates rose by 8% between the 1940 and 1960 birth cohorts. The increase is the net result of a 15% increase driven by a fall in the value of time at home primarily affecting the education choices of women, and a 5% decrease due to a rise in the price of science and business skills relative to humanities and teaching. We find that the relative returns to majors worked in favor of men, which is consistent with Gemici and Wiswall's results, although our methodology is very different.

The paper most closely related to our work is Sloane et al. (2021). Like Altonji et al. (2012), they use the ACS to study gender differences in college major and how these evolve over time. They study both differences in college major and differences in the mapping of college major into occupation by gender. They find that women choose majors with lower average earnings (based on the men in that major), and then sort into occupations with lower earnings conditional on their chosen major. They find that these differences narrow across birth cohorts, but that women still choose majors and occupations with lower expected earnings (for men). Because Sloane et al. use ACS data, they cannot study trends as far back in time or consider graduate fields of study or degrees, but they are able to more exhaustively consider the interactions between major and occupation given the large sample sizes in the annual ACS surveys.

majors likely changed substantially for women as employment levels rose, leading to shifts to higher paying fields. Altonji et al. (2016) and Patnaik (2020) survey the literature on the response of college major to relative returns. Recent papers include Abramitzky et al. (2024) who find that a policy change whereby Israeli kibbutzim switched from a policy of equal sharing to distribution based on individual earnings led to a substantial shift toward bachelor's degrees in higher paying fields for both men and women, Bronson (2019) who considers major choice within a structural model of education choice, marriage and the labor market, and Blom et al. (2021) who study effects of business cycle conditions on major choice by gender. Goldin et al. (2006) study long term trends in the gender gap in BA attainment rates but do not focus on field of study.

2 Data

We use three datasets. The National Center for Science and Engineering Statistics' National Survey of College Graduates (NSCG) is our primary data source. We draw heavily upon the data construction for earnings, occupation, college major, graduate degree attainment and field described in Altonji and Zhong (2021) and Altonji et al. (2023). We use surveys that are nationally representative, namely the NSCG data from 1993, 2003, and 2010 to 2019. With the NSCG 1993 wave, we also use retrospective information about occupation in 1988 and matched 1990 Census data on earnings and occupation. In most of the empirical analyses we restrict the sample to people born between 1931 and 1984. We deflate earnings and wage measures used in the paper to 2013 dollars. The earnings analysis is restricted to full time workers aged 23 to 59 who are not enrolled in school and who earn more than $5,000. Details are in Appendix A along with summary statistics by decade of birth.

The second dataset is the Census/ACS data. We use the Census 5% data from 1960 to 2000, and the 2001 to 2018 ACS data, restricting the sample to full time workers with four or more years of college in the 1960 to 1990 samples, and people with a bachelor's degree or higher in the samples after 1990. In the empirical analysis we restrict the sample to full time workers who were not enrolled in school and who earn more than $5,000. We exclude people who work less than 35 hours per week or 40 weeks per year. Top-coded earnings are adjusted by multiplying the top coded value with 1.5. We remove all imputed values in the variables we use.

We use the Census/ACS data to estimate occupational earnings premiums, which serve as the dependent variable in some of our analyses using the NSCG. As we explain in Section 4.3.2, we also use the Census/ACS to estimate the regression coefficients relating earnings to a polynomial in birth cohort and age. We use the coefficients to constrain the interactions between birth cohort and age when we estimate earnings functions in the NSCG.

While the NSCG is our main source of information about field of study, we also provide supplementary analyses using estimates of the distribution of fields of study by gender, degree type, and year of graduation calculated from institutional level surveys. The 1966 to 1985 Higher Education General Information Survey (HEGIS) and the 1985 to 2019 Integrated Postsecondary Education Data System (IPEDS) provide the annual number of degrees conferred from 1966 to 2019 (see Appendix C). We use HEGIS and IPEDS to generate alternative estimates of the marginal distribution of BA majors by gender for birth cohorts after the early 1940s.[9] We refer to this data source as the HEGIS/IPEDS data in the rest

[9] We make minor adjustments to achieve consistent degree type and field of study classifications across the HEGIS, IPEDS, and NSCG data. Data before 1966 is only available for some fields of study.

of the paper.

3 Descriptive trends in the gender gap

3.1 Trends in educational attainment

Figure 1 shows the trends in college majors for men and women by birth cohort. The blue dash line shows the proportion of men and the orange solid line is for women. Panels A-E report the three year moving average over birth cohorts by gender for the five majors that account for most of the changes. Panel F shows the trends in graduate degree attainment (among college graduates) at age 35 by birth cohort and gender. In Appendix Figures B.1 and B.2, we show the college major trend for 19 majors and 19 graduate fields.[10] The two vertical dashed lines separate the birth cohorts from 1931 to 1984 into three periods in which the patterns of the trends vary.

More than half of women born in the 1930s and 1940s were majoring in Education, English, Nursing, or Other Humanities. Women's major distribution remained fairly constant across these cohorts except for a slow decline in Education, which fell from 31.1% of female graduates born in 1935 to 26.7% by 1947. There are also notable fluctuations in English/Languages/Literature, which reached its all-time high of about 14.6% of women in 1932 and again in 1945 before falling to 12.5% by 1949. Men in these early cohorts were primarily majoring in Business (25% in 1935) and Engineering (19.2%), but these majors declined to 22.1% and 10.8% by 1949, respectively. Between the 1935 and 1950 birth cohorts, men diversified into lower-paying majors such as Humanities, Political Science, Social Sciences, and the Arts; these majors combined comprised 13.5% of men born in 1935 and 20% in 1949. Education and Humanities remained the third and fourth most popular majors for men in this period, making up a combined 15.3% in 1935. Both men and women of these cohorts saw some early growth in Computer Science and Math.

College graduates born in the 1950s and early 1960s were moving into higher-paying college majors. Women drastically shifted away from their early majors and into Business, alongside modest growth in Marketing, Communications & Journalism, Biology, Health, and Engineering. In fact, as Education fell from 27% of female graduates born in 1950 to 12% by 1962, Business grew from less than 10% to 22%, overtaking Education as the most popular major for women in 1958. Meanwhile, men saw renewed growth in Engineering and

[10]We do not report college major trends for men in Nursing and women in Marketing and Other Social Sciences for the earliest birth cohorts because the cell counts are too low. For graduate degrees, we merge together Nursing and Health Administration for male graduates and suppress values for some early birth cohorts for similar cell count reasons.

Business and sustained growth in Computer Science & Math, which emerged as the third most popular major for men born in the 1960s. The once-steady male representation in the Education and Humanities majors declined below 5% by the 1960s.

The late 1960s through the 1980s birth cohorts saw a notable decline in female Business majors, which fell to 12.2% of those born in 1984 but remained the most popular major for women. Women experienced growth in Other Social Sciences (9.5% in 1984), Biology (8.4%), and Psychology (7.1%), which emerged as the third, fourth, and fifth most popular majors for women, trailing behind the declining Education major (11.1%). During the same time period, the share of men majoring in Computer Science & Math rose from 6.3% to 10.7%. The proportion of male graduates also grew in Biology. It stagnated in Engineering at around 16% and declined in Business, which dipped below Engineering in 1981.

Panel F of Figure 1 shows that only 10% of female college graduates born in 1932 obtained advanced degrees, compared to 25% of their male counterparts. Female college graduates saw huge gains in graduate school attendance across these six decades, surpassing men in the 1966 cohort and reaching almost 40% advanced degree attainment for college graduates born in 1984. Men, on the other hand, have had a fairly flat profile, reaching their highest levels of graduate degree attainment of 34% in 1942 and regaining this rate in 1983. Both men and women saw growth in graduate school between the 1930s and early 40s cohorts, followed by a fifteen-year period of decline starting with the 1945 birth cohort.[11] Women's rates grew rapidly from the 1960s onward and more than offset the earlier decline, while men's rates stagnated until 1970 before seeing growth again in the late 1970s and throughout the 1980s. Appendix Figure B.2 documents large shifts across cohorts for both men and women in graduate degree fields of study.

3.2 Trends in gender differences in earnings and occupational sorting

To provide some initial evidence on the gender gap, we use a simple log earnings regression to study how the gap changes across cohorts as we add education and occupation controls. The rows of Table 1 show estimates of the coefficient on a male indicator for those born between 1931 and 1939, those born between 1940 and 1947, those born between 1948 and 1963, and those born between 1964 and 1984. Column 1 includes the baseline controls, which consist of parents' education dummies, a cubic in age interacted with a male dummy, a cubic in birth year, and race indicators interacted with gender. The age cubics are constructed such

[11]This downturn among college graduates beginning with the late 1940s cohorts is also evident in CPS data for college graduates aged 34-36. (Not shown)

that the coefficients in the table capture the estimated residual earnings gap averaged over age 28 to 52.[12] With the first regression as the base, we add detailed college major dummies (column 2), detailed graduate field dummies (column 3), and occupation dummies (column 4). The nested regression specifications allow us to identify the additional contribution of each set of controls to explaining the gender gap. All regressions compare individuals with college degrees working full time.

The baseline regression coefficients show that, without education or occupation controls, the estimated log earnings gap is 0.49 (0.025) for the first birth cohort group. The gap declines to 0.45, 0.37, and 0.34 for the second, third, and fourth birth cohort groups, respectively. Next, column 2 shows that controlling for college majors shrinks the gender gaps for the four cohort groups to 0.35, 0.33, 0.28, and 0.27 respectively. This is a 29% reduction in the earliest cohort and a 21% reduction in the fourth (1964-84) cohort group. The difference between column 2 and column 3 shows that graduate degree attainment and field explain another 0.023 to 0.032 of the gender gap for the first three cohorts, but only 0.014 for the 1964-84 group. This result suggests that, among college graduates, differences in college major and graduate degrees account for a substantial portion of the gender gap in earnings, but the contribution decreases in the younger cohorts. Moreover, after controlling for college major and graduate field, large gaps still persist across all four birth cohort groups. Adding occupation dummies to the controls further reduces the gaps for all four cohorts by approximately 0.10 to 0.21, 0.20, 0.17 and 0.16, respectively. This suggests that occupational sorting also plays an important role in the gender pay gap of highly educated workers, consistent with prior work.

Overall, Table 1 provides initial evidence on how the gender gap among college graduates has evolved across birth cohorts. However, these regressions impose several strong functional form assumptions, such as the returns to field of study and graduate school being the same for men and women, and that the returns to graduate school do not depend on college major. We extend this simple decomposition in the next section.

4 Birth cohort specific decompositions of the gender gap in earnings: methods

This section develops our decomposition of the gender gap in earnings. In section 4.1, we present the regression model of earnings that is used in the decompositions. In section 4.2, we present the decomposition formula for the case in which gender differences in relative

[12] We also use a regression index of interactions between age and year of birth as a control. The parameters that define the index are estimated using the Census/ACS data, as discussed in 4.3.2.

returns to degrees are constant across cohorts. Sections 4.3 and 4.4 discuss estimation of the key inputs into the decomposition: the earnings model parameters and the cohort and gender specific college major and graduate field probabilities. Section 4.5 considers the case in which gender differences in effects of c and g on earnings vary across birth cohorts. Section 4.6 discusses the decomposition of the occupation specific component of earnings.

4.1 The model of earnings and the occupation component of earnings

We start with some notation. Let i denote the individual, $b(i)$ denote birth cohort, t denote the calendar year, and a_{it} denote age. We use $c \in \{1, ..., \mathcal{C}\}$ as the index of the undergraduate major, and $g \in \{0, ..., \mathcal{G}\}$ as the index of graduate degree type. We aggregate undergraduate degrees into 19 majors and aggregate graduate degrees into 20 fields, so in the empirical work $\mathcal{C} = 19$ and $\mathcal{G} = 20$. The value $g = 0$ indicates no graduate degree. Later we use the dummy variables $C_{c(i)}$ and $G_{g(i)}$ for $c(i) = c$ and $g(i) = g$. We use $G_i = 1(g(i) > 0)$ to indicate that the individual has a graduate degree. The gender index s is f for females and m for males. The gender dummy variables $S_{s(i)}$ are 1 if $s(i) = s$ and zero otherwise. We use $o \in \{1, ..., \mathcal{O}\}$ as the index of occupation. Variables $c(i)$, $g(i)$ and $o(it)$ denote i's choice, but we usually leave the i and it arguments implicit.

In anticipation of decompositions of the earnings gap into within occupation and between occupation components, we start by writing log earnings Y_{it} as the sum of an occupation component $\bar{y}^{ba(it)}_{o(it)}$ and a within occupation component $\tilde{y}_{it}$:

$$Y_{it} = \bar{y}^{ba(it)}_{o(it)} + \tilde{y}_{it}.$$

We define the occupation specific components to be the same for men and women, meaning all within-occupation gender differences are captured in $\tilde{y}_{it}$. We return to this point below. We consider a specification in which the $\bar{y}^{ba(it)}_{o(it)}$ are constant across birth cohorts, which simplifies the decompositions, and a specification in which they vary with birth cohort and age.[13]

The regression equation for log earnings is

$$Y_{it} = \alpha^{sb}_{cg} + X^s_{1it}\beta^s_1 + X^s_{2it}\beta^s_2 + Z^s_i\Gamma^s + u_{it}. \tag{1}$$

There is no separate constant term in equation (1).

[13]Because $t = a + b$, we implicitly allow the occupation specific components to depend on both t and b. See section 4.5 below.

In our "constant returns" specification, we assume that the gender specific returns to degrees, α_{cg}^{sb}, shift by the same amount, α^{sb}, across cohorts. Thus relative returns across cg pairs are constant, and we can express α_{cg}^{sb} as

$$\alpha_{cg}^{sb} = \alpha_{cg}^{s0} + \alpha^{sb}, \tag{2}$$

where we normalize around the 1961 birth cohort returns, α_{cg}^{s0}. We specify α^{sb} to be a gender specific cubic birth year polynomial, so it captures differences across cohorts in career earnings that are independent of choice of cg. The vector X_{1it} contains the triple interaction among gender, college major, and a cubic age polynomial. The demographic control vector Z_i^s contains parental education levels and interactions between gender, race, and Hispanic dummies. The excluded categories for both men and women refer to white non-Hispanics whose parents have high school degrees.

The control vector X_{2it}^s contains interactions between b_i and a_{it} up to the second order plus $b_i^3 \times a_{it}$ and $b_i \times a_{it}^3$, all interacted with gender $S_{s(i)}$. We include these terms because the gender specific age profiles are likely to vary with birth cohort. If age profiles differ by cohort, differences in earnings gaps at a particular age may be a poor guide to changes in cohort differences in life cycle earnings. For this reason, we normalize the gender specific age polynomial terms in X_{1it}^s and X_{2it}^s so that the intercepts α_{cg}^{sb} are defined to give equal weight to each age between 28 and 52.[14]

Note that because calendar time t equals $b_i + a$, t, $t^2$and t^3 are perfectly collinear with the cohort cubic, age cubic, and cohort-age interaction terms that we include in the earnings model.[15] This means that age and cohort variables account for secular trends in earnings due to general productivity changes, changes in gender discrimination, changes in gender norms, and other factors in ways that depend on gender but not cg. Therefore, the cohort differences in labor market outcomes capture both secular change and true cohort effects.

We normalize birth cohort around 1961, so that b_i is 0 for the 1961 birth cohort. This normalization of b_i, our choice of reference groups for Z_i^s, and our treatment of the age polynomials together imply that the α_{cg}^{s0} refer to the 1961 birth cohort for a non-Hispanic white individual whose mother and father have only high school diplomas. The coefficients on parental education do not depend on gender. Given Z_i^s, the mean of log earnings between age 28 and 52 for an individual with a degree in cg from cohort b is $\alpha_{cg}^{s0} + \alpha^{sb} + Z_i^s\Gamma^s$.

[14]Our choice balances a desire to cover most of the period when people normally work against concerns about the age distribution of our sample for the early and late cohorts. See section 4.3.2. We obtain very similar results when we define α_{cg}^{sb} to refer to 26-59 (not reported).

[15]Even t^4 is also almost perfectly collinear with the cohort and age cubics and the cohort-age interactions in our model. The R^2 of a regression of t^4 on them is 1 to five decimals over the range of t, b and a in our sample.

The occupation component $\bar{y}_{o(it)}^{ba(it)}$ is equal to its expectation conditional on a_{it}, t, b, cg, b, race/ethnicity and s plus the error term $\bar{u}_{it}$:

$$\bar{y}_{o(it)}^{ba(it)} = \bar{\alpha}_{cg}^{sb} + X_{1it}^{s}\bar{\beta}_1^s + X_{2it}^{s}\bar{\beta}_2^s + Z_i^s\bar{\Gamma}^s + \bar{u}_{it}\,. \tag{3}$$

Note that the value of $\bar{y}_{o(it)}^{bat}$ itself does not depend on c(i), $g(i)$ or $s(i)$. The variables c, g and s influence the conditional mean of $\bar{y}_{o(it)}^{bat}$ only through their influence on occupation choice. The regression model for $\tilde{y}_{it}$, the within occupation component of earnings, is implicitly defined as the difference between equations (1) and (3).

4.2 Formulas for birth cohort specific decompositions of the gender gap in earnings

In this section, we provide formulas to decompose the gender gap in earnings into the contributions of (1) gender differences in choice of college major and graduate education, (2) gender differences in relative returns to degrees, (3) cross cohort changes in gender differences in earnings that are common to all degree choices, and (4) cohort differences in demographic characteristics. Here we present the formula for the constant return case in which the relative returns across cg pairs are constant across cohorts. We turn to the varying relative returns case in section 4.4.

Define the gender gap to be

$$GAP(b) = E[Y|b,m] - E[Y|b,f].$$

When relative returns to cg are constant, the expected value of career log earnings for a person with a degree in cg from cohort b and characteristics Z_i^s is $\alpha_{cg}^{s0} + \alpha^{sb} + Z_i^s\Gamma^s$. Let

$$\Delta Z_b = E[(Z_i^m\Gamma^m - Z_i^f\Gamma^f)|b]$$

denote the mean difference between males and females in the earnings regression indices of the demographic variables, and let P_{cg}^{sb} denote the conditional probability $Pr\,(c(i) = c, g(i) = g \mid s, b)$. Then given equation (2), one may write $GAP(b)$ as

$$GAP(b) = \sum_{cg} \left(\alpha_{cg}^{m0}P_{cg}^{mb} - \alpha_{cg}^{f0}P_{cg}^{fb}\right) + (\alpha^{mb} - \alpha^{fb}) + \Delta Z_b.$$

$GAP(b)$ may be rearranged to provide a Blinder-Oaxaca style decomposition using the male education coefficients and the female cg probabilities:

$$
\begin{aligned}
GAP(b) = &\sum_{cg} \left(\alpha^{m0}_{cg} - \alpha^{f0}_{cg}\right) P^{fb}_{cg} && \text{relative return gap} \\
&+ \sum_{cg} \alpha^{m0}_{cg} \left(P^{mb}_{cg} - P^{fb}_{cg}\right) && \text{education gap} \\
&+ \alpha^{mb} - \alpha^{fb} && \text{cohort } b \text{ residual gap} \\
&+ \Delta Z_b. && \text{demographic control gap}
\end{aligned} \tag{4}
$$

In Appendix E, we provide results using the female education coefficients and male probabilities. The first term, which we call the relative return gap, is the portion of the gap for birth cohort b explained by the gender differences in relative returns to degrees, evaluated using the female degree probabilities for cohort b. Each of the terms $(\alpha^{m0}_{cg} - \alpha^{f0}_{cg})$ includes a component that is common to all fields in 1961.

The second term, which we refer to as the education gap, is the portion explained by the gender gap in degree distributions evaluated using the returns for males. Recall that in the constant returns specification, we assume that the gender-specific returns to degrees are cohort-invariant, while the distribution of degrees fluctuates across cohorts. Therefore, the fluctuations in the first term across cohorts will be due entirely to the varying degree distribution for females. Fluctuations in the second term reflect shifts in the gender difference in the education distributions. It is the key term in our analysis.

The third term captures gender differences in cohort specific shifts in earnings affecting all cg categories equally. It is 0 when $b = 1961$, because both α^{mb} and α^{fb} are normalized to 0 for that cohort. The cohort b residual gap term captures the change in the unexplained gender gap in earnings of college graduates. It could arise from a number of factors mentioned in the introduction, including changes in discrimination. The final term, ΔZ_b, represents changes in the gender gap that are due to changes across cohorts in gender differences in race and ethnicity, and parental controls. It turns out to be small in magnitude.

We further decompose the education gap into the contribution of differences in college major, differences in graduate school attendance conditional on college major, and differences in graduate field conditional on graduate school attendance and college major. Because G_i is a binary indicator equal to 1 when $g_i > 0$, we have

$$
\begin{aligned}
P^{sb}_{cg} &\equiv \Pr{}^{s}_{b}\left(c(i) = c, g(i) = g\right) \\
&= \Pr{}^{s}_{b}\left(g(i) = g | G(g(i)), c_i = c\right) \times \Pr{}^{s}_{b}(G(g_i)|c_i = c) \times \Pr{}^{s}_{b}(c_i = c).
\end{aligned}
$$

Simplifying the notation, the education gap may be decomposed as follows:

$$
\begin{aligned}
&\text{Education Gap } (b) = \\
&\sum_{cg} \alpha^{m0}_{cg} \times \left(P^{mb}_{g|G,c} - P^{fb}_{g|G,c}\right) \times P^{fb}_{G|c} \times P^{fb}_{c} && \text{grad field gap} \\
&+ \sum_{cg} \alpha^{m0}_{cg} \times P^{fb}_{g|G,c} \times \left(P^{mb}_{G|c} - P^{fb}_{G|c}\right) \times P^{fb}_{c} && \text{grad enroll gap} \\
&+ \sum_{cg} \alpha^{m0}_{cg} \times P^{fb}_{g,G|c} \times \left(P^{mb}_{c} - P^{fb}_{c}\right) && \text{BA field gap} \\
&+ \Delta ED^{23}_{b}. && \text{approx. error}
\end{aligned}
\tag{5}
$$

The first term is the contribution of differences in graduate field conditional on graduate school attendance and c. The second term is the contribution of differences in graduate degree attainment conditional on c. The third term is the contribution of gender differences in c. ΔED^{23}_{b} is an approximation error that turns out to be negligible.[16]

4.3 Estimation of the Degree Probabilities and the Earnings Model

In this section we first discuss how we estimate the degree probabilities. We then turn to earnings. We describe how we estimate the age-birth cohort interaction term and discuss our use of OLS. Finally, we present evidence on trends in selection into the sample of college graduates who work full time.

4.3.1 Estimation of the Degree Probabilities

For each major c, we estimate P^{fb}_{c} by fitting a b-spline to the microdata on $C_{c(i)}$ for the female sample. The spline basis is major and gender specific. We estimate $P^{fb}_{g|c}$ by fitting a b-spline to $G_{g(i)}$ using females who majored in c. $P^{fb}_{G|c}$ is $(1 - P^{fb}_{g=0|c})$, recalling that we define the graduate field g to be 0 for those with no graduate degree. We use these estimates to construct estimates of $P^{fb}_{g|G,c}$ and $P^{fb}_{g,G|c}$. We use the same approach for males. We employ sample weights in this analysis, which is important because the NSCG over represents STEM fields. Appendix G.2 shows that the decompositions of the earnings gap are similar when we estimate P^{fb}_{c} and P^{mb}_{c} as a three year moving average of b, s specific major probabilities.

[16]This approximation error is the sum of terms involving second order and third order interactions among the gender differences in P_c, $P_{G|c}$, and $P_{g|G,c}$.

4.3.2 Estimation of Interactions Between Age and Birth Cohort using the Census/ACS data

In the NSF data, the age-birth cohort interaction term $X^s_{2it}\beta^s_2$ is identified for the earliest and latest cohorts only through the imposed functional form. Our regression sample covers graduates aged 23–59, with earnings observed from 1989 to 2019. Consequently, individuals born in 1931 appear only at age 58, while those born in 1984 appear only between ages 23 and 35. Although the functional-form assumption technically identifies the full age-cohort profile, it requires substantial extrapolation for combinations of b_i and age_{it} that fall well outside the observed data. To address this, we construct an age-birth cohort regression index using the nationally representative decennial Census and ACS waves between 1960 and 2018 and use it to restrict $X^s_{2it}\beta^s_2$. The first step is to estimate a regression model that resembles equation (1) in the decennial Census and ACS data from 1960 to 2019, imposing the same age and education restrictions we impose on the regression sample in the NSF. The model is

$$Y_{it} = \varphi^s_0 S_{s(i)} + \varphi^s_1 S_{s(i)} G_{it} + X^s_{1it}\beta^{s*}_1 + X^s_{2it}\beta^{s*}_2 + Z^{s*}_i \Gamma^{s*} + u_{it}, \tag{6}$$

where G_i is an indicator for whether i has graduate education.[17] We use G_i rather than $C_{c(i)}G_{g(i)}$ because college major and graduate degree are not available across the Census/ACS waves. The control vector Z^{s*}_{it} consists of race and Hispanic dummies interacted with $S_{s(i)}$, a gender-specific cubic birth year polynomial, and a gender-specific cubic age polynomial. The control vector X_{2it} contains gender-specific age-birth year interactions up to the second order plus $b^3_i \times age_i$ and $b_i \times age^3_i$. We impose the restriction

$$X^s_{2it}\beta^s_2 = \beta^s_3[X^s_{2it}\hat{\beta}^{s*}_2]$$

and replace $X^s_{2it}\beta^s_2$ with $\beta^s_3[X^s_{2it}\hat{\beta}^{s*}_2]$ in equation (1). The use of the index constrains the shape of the interactions between age_i and b_i to conform to what is observed in the Census/ACS between 1960 and 2019. Note, however, that we are extrapolating beyond the Census/ACS sample for the most recent cohorts.

[17] For the 1960-1990 Census, $G_i = 1$ if the individual has five or more years of postsecondary education. For the 2000 Census and the ACS waves, it is 1 if the individual has a master's degree, a graduate professional degree, or a doctoral degree.

4.3.3 Use of OLS

In our base specification, we estimate the earnings model by OLS.[18] We do so despite concerns about bias due to selection into particular fields of study. In the case of college major, there is no practical alternative to OLS in our data.[19] In the case of the return to graduate degrees conditional on undergraduate field, Altonji and Zhong (2021) and Altonji et al. (2023) assume the $\alpha_{cg}^{s0} = \alpha_{c}^{s0} + \alpha_{g}^{s0}$ and use a group fixed effects strategy that they call FEcg. The approach is to add fixed effects for c interacted with whether the individual eventually obtains a graduate degree in g. It amounts to using experience-adjusted comparisons of the pre and post graduate school earnings of individuals who obtain a graduate degree to identify α_{cg}^{s0} , like an individual fixed effects approach. The big advantage of FEcg in our application is that it does not require that both pre and post graduate school earnings be observed for a given individual. Our results for the decomposition are similar when we impose the additive specification and use FEcg to estimate α_{g}^{s0} (results not included). For simplicity, we use the OLS estimates throughout our analyses. However, the FEcg approach does not address bias in the estimates of the returns to college major.

4.3.4 Selection into college, graduate school, and working full time

Our analysis focuses on college graduates who work full-time, a population that has changed over the last several decades. These changes may introduce selection bias into our gender decomposition, particularly if gender differences in who goes to college or works full time change across cohorts. Here we briefly summarize our evidence on selection into our sample, which we develop more fully in Appendix L.

The NSF data does not contain measures of ability and only includes college graduates. We therefore use data from Project Talent (PT), the NLS72, NLSY79, and NLSY97 to study

[18] We use sample weights to address choice based sampling arising from the fact that the sample selection probabilities in the 1993 and 2003 NSCG are based in part on occupation in the 1990 and 2000 Census (respectively). For some cg combinations we have fewer than 30 people, in which case the estimate of α_{cg}^{f} may be inaccurate. These combinations account for 2.5% of men and 3.5% of women in the sample, and 2.3% of men and 2.0% of women in the 1931-1984 birth cohorts considered in the decompositions below. To handle these cases, we first estimate a version of equation (1) in which we replace the term $\sum_{g=0}^{\mathcal{G}} \alpha_{cg}^{s} S_{s(i)} C_{c(i)} G_{g(i)t}$ with the additively separable specification $\sum_{c=1}^{\mathcal{C}} \alpha_{c}^{s} S_{s(i)} C_{c(i)} + \sum_{g=1}^{\mathcal{G}} \alpha_{g}^{s} S_{s(i)} G_{g(i)t}$. We use $\hat{\alpha}_{c}^{s} + \hat{\alpha}_{g}^{s}$ as the estimate of α_{cg}^{s} for the c, g combinations with fewer than 30 observations in the decompositions below.

[19] See Altonji et al. (2012), Altonji et al. (2016) and Lovenheim and Smith (2023) for discussions of the methodological challenges and surveys of empirical studies of the return to undergraduate field, some of which report return estimates by gender. Kirkeboen et al. (2016) provide a strategy for estimating returns in settings such as Norway and Chile, where students provide preference rankings of programs and admissions is based on grades and tests. Bleemer and Mehta (2022) use a fuzzy RD design based on a minimum grade requirement at the University of California Santa Cruz to estimate the return to majoring in economics and obtain results similar to OLS. Altonji and Zhong (2021) provide a formal discussion of bias in the use of OLS to estimate the return to graduate degrees.

selection on ability (proxied by test scores) into college, graduate school, and working full time. Each of these datasets administer achievement tests and include details on college and graduate school attendance. Comparing the test score percentiles between male and female college graduates, Appendix Figure L.2 shows that the gap between male and female test scores grew from the PT cohort born in the early 1940s to the NLSY79 cohort born in the early 1960s, before shrinking moderately for the NLSY97 cohort born in the early 1980s. This increase corresponds to a period during which the share of men earning BAs is relatively stable, while the share of women earning BAs is increasing rapidly (See Appendix Figure L.3).[20] We find similar selection patterns on test scores for graduate degree attainment and for working full time (among college graduates). If test scores reflect earnings potential, these differential trends imply that we may be missing compositional changes to our study sample, which would increase the gender gap over time, suggesting the convergence would be smaller if we could correct for selection. Consistent with this, Blau et al. (2024) show that accounting for selection results in larger declines in the gender gap between calendar years 1980 and 2015 using the PSID and the full population of workers. This compositional change would contribute to cohort trends in the residual component of the gap. The fact that the gender gaps in math and verbal scores change in parallel provides a little reassurance that the changes do not have differential effects across majors on the α_{cg}^{sb} estimates. Furthermore, the changes in test score differentials are relatively small.

Lastly, in Appendix L we show that, in the decennial census and ACS, women working full time report working fewer hours per week, ranging from 1.5 to 5.5 hours per week depending on the age and birth cohort. The gaps are the largest around age 35 and are notably smaller for birth cohorts born after 1975. This closing of the hours-gap among full-time workers may account for some of the reduction in the gender gap across birth cohorts that we measure, which is consistent with the results for hourly wage rates discussed in Section 5.1.3 and Appendix K.

4.4 Allowing relative returns to degrees to vary by cohort

In the decomposition above, we assumed that the gender-specific returns to college majors and graduate fields remain constant for birth cohorts ranging from the 1930s to the 1980s.

[20] Appendix Figure L.1 reports trends in SAT math and verbal scores of college bound high school seniors who took the exam, starting with the 1950 birth cohort. It shows a decline in scores for both men and women, but the male-female difference increases by about 10 points in math and about 13 points in verbal (0.1 and 0.13 standard deviations) between the 1950 and 1960 birth cohorts followed by a partial reversal between 1960 and the early 80s cohorts. Goldin et al. (2006) document trends in the gender differences in BA attainment rates from the mid 1870s to 1975. They also explore the role of high school test scores, high school class rank, math and science courses in high school, and family background in trends in the college gap using micro data on individuals born around 1939, 1950, and 1974.

This assumption simplifies the decompositions but is strong. We relax this constraint and write α_{cg}^{sb} in equation (1) as

$$\alpha_{cg}^{sb} = \alpha_{cg}^{s0} + \left(\alpha^{sb} + \delta_{cg}^{sb}\right) \tag{7}$$

where δ_{cg}^{sb} captures gender specific changes in the relative return to cg across cohorts. In the constant returns case, $\delta_{cg}^{sb} = 0$ for all b. Given the other normalizations, the α_{cg}^{s0} are gender and cg specific intercepts for the 1961 birth cohort, and the common cohort component α^{s0} and the δ_{cg}^{s0} are equal to 0. Given sample size limitations, we restrict α^{sb} to equal a cubic polynomial in b, as in the constant returns case. We also restrict δ_{cg}^{sb} to equal the sum of a c specific and a g specific cubic polynomial in b, where we recall that b is normalized to 0 for the 1961 birth cohort. Specifically,

$$\begin{aligned}\alpha^{sb} + \delta_{cg}^{sb} = {} & [(\alpha_1^s + \delta_{c1}^s)b_i + (\alpha_2^s + \delta_{c2}^s)b_i^2 + (\alpha_3^s + \delta_{c3}^s)b_i^3] \\ & + \mathbb{1}\{g > 0\} \times [\delta_{g1}^s b_i + \delta_{g2}^s b_i^2 + \delta_{g3}^s b_i^3]\end{aligned} \tag{8}$$

We substitute the expression $\alpha^{sb} + \delta_{cg}^{sb}$ into equation (1) and estimate by least squares. The regression model identifies $\alpha^{sb} + \delta_{cg}^{sb}$ and α_{cg}^{s0}. We need to impose a normalization to distinguish the α^{sb} (common cohort effects) from δ_{cg}^{sb} (field-specific cohort effects). Because α^{sb} is defined to be a common component that shifts returns in all fields by the same amount, we set α^{sb} to the weighted average of the estimates of $\alpha^{sb} + \delta_{cg}^{sb}$ using the same cg weights for all birth cohorts and both genders. For this reason, we implicitly define α^{sb} to be

$$\alpha^{sb} \equiv \sum_{cg} \left((\alpha^{sb} + \delta_{cg}^{sb}) \times \frac{1}{2} \left(\bar{P}_{cg}^{f} + \bar{P}_{cg}^{m} \right) \right) \tag{9}$$

where $\bar{P}_{cg}^{f} = \frac{1}{54}\sum_b P_{cg}^{fb}$ and $\bar{P}_{cg}^{m} = \frac{1}{54}\sum_b P_{cg}^{mb}$ are the unweighted averages over the 54 birth cohorts of the cg probabilities for women and men respectively. The above equation implicitly defines the $(\bar{P}_{cg}^{f} + \bar{P}_{gc}^{m})/2$ weighted average of δ_{cg}^{sb} to be 0 for each birth cohort and gender. We are assigning any deviation of the weighted average of the δ_{cg}^{sb} from 0 to α^{sb}.

4.4.1 The gender decomposition formula with cohort varying relative returns

The presence of cohort varying relative returns adds two additional terms to the gender gap decomposition formula. Given equation (7), the gap is

$$GAP(b) = \sum_{cg} \left((\alpha_{cg}^{m0} + \delta_{cg}^{mb})P_{cg}^{mb} - (\alpha_{cg}^{f0} + \delta_{cg}^{fb})P_{cg}^{fb} \right) + (\alpha^{mb} - \alpha^{fb}) + \Delta Z_b.$$

Decomposing the double sum, we have

$$
\begin{aligned}
GAP(b) = & \sum_{cg} \left(\alpha_{cg}^{m0} - \alpha_{cg}^{f0}\right) P_{cg}^{fb} && \text{rel. return gap, base year returns} \\
& + \sum_{cg} \alpha_{cg}^{m0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{education gap, base year returns} \\
& + \alpha^{mb} - \alpha^{fb} && \text{cohort } b \text{ residual gap} \\
& + \Delta Z_b && \text{demographic control gap} \\
& + \sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb} && \text{rel. return gap, varying returns} \\
& + \sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{education gap, varying returns}
\end{aligned} \tag{10}
$$

The first four terms correspond to the terms of the decomposition in the constant returns case. Our prior discussion of them applies. The fifth and sixth terms involve the δ parameters and are equal to zero in the constant relative returns case. The fifth term captures the fact that changes across cohorts in gender differences in relative returns, $\delta_{cg}^{mb} - \delta_{cg}^{fb}$, interact with the degree shares P_{cg}^{fb}. To see this, consider the case where the P_{cg}^{fb} are constant across cohorts. Then the fifth term would capture the degree to which movements in $\delta_{cg}^{mb} - \delta_{cg}^{fb}$ tend to be more positive in fields that are common among females. The sixth term captures the degree to which trends in relative returns for males are associated with the gender gap in c, g probabilities. If returns are rising (falling) in male (female) dominated fields, this widens $GAP(b)$ (holding the other terms equal). In the case where relative returns change identically for men and women, with $\delta_{cg}^{mb} = \delta_{cg}^{fb}$ for all b and cg, the sixth term would capture the degree to which cross cohort changes in returns are more positive in fields that are more popular among males than females.[21]

One can measure the contribution of changes in $P_c^{mb} - P_c^{fb}$, $P_{G|c}^{mb} - P_{G|c}^{fb}$ and $P_{g|c,G}^{mb} - P_{g|c,G}^{fb}$ to each of the terms that comprise the education gap using a formula analogous to equation (5) for the education gap in the constant relative returns case, but with 3 additional terms added in which the base year returns α_{cg}^{m0} are replaced with the varying return parameters involving δ_{cg}^{mb}.[22] We discuss the decomposition of the education gap below, focusing on how the decomposition of the α_{cg}^{m0} terms change when we allow for gender differences in the relative returns to vary across birth cohort.

[21] See Kim (2010); Cheng et al. (2019); Kröger and Hartmann (2021) for a discussion of alternative dynamic decompositions that have been proposed.

[22] The formula is provided in Appendix D, equation (12).

4.5 Decompositions of the Occupation Specific Component of Earnings

To estimate the across-occupation regression in equation (3), we first construct the dependent variable $\bar{y}^{ba}_{o(it)}$.[23] One way to do this would be regress Y_{it} on occupation fixed effects, and interactions between age and occupation and birth cohort and occupation. We use this option to supplement our main analysis rather than as our preferred approach. The main reason is that the mapping from college majors and graduate degrees to occupations almost certainly changes across birth cohorts, and data limitations prevent us from controlling for college major and graduate field in the Census/ACS data when estimating the birth cohort specific occupation fixed effects. The changes in the mapping will lead the birth cohort specific component of the occupation premiums to pick up the effects of c and g on earnings within that occupation. Our preferred alternative is to estimate occupation-by-age fixed effects where we parameterize age using a third-order polynomial of age interacted with occupation. This approach imposes the assumption that $\bar{y}^{ba}_{o(it)}$ does not depend on b and so can be written as $\bar{y}^{a}_{o(it)}$. Using these estimates, we then construct $\bar{y}^{a}_{o(it)}$ as our dependent variable based on the individual's occupation and age. See Appendix H.1.

Next, substituting $\bar{y}^{a}_{o(it)}$ for Y_{it} as the dependent variable, we estimate and decompose the gender gap for $\bar{y}^{a}_{o(it)}$ using the same steps as when decomposing earnings.[24] First, we use the Census/ACS data to construct $X^{s}_{2it}\bar{\beta}^{s}_{2}$ using our new dependent variable. Second, we estimate equation (3). For the constant-returns case, we assume $\bar{\alpha}^{sb}_{cg} = \bar{\alpha}^{s0}_{cg} + \bar{\alpha}^{sb}$, while for the dynamic returns case, we assume $\bar{\alpha}^{sb}_{cg} = \bar{\alpha}^{s0}_{cg} + (\bar{\alpha}^{sb} + \bar{\delta}^{sb}_{cg})$.[25] Third, we perform the gender gap decompositions using the same formulas as for the earnings decomposition, but replacing α^{s0}_{cg}, α^{sb}, and δ^{sb}_{cg} with $\bar{\alpha}^{s0}_{cg}$, $\bar{\alpha}^{sb}$, and $\bar{\delta}^{sb}_{cg}$.[26]

For both the earnings and occupation decompositions, we calculate standard errors using 200 bootstrapped samples stratified by birth cohort. Since we have panel data, we sample individuals with replacement, keeping all observations for each sampled person. We do not bootstrap the Census/ACS data because sampling error from these datasets is likely to be second order given their large size. Thus, the occupational premiums and the age-birth cohort regression index for the earnings and occupation regressions remains unchanged across

[23]See Appendix H.1 for technical details.

[24]We rescale the estimates of age specific occupational premiums $\bar{y}^{a}_{o(it)}$ using the coefficient estimate from a simple regression of $ln\,(earnings)$ on $\bar{y}_o$ (the occupation dummies constructed in the Census/ACS) and a female indicator. This addresses differences in earnings measures and discrepancies across the occupation measures in the Census/ACS and the NSCG. This rescaling factor is 0.816.

[25]When estimating equation (3), we impose the same polynomial restrictions and normalizations on $\bar{\alpha}^{s0}_{cg}$, $\bar{\alpha}^{sb}$, and $\bar{\delta}^{sb}_{cg}$ that we impose on α^{s0}_{cg}, α^{sb}, and δ^{sb}_{cg} in the earnings case. See Section 4.4.

[26]For the constant returns case we use the modified versions of equations (4) and (5), and for the dynamic returns case we use the modified versions of equations (10) and (12).

bootstrap samples.

5 Decompositions of the Gender Gap: Results

We now present the results of the empirical decompositions of the long-term trends in the gender gap. The first two subsections consider the earnings gap. The second two subsections consider the gap in the occupation premium, and how much of the overall earnings gap this can explain. For both earnings and the occupation premium, we begin with the assumption that relative returns to degrees are constant across cohorts, which we then relax.

5.1 Decompositions of the Gender Gap in Earnings Using Constant Relative Returns

Figure 2 panel A plots the terms of the decomposition in equation (4), where we assume constant relative returns.[27] The solid black line is $GAP(b)$, the total gap. The gap starts at 0.645 in 1931 and then declines rapidly and almost linearly until 1944, when the gap is 0.411. The decline is 0.017 per year. The rate of decline slowly falls until about 1953, when the gap reaches 0.333. After that the decline is much more gradual, averaging only 0.004 per year until the 1978 birth cohort, when the gap is 0.273. Between 1978 and 1984 the gap declines by 0.006 per year, and ends at 0.235 for the 1984 birth cohort.

It is important to keep in mind that $GAP(b)$ is our estimate of the gap in average earnings between the ages of 28 and 52 by birth cohort. As we do not observe this average for every birth cohort in our data, the gap is constructed from our regression model and estimates of the postsecondary education outcome probabilities.[28]

Next we turn to the relative return gap (Figure 2 panel A, orange line). The formula is $\sum_{cg}\left(\alpha^{m0}_{cg}-\alpha^{f0}_{cg}\right)P^{fb}_{cg}$, where$\left(\alpha^{m0}_{cg}-\alpha^{f0}_{cg}\right)$ includes the 1961 value of the gender gap com-

[27] Table J.1 in Appendix J reports the point estimates and standard errors that underlie Panel A of Figure 2 for each value of b. Tables J.2, J.3, and J.4 in Appendix J report point estimates and standard errors for Figure 6 panel A, Figure 4 panels A, B, and C, and Figure 8 panels A, B, and C.

[28] In Appendix F, we report estimates of the earnings gap for the college plus population based on the Census/ACS for 1960 through 2019. These are also based on the regression discussed in section 4.1, which permits us to perform an age adjustment, so that the Census/ACS based estimates of $GAP(b)$ also correspond to career earnings of full time workers between age 28 and 52. The critical difference is that we do not know college major or graduate field, and so the education measure is limited to graduate attendance. However, the analysis serves as a check on whether the fact that the NSCG earnings data starts in 1990 distorts results for the early birth cohorts despite our use of the index $X^s_{2it}\hat{\beta}^{s*}_2$ as a control. Fortunately, Census/ACS based estimate of the total earnings gap starts at 0.65 for the 1931 birth cohort and is very close to the NSCG estimates displayed in Figure 2 until the late 50s cohorts. It declines very little after that and is about 0.06 above the NSCG estimates for the early 80s cohorts. The return gap is larger and the education gap is much smaller and flatter, as one would expect. The residual gap follows the same path as the path based on the NSCG, but is a bit larger in the early years.

ponent that affects all cg categories equally. The return gap starts at 0.204 for the 1931 cohort, peaks at about 0.211 for 1959, and then slowly declines to 0.193 for the 1984 cohort. Since the relative returns are constant, the small amount of variation in this term is driven entirely by changes in the mix of degrees that women choose. Although the major shares for women shift substantially across cohorts, the graph indicates that those shifts are not large enough or sufficiently correlated with gender differences in relative returns $\left(\alpha_{cg}^{m0} - \alpha_{cg}^{f0}\right)$ to induce significant shifts in the relative returns gap.[29] We obtain the same result when we weight by the male shares.

In contrast, the cohort residual gap $\alpha^{mb} - \alpha^{fb}$ (gray line, normalized to 0 when $b = 1961$) changes dramatically over the early cohorts. Recall that this gap is a change in the gender gap across cohorts that is shared across all undergraduate and graduate degrees. The residual gap starts at 0.253 for 1931, which is 39% of the total gap in that cohort. It drops very rapidly across the 1930s cohorts. After the 1930s, it continues to decline, but at a decreasing rate until it reaches 0 for the 1960 cohort. It rises to 0.014 in 1975 and falls to -0.01 in 1984. The main takeaway is that the unexplained component of the gender gap dropped dramatically across birth cohorts until the late 40s cohorts, and very little after that. Evidence from Goldin (2014, 2006, 2021), Lemieux (2006), Ruggles (2015), Blau and Kahn (2017), and many other suggest several trends that may help explain the convergence in the residual gap we estimate, such as greater cumulative work experience at each age, longer weekly hours, declining fertility, shifting gender norms and occupational preferences, and reduced discrimination.

We now turn to the education gap, $\sum_{cg} \alpha_{cg}^{m0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$, which is the green line. It is very large for the early birth cohorts, starting at 0.191 in 1931 and peaking at 0.206 in 1935. The latter value is only 0.029 less than the *total gap* for the 1984 birth cohort. The education gap then begins to fall after the 1936 cohort. The decline is particularly rapid between the 1940 and 1952 birth cohorts, averaging 0.005 per year. The gap reaches 0.124 for the 1952 cohort. As we discussed earlier, during this period women moved out of education, the humanities, and the arts toward business and other higher paying fields. The education gap continues to decline between 1952 and 1972 cohorts, but only by an average of 0.003 per year. The education gap actually increases slightly between 1972 and 1977, and then declines at an accelerating rate for the most recent cohorts. It ends at 0.067 for the 1984 cohort (27.6% of the total gap). Overall, the residual gap and education gap both contribute to the rapid decline in the total earnings gap from 1931 through the mid 1950s, while most

[29] Appendix Figure I.1 disaggregates the return gap in Figure 2 into the contributions $\sum_g \left(\alpha_{cg}^{m0} - \alpha_{cg}^{f0}\right) P_{cg}^{fb}$ for each major c. The figure shows that for a few majors the increases or decreases in the return gap across cohorts are substantial even though the sum is relatively constant.

of the decline since 1955 comes from declines in the education gap.

In Figure 2 panel B, we use equation (5) to decompose the education gap into the contributions from college major, graduate school attendance given college major, and graduate field given graduate school attendance and college major. The BA field gap (pink dotted line) starts at 0.143, peaks at 0.146 for the 1933 cohort, and then declines until 1969. The steepest decline is between 1940 and 1949. After the 1969 birth cohort the BA field gap increases slowly until 1977, when it starts to decline. The lack of monotonicity is consistent with findings in Altonji et al. (2012) and Sloane et al. (2021), though these papers do not include graduate education and do not consider cohorts before 1950, when the largest changes were occurring.

The graduate attendance gap (purple line) peaks at 0.030 in 1932 and then declines slowly across birth cohorts, turning negative in 1968. The decline accelerates after 1980. Between 1931 and 1984, the graduate attendance gap accounts for 40% (0.049) of the decline in the overall education gap, which compares to 54% (0.067) for the BA field gap and 6% (0.007) for the Grad field gap (blue line).

Overall, the steep decline in the education gap in the earlier birth cohorts is driven by changes in BA field, while the continued decline since the 1950s is driven by a combination of changes in college majors and graduate attendance, with graduate field playing a significant role from the early 1940s to the late 1950s.

5.1.1 Contributions of specific BA fields to the trend in the Education Gap

In Figure 3 we disaggregate the education gap in Figure 2 by BA field. For each BA field, we aggregate the contribution to the gap across all graduate fields, including no graduate degree ($g = 0$). The sum of the major specific curves by birth year for the education gap equals the green line for the education gap shown in Figure 2.

The field specific results show that the decline in the education gap between the 1930s and 1960s cohorts is largely caused by the decline in the gender difference in the probability of majoring in Education, English/Languages/Literature, and Other Humanities, with Business and Fine Arts also playing a role. The flattening of the education gap curve from the late 60s until the late 70s reflects a slower decline in the contribution of Education, and Other Humanities that is offset by small increases in the contribution of Computer Science/Math, Engineering, and Psychology. The decline in the education gap after the late 1970s is driven by several fields, with Business, Biology, Nursing, and Health contributing to the decline. Engineering and Computer Science/Math work in the opposite direction.

5.1.2 Robustness Checks

We obtain very similar results when we perform the decompositions using the female education coefficients and male probabilities (see Appendix E). In Appendix G, we explore the sensitivity of the earnings gap decompositions to alternative measures of the undergraduate degree probabilities P_c^{fb} and P_c^{mb}. First, we replace estimates of P_c^{fb} and P_c^{mb} based on the NSCG data with estimates based on the HEGIS/IPEDS data discussed in Section 2 and in more detail in Appendix C. The analysis is restricted to cohorts after 1944 because HEGIS starts in 1966. The results in Appendix Figure G.1 are very similar to the NSCG based estimates even though the gender and cohort specific HEGIS/IPEDS and the NSCG based estimates differ substantially in some cases. Second, we obtain similar results when we use simple 3 year moving averages to estimate P_c^{fb} and P_c^{mb} rather than using b-splines (see Appendix Figure G.3).[30]

5.1.3 Hourly Wage Rates

To isolate the role that gender differences in annual hours play in the gender gap among full time workers, in Appendix K we repeat the analysis replacing log annual earnings with log average hourly earnings. Comparing the log wage rate decomposition to the log annual earnings decomposition, the hourly wage gap is 0.157 smaller in 1931 and 0.023 smaller in 1984. However, the make up of the decreasing trends in the gaps are similar. For log annual earnings, the total gap, relative return gap, and education gap decline by 63%, 9%, and 67% respectively from 1931; for log hourly wage the corresponding declines are 57%, 8%, and 64%. The main difference is that the decline in the cohort residual gap between 1931 and 1984 in the log wage case is about 0.11 smaller than the decline in the log annual earnings case. Most of the difference is between 1931 and 1961 (the year in which the cohort residual gap is normalized to zero in all of our analyses).

5.2 Decompositions of the Gender Gap in Earnings Using Cohort Specific Relative Returns

In this subsection, we extend our prior decomposition to allow the relative returns to vary across birth cohort. This is captured by the additional term δ_{cg}^{sb}, as described in equation (10). Figure 4 panel A shows the result of this decomposition. The black line is the estimate of the gender gap in log earnings across birth cohorts. When allowing for cohort specific

[30] The decomposition of the gender gap in the occupation component of earnings that we discuss in section 5.3 are also robust to using the HEGIS/IPEDS data and to the use of simple 3 year moving averages to estimate P_c^{fb} and P_c^{mb} rather than using b-splines. The results are in Figure G.2 and Figure G.4.

relative returns, we find a larger gender gap in the early cohorts and a smaller gap in the later cohorts. In the 1931 birth cohort, the total gender gap is 0.68. The total gap decreases steadily through the 1940s birth cohorts and then continues to decrease at a slower rate, reaching 0.238 by 1984. Most of the decline is accounted for by changes in the birth cohort residual gap ($\alpha^{mb} - \alpha^{fb}$) in favor of women. Recall that the residual gap is common to all education choices and is normalized to 0 in 1961. It starts at 0.336 in 1931 and decreases sharply from the 1930s to the late 1940s birth cohorts and then more slowly from 1950 to 0 in 1961. It remains near 0 until 1978 and then declines to -0.021 in 1984.

The contribution of the relative returns to cg to the gender gap is close to 0.2 for all cohorts (orange dashed line). The education gap (green dashed line) drops from 0.144 in 1931 to 0.089 in 1984. It accounts for 36% of the gender gap on average across cohorts and 37.4% of the total gap for the 1984 birth cohort. But it contributes only 0.055 to the decline in the gender gap, which is less than half of the 0.128 drop in the constant returns case.[31]

5.2.1 The Effect of Cohort Specific Relative Returns on the Education Gap

When we allow the relative returns to vary by birth cohort, we surprisingly find that the education gap declines by only 0.055. In panel C of Figure 4, we decompose the education gap based on the cohort-specific relative return specification (green line) into its two parts. The first is $\sum_{cg} \alpha_{cg}^{m0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$, which evaluates the cg specific education gaps using the base year return to cg, α_{cg}^{m0} (light green long dashed line). The second part is $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$. This term evaluates the cg specific education gaps using δ_{cg}^{mb}, the male specific changes in the relative return to cg across cohorts (yellow line).[32] The relatively flat education gap (green line) is the sum of a decreasing base year return education gap and an increasing varying return education gap. The base year return component of the education gap decreases from 0.195 in 1931 to 0.061 in 1984. The curve is similar to the curve for the education gap based on the constant returns specification of the earnings model (Figure 2 panel A). The decline means that the male-female difference in choice of cg, $P_{cg}^{mb} - P_{cg}^{fb}$, tends to be falling across cohorts in fields that were "high-paying" in 1961 (high values of α_{cg}^{m0}) and rising in lower-paying fields. The large shifts of women into Business and out of Education, English,

[31] Appendix Figure I.5 disaggregates the return gap in Figure 4 panel A into the contributions $\sum_g \left(\alpha_{cg}^{mb} - \alpha_{cg}^{fb}\right) P_{cg}^{fb}$ for each major c. The figure also disaggregates the education gap by major.

[32] In panel C of Figure 5, we graph $\sum_{cg} \delta_{cg}^{mb} P_{cg}^{mb}$, $\sum_{cg} \delta_{cg}^{fb} P_{cg}^{fb}$, and the difference between them. The difference is the sum of $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ and $\sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb}$. This graph shows the combined effects of the interaction between changes in relative returns and field choice on the earnings paths of men and women. One can see that $\sum_{cg} \delta_{cg}^{mb} P_{cg}^{mb}$ slowly rises from about -0.018 to 0.017 over the whole period, while $\sum_{cg} \delta_{cg}^{fb} P_{cg}^{fb}$, rises initially, falls from about 0.038 (relative to 1961) to 0 in 1961 and is near zero after that. The net result is to increase the gender gap in earnings by about 0.06 between the 1930s and 1984.

and the Humanities that we documented earlier are part of the story.

On the other hand, the rise in $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ across cohorts indicates that returns to degree types that are dominated by women declined across cohorts. The two effects partially cancel out and generate the relatively modest decline in the cohort-specific education gap.

Using the estimates of the cohort specific component of the occupation premiums based on the Census/ACS data, we have confirmed that relative earnings fell in occupations typically associated with a degree in Education (such as teachers) and rose in occupations associated with Engineering degrees (such as engineers), and that changes across cohorts in the field specific returns favored men relative to women (not shown).

To shed further light on the upward trend in the varying relative return component of the education gap, $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$, we rewrite it as

$$\begin{aligned}
&\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{Edu gap, varying return} \\
=&\sum_{cg} \delta_{cg}^{mb} \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right) && \text{Edu gap, constant probability gap} \\
+&\sum_{cg} \delta_{cg}^{mb} \left(\left(P_{cg}^{mb} - P_{cg}^{fb}\right) - \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)\right) && \text{interaction term,}
\end{aligned} \tag{11}$$

where we recall that $\bar{P}_{cg}^{m}$ and $\bar{P}_{cg}^{f}$ are the cohort invariant gender specific average probability of obtaining cg, as defined below equation (9). Figure 5 panel A graphs the three terms in (11). The solid yellow line is the graph of $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ from Figure 4 panel C. As we have already noted, it increases from about -0.052 to 0.025, favoring men. The graph of $\sum_{cg} \delta_{cg}^{mb} \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)$ (dark blue dashed line) shows that about 0.03 of that increase is because relative returns (measured by the returns for males) rose for degrees in fields that are more popular among men. This increase is reinforced by a positive interaction between cohort trends in the relative returns and trends in gender differences in education choice $\left(\left(P_{cg}^{mb} - P_{cg}^{fb}\right) - \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)\right)$ (turquoise dashed line). Relative to men, women increasingly moved away from majors with rising returns.[33]

[33] Appendix Figure I.10 graphs the contribution of each major to the varying return component of the education gap, $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$. For each major, it reports $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$, $\sum_{g} \delta_{cg}^{mb} \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)$ and the interaction term $\sum_{cg} \delta_{cg}^{mb} \left(\left(P_{cg}^{mb} - P_{cg}^{fb}\right) - \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)\right)$. The trends in $\sum_{g} \delta_{cg}^{mb} \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)$ worked in favor of men in education, engineering, and to a lesser extent nursing and fine arts, with smaller changes in both directions in other fields. The trend in nursing favors men until the early 1950s and then partially reverses.

5.2.2 Decomposing the Return Gap

In Figure 4 panel B, we decompose the return gap (orange line). From equation (10), the return gap is the sum of two terms, a base year return gap and a varying return gap. The base year return gap (light green long dashed line) represents the constant component of the cohort-specific returns across birth cohorts. Its stable magnitude aligns with our findings from the constant returns specification. It shows that the value of the average return difference across cg is about 0.2 and that the changes in cg choices of women over birth cohorts have only a minor impact on their average return relative to men. Interestingly, the gender gap in the varying returns to cg, $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)P_{cg}^{fb}$ (yellow dotted line) is close to zero for all cohorts. This holds true even though the values of $\delta_{cg}^{mb}-\delta_{cg}^{fb}$ are nontrivial for some fields (not shown). This implies that the weighted average across cg of the gender gap in δ_{cg}^{sb} is negligible.

In Figure 5 panel B, we decompose the varying return component (solid yellow line) into its two subcomponents. The blue dashed line is the graph of $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)\bar{P}_{cg}^{f}$, which weights changes in $(\delta_{cg}^{mb}-\delta_{cg}^{fb})$ by $\bar{P}_{cg}^{f}$, the average probability of cg for women. It rises slowly from about -0.025 to 0 in 1960 and varies little after that. Changes in relative returns benefited men relative to women. However, the rise in $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)\bar{P}_{cg}^{f}$ is more than offset by a decline in the interaction term $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)\left(P_{cg}^{fb}-\bar{P}_{cg}^{f}\right)$, the turquoise dashed line. Across the early cohorts, women moved toward majors in which $(\delta_{cg}^{mb}-\delta_{cg}^{fb})$ was falling, leading to a decline in the interaction term of about 0.035 between the 1931 and 1946 cohorts. After that, the term slowly rises to zero.[34]

5.3 Decompositions of the Gender Gap in the Occupation Premium Using Constant Relative Returns

So far we have decomposed the gender gap in log earnings. In Section 4.1, we expressed log earnings as the sum of a within-occupation component and an occupation component, which we also refer to as occupation premium. Here we study the role of occupation choice in shaping the gender gap by decomposing the occupation component of log earnings. Since gender differences in earnings within an occupation are part of the within-occupation component, all differences discussed here capture gender differences related to occupation choice.

Figure 6 panel A shows the decomposition of gender differences in the occupation pre-

[34] Appendix Figure I.9 decomposes the line by reporting $\sum_{g}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)\bar{P}_{cg}^{f}$ for each major. One can see that a trend against women in $\delta_{cg}^{mb}-\delta_{cg}^{fb}$ for Education, Biology, and Other Social Sciences contribute to the rise $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)\bar{P}_{cg}^{f}$. Appendix Figure I.9 also graphs the major specific contributions to $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)\left(P_{cg}^{fb}-\bar{P}_{cg}^{f}\right)$.

mium, $\bar{y}^a_{o(it)}$. Across birth cohorts, the total gap (black line) drops from 0.202 to 0.098, while its share of the overall gap in earnings increases from 31% to 42%. From 1931 to 1963, the decline is almost linear, with a slope of -0.003 per year. From 1963 to 1984, the gap stabilizes at around 0.108.

The cohort residual gap (gray long dashed line) starts at 0.035 in 1931, declines to zero by 1955, and remains at zero through the 1984 birth cohort. Keep in mind that the cohort residual gap is normalized to be zero in 1961.

The relative return gap in the occupation premium (orange dashed line) is the average of the cross-occupation returns for men and women, weighted by P^{fb}_{cg}, the birth cohort specific cg probabilities for women. The gap is around 0.060. Since the occupation premiums we use are the same for men and women, the 0.060 gap arises because, for a given choice of c and g, women end up in lower paying occupations than men. Across cohorts, changes in the female cg probabilities have little overall impact on this gap, even though we find modest variation across cohorts for a few specific majors (see Appendix Figure I.3).

The education gap (green long and short dashed line) drives most of the changes in the occupation premium across birth cohorts. It starts at 0.104, rises slightly to 0.107 in 1936, and then steadily declines to 0.053 by 1963. From 1963 to 1978, it is relatively stable, before falling further to 0.043 by 1984.

Figure 6 panel B decomposes the education gap into gender differences in college major, graduate school attendance given college major, and graduate field of study given graduate school attendance and college major.[35] On average across cohorts, undergraduate field (magenta dotted line) accounts for 81% of the level of the education gap in the occupation premium and for most of its decline. Gender differences in graduate attendance conditional on major (purple dashed line) play a secondary role, hovering around 0 for most of the period. The gap due to gender differences in graduate field choice conditional on college major choice (light blue line) starts at 0.015 in 1931 and ends at 0.013 in 1984. This component grows to 0.029 in 1940, followed by a slow decline to 0.008 in 1964 and a slight increase afterward. Overall, the decomposition in Figure 6 panel B shows that differences in college major choices and graduate field conditional on college major give men better access to high-paying occupations, with the narrowing of the gender gap in college major choice contributing the most to long term trends.

Appendix Figure I.3 reports the contribution of specific college majors (and associated graduate degrees) to the education gap. This figure is the occupation counterpart to Figure 3 for earnings. Business and Education contribute the most to the decline in the college major

[35]It is based on equation (5) but substitutes the occupation premium coefficients $\bar{\alpha}^{m0}_{cg}$ in place of the earnings coefficients α^{m0}_{cg}.

gap, and English/Languages/Literature, Fine Arts, and health related fields also contribute. Business contributes 0.015 in 1931 and -0.001 in 1984. Education contributes 0.03 in 1931 and 0.01 in 1984. Psychology and Social Work and Computer Science/Math contribute a partially offsetting increase in the education gap after 1960.

5.3.1 Within and Across Occupation Sources of the Earnings gap

In Figure 7 panel A, we use the earnings and occupation decompositions to separate the gender gap in earnings into the portion that is within occupations and the portion that is across occupations. The estimates are based on the constant returns specification. The thickness of the color bands shows the size of the within and across occupations portions of the education gap, the relative return gap, the cohort residual gap, and the demographic gap. The trend in the occupation premium gap (magenta dashed line) is much flatter than the trend for earnings (black solid line). On average across cohorts, the education gap is about 59% across occupation and about 41% within occupation. In contrast, 71% of the relative return gap is within occupation. In 1931, 86% of the cohort residual gap is within occupation. By 1955, both the within and across occupation components of the residual gap are near zero (relative to 1961).[36]

Panel B performs a similar exercise for the education gap, decomposing the three parts of the education gap into "across" and "within" occupation components. The contribution of undergraduate field to the gender gap comes largely through occupation. The same is largely true for graduate field. In contrast, almost all of the effect of graduate attendance is within occupation. This finding suggests that getting a graduate degree increases the level of the job more than the nature of the job. Note that the graduate attendance effect is positive in early years but turns negative after the 1969 birth cohort, reflecting the growth in female graduate degree attainment.

5.4 Decomposing the Gender Gap in the Occupation Premium Using Cohort Specific Relative Returns

Figure 8 panel A presents the decomposition of the occupation premium $\bar{y}^a_{o(it)}$ when allowing for cohort-specific relative returns. The occupation premium gap (black line) starts at 0.206 in 1931 and declines to 0.102 in 1984. The occupation premium gap accounts for 30.2% of the overall gender gap in log earnings in the 1931 cohort and 41.0% in the 1980s cohorts. These estimates are similar to the estimates based on the constant returns specification.

[36] After 1961, the within occupation cohort residual gap has a small temporary increase up to 0.015 through the 1960s to the mid 1970s cohorts, but returns to zero by 1984.

The paths of the cohort residual gap in the occupational premium (relative to 1961) (gray dotted line) and the education gap (green dashed line) are also similar to the results for the constant returns specification reported in Panel A of Figure 6. The decline in the education gap contributes 0.087 to the decline in the occupation premium gap.[37]

The cohort-specific return gap (orange line) in the occupation premium starts at 0.035 in 1931 and rises to 0.059 by 1984. Panel B shows that the return gap evaluated at the base year difference in returns $\bar{\alpha}_{cg}^{m0} - \bar{\alpha}_{cg}^{f0}$ averages 0.059 and declines slightly across cohorts (light green dashed line). This is only 0.001 smaller than the constant relative return gap from Section 5.3.

The increase in the cohort-specific return gap is primarily driven by the cohort varying return component $\sum_{cg} \left(\bar{\delta}_{cg}^{mb} - \bar{\delta}_{cg}^{fb}\right) P_{cg}^{fb}$ (Panel B of Figure 8, yellow line). The line is upward-sloping from 1931 to 1953 and constant after that. Notably, it starts with negative values and then increases to zero. Since $\bar{y}_{o(it)}^{a}$ does not depend on gender or birth cohort, the term $\sum_{cg} \left(\bar{\delta}_{cg}^{mb} - \bar{\delta}_{cg}^{fb}\right) P_{cg}^{fb}$ captures cross cohort trends in gender differences in the relationship between occupation and cg. The upward slope means that although the cohort-specific return gap initially favors women (i.e., women in the early cohorts chose higher-paying occupations than men with the same education fields, relative to the 1961 birth cohort), this advantage diminishes over time and eventually vanishes, increasing from -0.026 to -0.002.

As a robustness check, we also perform decompositions using cohort and age specific occupation premiums, $\bar{y}_{o(it)}^{ba}$. This check also allows us to examine whether changes across cohorts in occupation premiums also contribute to the trend in the gender gap in the occupational component of earnings. When estimating these premiums, we restrict the cohort specific variation in the occupation premiums around the value in 1961 to be additively separable from the occupation specific age profiles, as discussed in Appendix H.1. With this restriction, $\bar{y}_{o(it)}^{ba}$ can be written as

$$\bar{y}_{o}^{ba} = \bar{y}_{o}^{0a} + \bar{y}_{o}^{b},$$

where $\bar{y}_{o}^{0a}$ is the age specific value of the occupation premium in the base year (1961), $\bar{y}_{o}^{b}$ is the cohort specific occupation component, and we have suppressed the i and t subscripts.

Appendix Figure H.1 presents decompositions using $\bar{y}_{o}^{ba}$ as the dependent variable. The total gap, education gap, and residual gaps are slightly large prior to the late 50s birth cohorts than the estimates in Figure 8 based on $\bar{y}_{o}^{a}$, but the trends are very similar, which is reassuring.

We now address the question of how much of the upward trend in the relative return

[37] The decomposition of the education gap into college major, graduate school attendance, and graduate field in Figure 8 panel D is also similar to decomposition in the constant returns case.

gap is driven by (1) changes in the mix of majors, (2) changes in the mapping from cg to o evaluated at the base year returns for each age, and (3) the change in the occupation premiums across birth cohorts.

Changes in the mix of majors evaluated at the gender difference in base year returns $\bar{\alpha}_{cg}^{m0} - \bar{\alpha}_{cg}^{f0}$ contribute very little to the trend in the return gap (Appendix Figure H.1, dashed light green line). To isolate the second factor, we decompose $\bar{y}_o^{0a}$ (in place of $\bar{y}_o^{ba}$). To isolate the third factor, we decompose $\bar{y}_o^{b}$, which captures the change in the payoff to occupation o. Appendix Figure H.2 compares the return gap decompositions for the entire occupation component $\bar{y}_o^{ba}$ (panel A), for $\bar{y}_o^{0a}$ (panel B), and for $\bar{y}_o^{b}$ (panel C). By construction, the sum of the lines in panels B and C equals the corresponding lines in panel A. Of the 0.029 increase in the varying return gap between 1931 and 1960 (panel A, yellow-short dashed line), 0.023 is caused by changes in the mapping from cg to occupations that on balance favor men (yellow short dashed line, Panel B). Only 0.006 of the increase is caused by changes in the occupation premiums that favor men (yellow short dashed Panel C). This further decomposition reveals that while most components in the gender pay gap decrease over birth cohorts and favor women, changes across cohorts in the mapping from cg to occupation choices have favored men by sorting them into better paying occupations. Without this component, which is -0.029 in 1931, the occupation gap would have been 18% higher, and the overall earnings gap would have been 4% higher.

6 Conclusion

We study the decline in the gender gap among full-time college-educated individuals born between 1931 and 1984, focusing on the role of college major choice, graduate degree attainment and field, and field-specific returns in explaining the gender gap. Recent papers such as Altonji et al. (2012) and Sloane et al. (2021) have shown that, for cohorts since 1950, gender differences in college major choice contribute substantially to the gender gap, and that differential trends in major choice lead to some narrowing of the gap. We go back much further in time and incorporate graduate education into the analysis. By going back to the early 1930s birth cohorts, we contribute new facts about the contribution of type of higher education to the large reduction in the gender gap that occurred in earlier cohorts. By incorporating graduate education, we can assess the importance of the 21 percentage point change in the gender gap in graduate degree attainment across birth cohorts. We can also assess how shifts in men's and women's graduate field choices have influenced the gender gap. The extension back in time is not straightforward because the earliest wave of the National Survey of College Graduates only includes earnings data back to the 1990 calendar year.

We address this limitation by supplementing the NSCG data with information about gender specific age and cohort effects on earnings based on the 1960-2000 Census and 2001-2018 ACS.

In brief, we find that much of the large gap in earnings between the 1931 and 1950 cohorts is due to a cohort specific "residual component" that shifts the gender gap in earnings by the same amount for all college graduates. Most of the decline is within occupation, especially for the early cohorts. The residual gap varies little for the 1951 to late 1970s birth cohorts, after which it resumes its decline. The factors behind the residual component are not the subject of our paper, but the extensive literature on the long term trend in the gender gap suggests changes in total labor experience at a given age, hours per week, lower fertility, shifting gender norms and preferences affecting occupation choice, and reduced discrimination against women have all played a role.

Second, we find that gender differences in the relative return to undergraduate and graduate degree combinations contribute to the gender gap, but very little to the decline in the gender gap over the full time period.

Third, we study and further decompose the education gap into the contributions from college major choice, graduate degree attainment, and graduate field. When evaluated at fixed relative returns to each degree type, we find that the education gap is large for the early cohorts but declines substantially and is an important part of the narrowing of the gender gap. However, to our surprise, this decline is mostly offset by cohort trends in the relative returns to specific fields that favored men. Overall, the education gap accounts for only 0.055 of the decline, which is very little of the total decline in the gender gap in log earnings.

How will the choices of college major and graduate field affect the gender gap in career earnings of more recent cohorts? Only time will tell, but we conducted a simulation in which we set the marginal probabilities P_c^{sb} of choosing each major to the average of the 1991, 1992, and 1993 cohorts while setting conditional probabilities of graduate school choices and the earnings model variables to the values for the 1984 cohort. The results show a small *widening* of the education gap by 0.012.[38] It remains to be seen whether changes in graduate education conditional on college major or in relative returns will amplify or offset this change.

We conclude with several caveats. First, we decompose the cohort specific gender gap for men and women who work full time. This is a well-defined question, but it would also be interesting to know how changes across cohorts in (1) who earns college degrees and (2) who works full time conditional on having a college degree contribute to changes in the earnings gap among college educated workers. In Appendix L we provide evidence based on test scores

[38] See Appendix Table J.7.

that differential trends in selection may have worked against the narrowing of the gender gap, although the changes are only modest. One could supplement our research by drawing on analyses from other papers that have studied changes in selection into higher education and into employment. Second, we rely on OLS to estimate the returns to undergraduate and graduate degrees. Our results for the specification that assumes constant returns across birth cohorts are robust to using the method developed in Altonji and Zhong (2021) that allows us to treat graduate degree choices as endogenous, but we do not have an alternative to OLS for college major. It is also possible that selection into the degrees has changed across cohorts in ways that alter the estimates of relative returns. Finally, more research is needed on the factors contributing to the large changes in education choices of both men and women that we document.

References

Abramitzky, R., V. Lavy, and M. Segev (2024). The effect of changes in the skill premium on college degree attainment and the choice of major. *Journal of Labor Economics 42*(1).

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016, January). Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 5, pp. 305–396. Elsevier.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics 4*(1), 185–223.

Altonji, J. G., J. E. Humphries, and L. Zhong (2023). The effects of advanced degrees on the wage rates, hours, earnings, and job satisfaction of women and men. *Research in Labor Economics 50*(1), 25–81.

Altonji, J. G. and L. Zhong (2021, April). The Labor Market Returns to Advanced Degrees. *Journal of Labor Economics 39*(2), 303–360. Publisher: The University of Chicago Press.

Black, D. A., A. Haviland, S. G. Sanders, and L. J. Taylor (2008). Gender Wage Disparities among the Highly Educated. *The Journal of human resources 43*(3), 630–659.

Blau, F. D. and L. M. Kahn (2017, September). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature 55*(3), 789–865.

Blau, F. D., L. M. Kahn, N. Boboshko, and M. Comey (2024). The impact of selection into the labor force on the gender wage gap. *Journal of Labor Economics 42*(4), 1093–1133.

Bleemer, Z. and A. Mehta (2022). Will studying economics make you rich? a regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics 14*(2), 1–22.

Blom, E., B. C. Cadena, and B. J. Keys (2021). Investment over the business cycle: Insights from college major choice. *Journal of Labor Economics 39*(4), 1043–1082.

Bronson, M. A. (2019). Degrees Are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women. *Working Paper*, 55.

Brown, C. and M. Corcoran (1997). Sex-based differences in school content and the male-female wage gap. *Journal of Labor Economics 15*(3), 431–465.

Cheng, S., C. R. Tamborini, C. Kim, and A. Sakamoto (2019). Educational variations in cohort trends in the black-white earnings gap among men: Evidence from administrative earnings data. *Demography 56*, 2253–2277.

Frazis, H., M. H. Ports, and J. Stewart (1995). Comparing measures of educational attainment in the cps. *Monthly Lab. Rev. 118*, 40.

Gemici, A. and M. Wiswall (2014). Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors. *International Economic Review 55*(1), 23–56.

Gicheva, D. (2013, October). Working Long Hours and Early Career Outcomes in the High-End Labor Market. *Journal of Labor Economics 31*(4), 785–824. Publisher: The University of Chicago Press.

Goldin, C. (2006, May). The Quiet Revolution That Transformed Women's Employment, Education, and Family. *American Economic Review 96*(2), 1–21.

Goldin, C. (2014, April). A Grand Gender Convergence: Its Last Chapter. *American Economic Review 104*(4), 1091–1119.

Goldin, C. (2021). *Career and family: Women's century-long journey toward equity*. Princeton University Press.

Goldin, C., L. F. Katz, and I. Kuziemko (2006). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic perspectives 20*(4), 133–156.

Imberman, S. A., M. F. Lovenheim, P. Massey, K. Stange, and R. J. Andrews (2024, September). Gender and racial earnings differences: Major choices, returns, and welfare consequences. Working paper.

Jaeger, D. A. (1997). Reconciling educational attainment questions in the cps and the census. *Monthly Lab. Rev. 120*, 36.

Kim, C. (2010). Decomposing the change in the wage gap between white and black men over time, 1980-2005: An extension of the Blinder-Oaxaca decomposition method. *Sociological Methods & Research 38*(4), 619–651.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The quarterly journal of economics 131*(3), 1057–1111.

Kominski, R. and P. M. Siegel (1993). Measuring education in the current population survey. *Monthly Lab. Rev. 116*, 34.

Kröger, H. and J. Hartmann (2021). Extending the Kitagawa–Oaxaca–Blinder decomposition approach to panel data. *The Stata Journal 21*(2), 360–410.

Lemieux, T. (2006). Postsecondary education and increasing wage inequality. *American Economic Review 96*(2), 195–199.

Lovenheim, M. and J. Smith (2023). Returns to different postsecondary investments: Institution type, academic programs, and credentials. In *Handbook of the Economics of Education*, Volume 6, pp. 187–318. Elsevier.

Park, J. H. (1996). Measuring education over time: A comparison of old and new measures of education from the current population survey. *Economics Letters 50*(3), 425–428.

Patnaik, A. (2020). Pricing, Income and College Major Choice. *Working paper*, 85.

Patnaik, A., M. J. Wiswall, and B. Zafar (2020, August). College majors. Working Paper 27645, National Bureau of Economic Research.

Ruggles, S. (2015). Patriarchy, power, and pay: The transformation of american families, 1800–2015. *Demography 52*(6), 1797–1823.

Sloane, C. M., E. G. Hurst, and D. A. Black (2021, November). College majors, occupations, and the gender wage gap. *Journal of Economic Perspectives 35*(4), 223–48.

Ureta, M. and F. Welch (1998). Measuring educational attainment: The old and the new Census and BLS taxonomies. *Economics of Education Review 17*(1), 15–30.

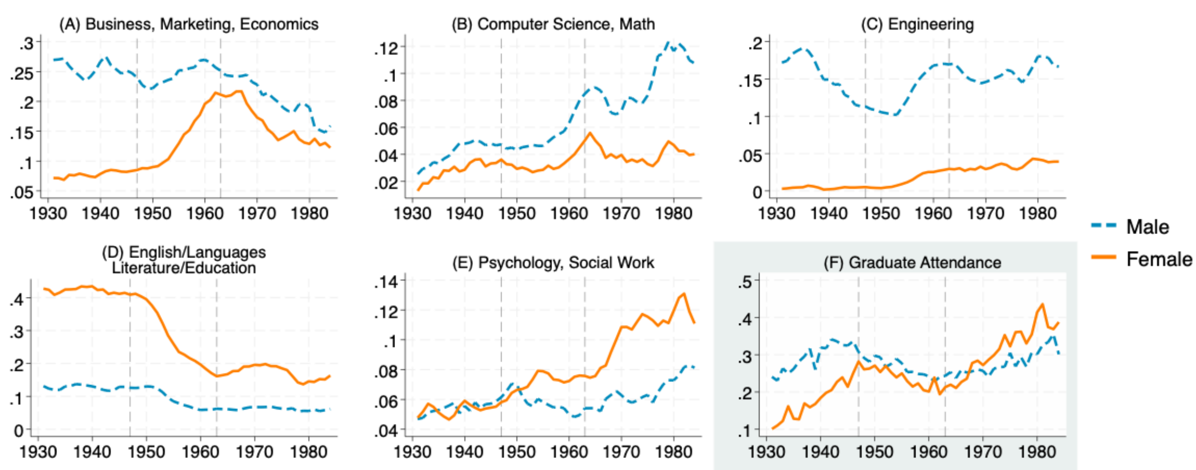
Zafar, B. (2013). College Major Choice and the Gender Gap. *Journal of Human Resources 48*(3), 545–595.

Table 1: Regression Estimates of the Earnings Gap between Men and Women

	(1)	(2)	(3)	(4)
Male (1931-39 birth cohorts)	0.494***	0.349***	0.317***	0.213***
	(0.025)	(0.024)	(0.023)	(0.021)
Male (1940-47 birth cohorts)	0.447***	0.328***	0.296***	0.196***
	(0.025)	(0.024)	(0.023)	(0.021)
Male (1948-63 birth cohorts)	0.371***	0.278***	0.255***	0.165***
	(0.026)	(0.025)	(0.024)	(0.023)
Male (1964-84 birth cohorts)	0.339***	0.271***	0.257***	0.162***
	(0.028)	(0.027)	(0.026)	(0.024)
Constant	10.88***	10.97***	10.82***	11.02***
	(0.026)	(0.027)	(0.026)	(0.025)
Baseline controls	Y	Y	Y	Y
College major		Y	Y	Y
Grad field of study			Y	Y
Occupation				Y
N	361,054	361,054	361,054	335,757

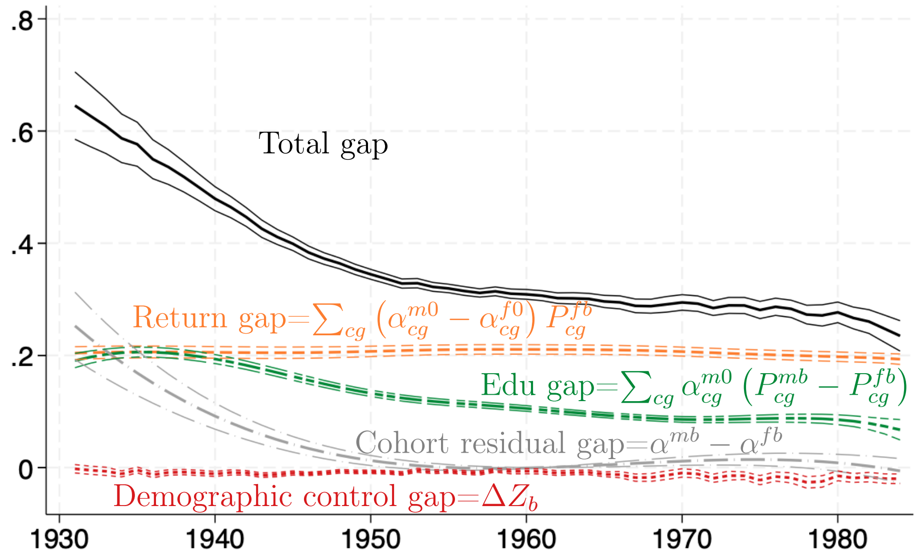
Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The reported estimates are for our race reference group, white and non-Hispanic individuals, averaged from ages 28 and 52. Baseline controls include a cubic in age interacted with a male dummy, a cubic in birthyear, race indicators interacted with gender, and parent education dummies. We also include a regression index of interactions between age and year of birth as a control. The coefficients that define the index are estimated using the Census/ACS data, as discussed in Section 4.3.2.

Figure 1: Aggregate Trends of College Majors and Graduate Attainment by Gender

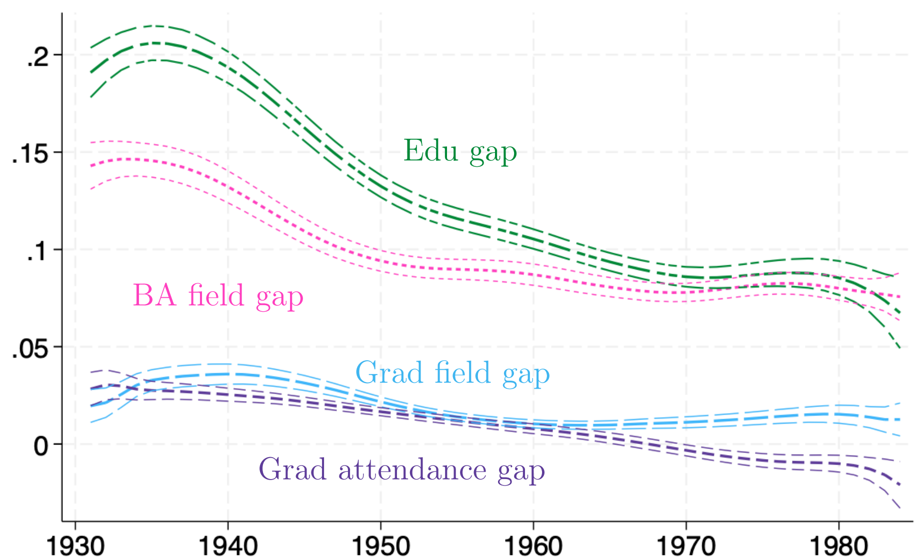


Notes: Panels A-E show the proportion of men and women in specific college majors by birth cohorts between 1931 and 1984. Panel A is Business, Marketing, and Economics majors. Panel B is Computer Science and Mathematics. Panel C is Engineering. Panel D is English/Languages/Literature and Education. Panel E is Psychology and Social Work. Panel F shows that among college graduates, the proportion of men and women with graduate degrees by age 35 by birth cohort. The blue dash line shows the male proportion and the orange solid line shows the female proportion. The proportions are based on 3 year moving averages and are calculated using the NSCG.

Figure 2: Decomposition of Log Earnings, Constant Returns



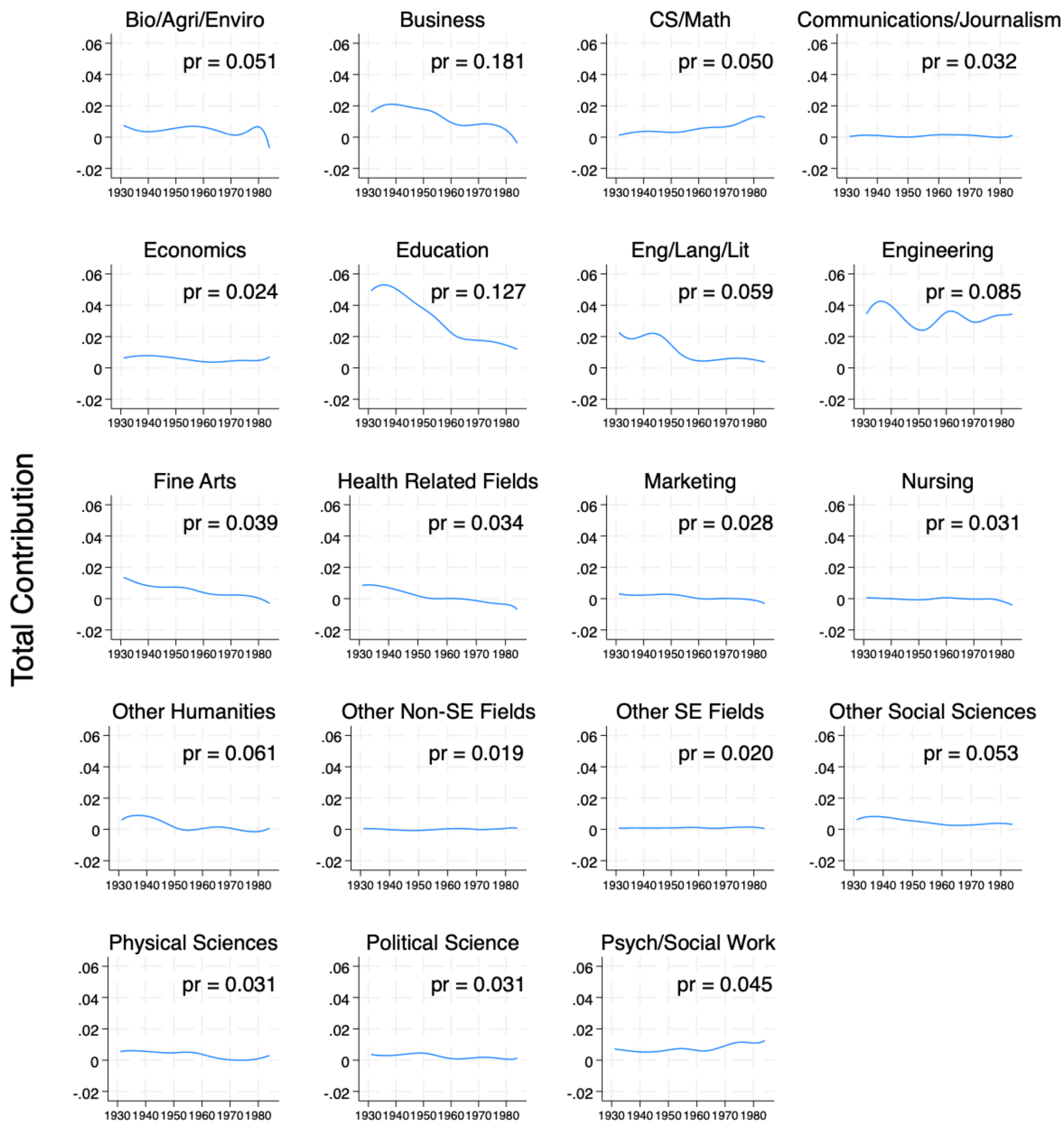
(A) Total gap



(B) Education gap

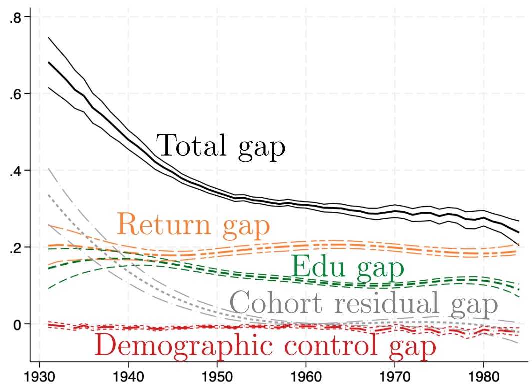
Notes: Panel A shows the decomposition of the predicted gender gap in log earnings for each birth cohort averaged from age 28 to 52. The black line shows the total gender log earnings gap, the orange line shows the portion of the gap at birth year b explained by the gender differences in returns to degrees (including the 1961 value of a component common to all college graduates), the green line shows the education gap, the gray line shows the cohort residual gap that is not related to education fields (normed to 0 in 1961), and the red line shows the demographic control gap. The estimates are constructed using OLS estimates of equation (4). The NSCG base year samples are used with cross sectional weights. Ages are restricted to be between 23 and 59. By construction, Total gap = Return gap + Education gap + Birth cohort residual gap + Demographic gap. Panel B shows the decomposition of the education gap based on equation (5). The green line is the education gap, copied from Panel A. The pink line shows the contribution of college majors, the purple line shows the contribution of graduate attendance, and the blue line shows the contribution of graduate degree field conditional on college major.

Figure 3: Disaggregating the Education Gap by College Major

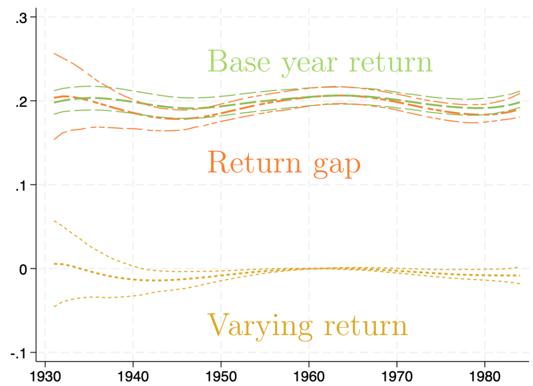


Notes: The figure shows the education gap disaggregated into the contribution of each college major. The subfigures plot $\sum_g (\alpha^{m0}_{cg} - \alpha^{m*})(P^{mb}_{cg} - P^{fb}_{cg})$ for each major c, where α^{m*} is the average of the earnings intercepts α^{m0}_{cg} weighted by average of the c,g probabilities for men and women. Specifically, $\alpha^{m*} \equiv \sum_{cg} \alpha^{m0}_{cg} \bar{P}_{cg}$, where $\bar{P}_{cg}$ is the average probability for each cg combination across sex and birth cohorts. We center α^{m0}_{cg} around α^{m*} to remove the influence of choice of earnings unit (annual), choice of the price deflator, and the presence in α^{m0}_{cg} of the 1961 value of a gender specific component that is common to all cg combinations. The α^{m*} component sums to zero in $\sum_{cg} \alpha^{m0}_{cg} \left(P^{mb}_{cg} - P^{fb}_{cg}\right)$ but not in the c specific components of that sum. The male and female probabilities for each c are presented in Appendix B.1. The curves in the major specific panels sum to the education gap graphed as the green line in Figure 2. The probability pr shown in each subfigure is $\bar{P}_{cg}$.

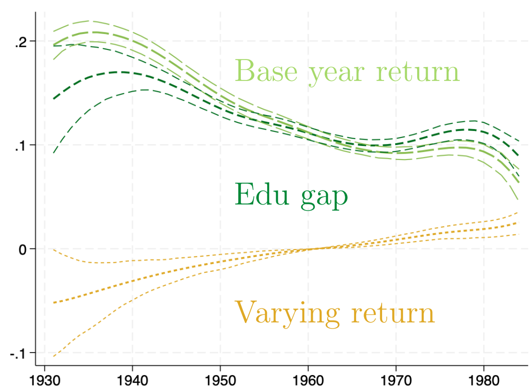
Figure 4: Decomposition of Log Earnings, Cohort Specific Relative Returns



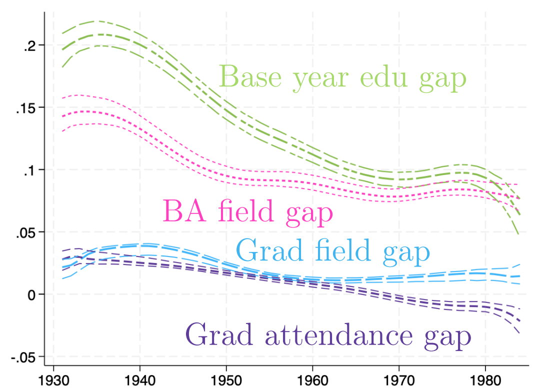
(A) Total gap



(B) The role in the return gap of base year and cohort varying relative returns



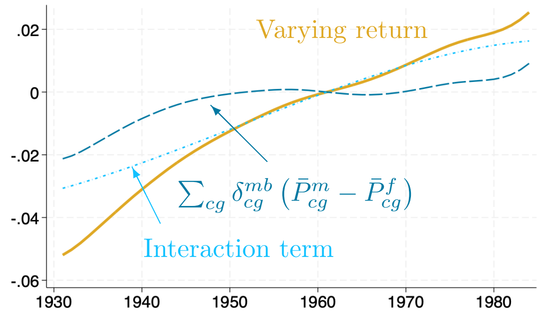
(C) The role in the education gap of base year and cohort varying relative returns



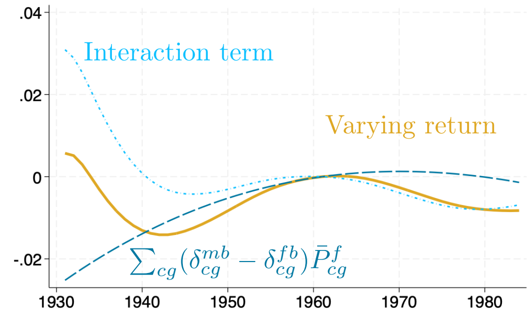
(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: Panel A shows the predicted gender gap in log earnings for each birth cohort averaged from age 28 to 52. The black line shows the total gender log earnings gap, the orange line shows the portion of the gap at b explained by the gender differences in returns to degrees, the green line shows the education gap, the gray line shows the cohort residual gap that is not related to education fields, and the red line shows the demographic control gap. The coefficient estimates are from regression model (10). Panel B decomposes the return gap, $\sum_{cg}\left(\alpha_{cg}^{m0}+\delta_{cg}^{mb}-\alpha_{cg}^{f0}-\delta_{cg}^{fb}\right)\times P_{cg}^{mb}$, into two components. The light green line uses the base year return α_{cg}^{m0}, so it is comparable with the orange in Figure 2. The yellow uses the gender specific, cohort varying relative return δ_{cg}^{mb}. Panel C decomposes the education gap, $\sum_{cg}\alpha_{cg}^{mb}\left(P_{cg}^{mb}-P_{cg}^{fb}\right)$, into two components in the same way as panel B. They sum up to the green line, the education gap. Panel D decomposes the base year return education gap into three components, the contributions of undergrad field in pink, grad attendance in purple, and grad field in blue. They sum up to the education gap with base year return.

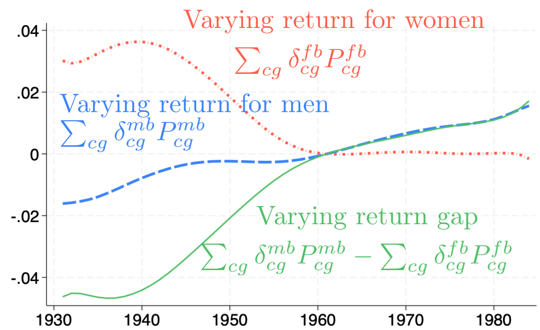
Figure 5: Decomposition of the Varying Return Components of the Education and Return Gaps in Log Earnings



(A) Education gap, log earnings



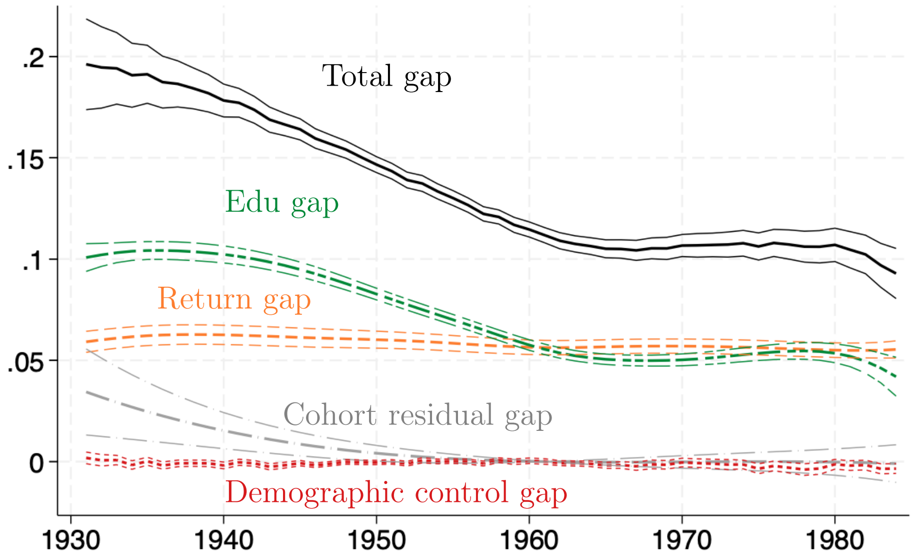
(B) Return gap, log earnings



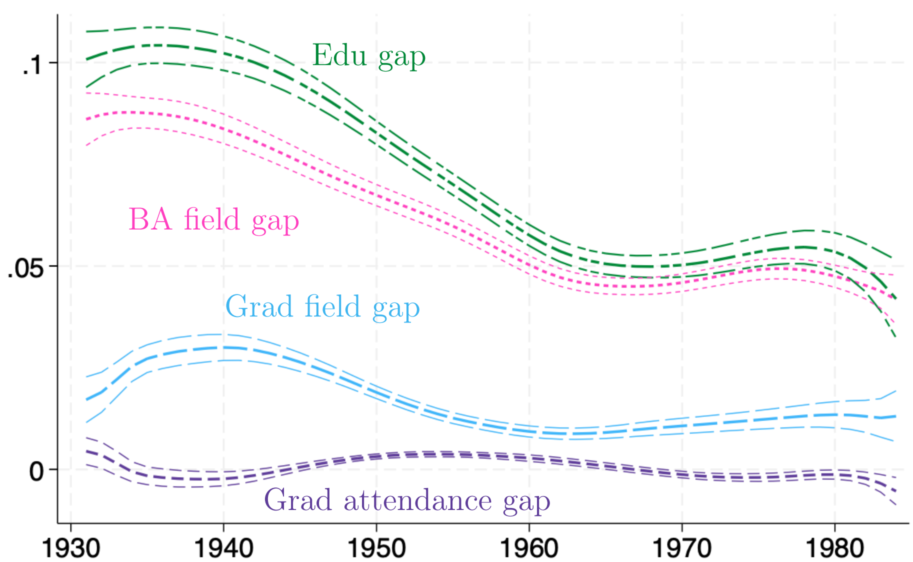
(C) Varying returns for men and women, log earnings

Notes: Panel A decomposes the varying return component of the education gap, $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$. The yellow solid line is, $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$, the same as in Figure 4 panel C. The dark blue dashed line is the graph of $\sum_{cg} \delta_{cg}^{mb} \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)$ where $\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}$ is the difference in average probability for men and women (summand labeled on the figure). The interaction term $\sum_{cg} \delta_{cg}^{mb} \left(\left(P_{cg}^{mb} - P_{cg}^{fb}\right) - \left(\bar{P}_{cg}^{m} - \bar{P}_{cg}^{f}\right)\right)$ is the light blue dotted line. Panel B shows the return gap using the gender-specific, cohort varying relative returns in the decomposition of log earnings (yellow solid line), $\sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb}$. This line is the same as the yellow line in Figure 4 panel B. The dark blue dashed line uses the average probability for women (summand labeled on the figure). The interaction term $\sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) \left(P_{cg}^{fb} - \bar{P}_{cg}^{f}\right)$ is the light blue dotted line. Panel C shows the overall gap in the contribution of the gender-specific, cohort varying relative returns in the decomposition of log earnings (green solid line) and the raw sums by gender (blue dashed line for men and red dotted line for women).

Figure 6: Decomposition of Occupation Premium, Constant Returns



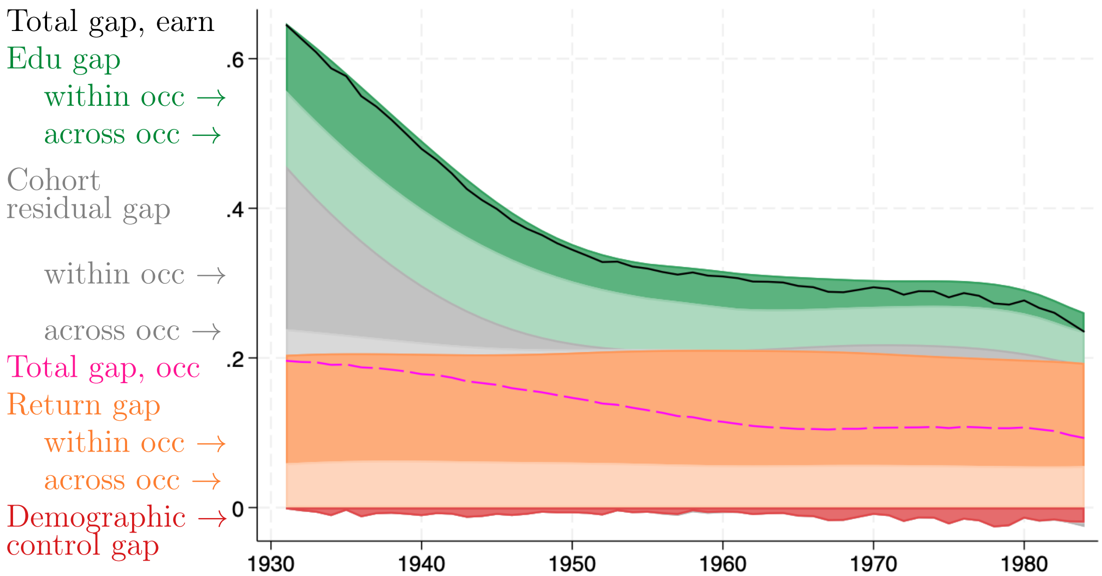
(A) Total gap



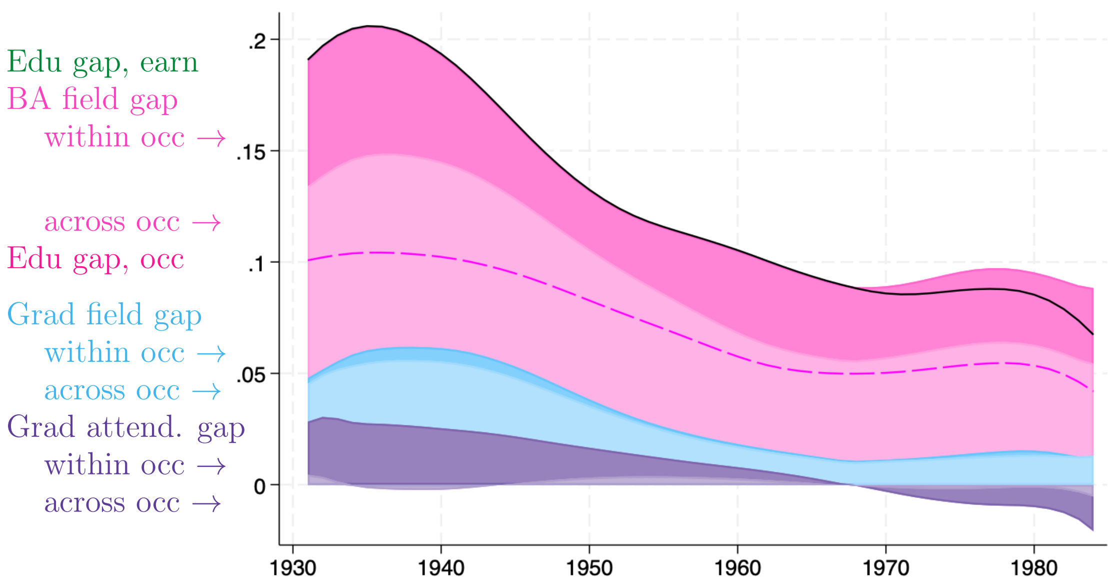
(B) Education gap

Notes: The figure shows the predicted gender gap in the occupation premium for each birth cohort averaged from age 28 to 52. The occupation premiums are estimated as described in section 4.5, and are used as the dependent variable in equation (4) to estimate the gender gap. The definitions of the lines are the same as Figure 2.

Figure 7: Within and Across Occupation Decomposition, Const. Rel. Returns Case



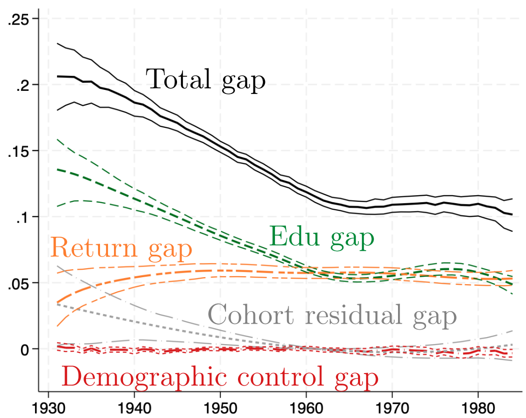
(A) Total gap by occupational effects



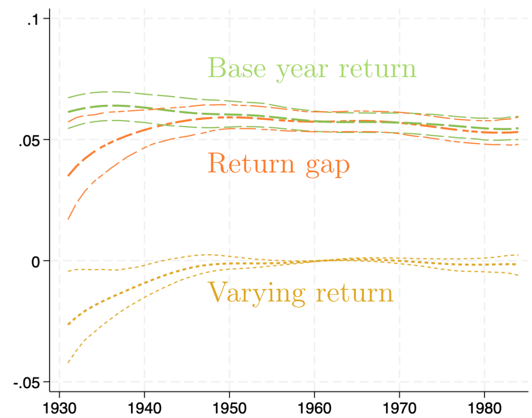
(B) Education gap by occupational effects

Notes: Panel A shows the predicted gender gap broken down by the sources of the gap, i.e. return, education, cohort, or demographic, and by within or across occupation effects. Panel B shows the education gap broken down by the three education choices, i.e. college field, graduate school attendance, or graduate school field, and by within or across occupation effects. The contributions are stacked to show how they constitute the total gap. Negative gaps are shown below the 0 axis, so the total earnings and occupation gaps are smaller than the sum of the positive gaps. The black solid line and pink dashed lines in panel A are the total gap in earnings and in the occupation premium, respectively. The corresponding lines in panel B are for the education gap.

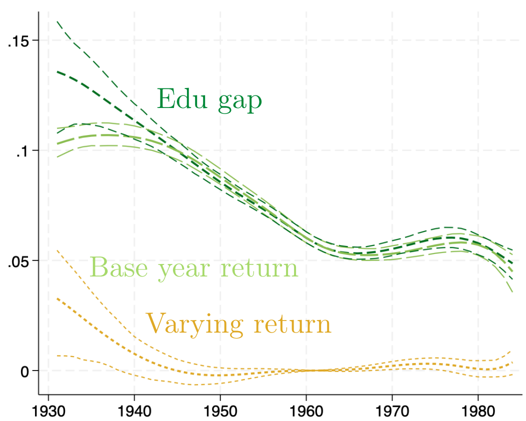
Figure 8: Decomposition of Occupation Premium, Cohort Specific Relative Returns



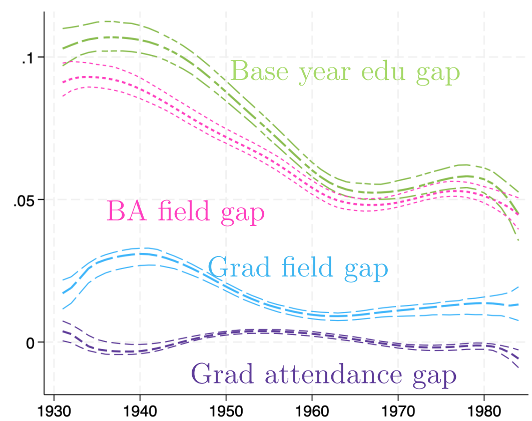
(A) Total gap



(B) The role in the return gap of base year and cohort varying relative returns



(C) The role in the education gap of base year and cohort varying relative returns



(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: This figure shows the predicted gender gap in occupation premium for each birth cohort averaged from age 28 to 52. The occupation premiums are estimated as described in section 4.5, and are used as the dependent variable in equation (10) to estimate the gender gap. The definitions of the lines are the same as Figure 4.

Online Appendix

A The NSCG and Census/ACS Data

This data appendix summarizes the NSCG data we use, the weights, our sample restrictions, variable definitions, and crosswalks. Because we used the NSCG in our previous papers, including Altonji and Zhong (2021) and Altonji et al. (2023), we draw heavily from those papers' data sections and appendices, with parts of the discussion taken verbatim.

A.1 NSCG Data

We employ data from the NSCG, which is part of the Scientists and Engineers Statistical Data System (SESTAT) sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF). The NSCG 1993, 2003, 2010-2019 can be weighted to be representative of the U.S. population with bachelor's degrees.

The NSCG 1993 and 2003 are, respectively, subsamples of the 1990 and 2000 decennial census long form respondents who had a bachelor's degree. From 2010 on, the NSCG employs a new rotating sampling strategy. The NSCG 2010 includes respondents from previous waves but is drawn primarily from respondents to the 2009 American Community Survey (ACS). For example, the sample for the NSCG 2013 combines a subset of the interviewees from the NSCG 2010 and a subset of interviewees with a BA degree from the ACS 2010; the sample for the NSCG 2015 combines a subset of NSCG 2013 and a subset of interviewees with a BA degree from the ACS 2013.

We also use information from a version of the NSCG 1993 that is available from the Inter-university Consortium for Political and Social Research (ICPSR). The ICPSR version includes several variables from the 1990 Census, including occupation based on the 1990 census classification, employment, earnings, and work hours in 1989. In addition to using the 1990 Census information, we obtain information about occupation in 1988 from an NSCG 1993 question. We created occupation categories that are consistent across the census and SESTAT.

A.2 Weights

We construct weights to make the pooled sample representative of the US population of college graduates over the years of our sample, and we use weights unless otherwise noted.

Let $weight_{is}$ denote the survey weights provided by NCSES, where i denotes the person and s denotes a specific wave of a survey, such as the NSCG 2003. We standardize weights across NSCG waves by preserving the relative weights within each survey while accounting for varying sample sizes across surveys. We divide the survey weight by the sum of weights of all observations from the survey and then multiply by the number of observations.

$$weight_{is}^{surv-adj} = \frac{weight_{it}}{\sum_{i=1}^{N_s} weight_{is}} \times N_s$$

A.3 Sample restrictions

We use two primary samples in our analysis. For both samples, we restrict the age of respondents to between 23 and 59. The first is the sample we use to calculate the c and cg specific probabilities. For this sample, we impose a few restrictions. First, we only include respondents from the 1988, 1990, 1993, 2003, and 2010-2019 waves of the NSCG. Second, for respondents who were surveyed in more than one of the NSCG waves that we draw from, we only include their most recent survey response. Third, we restrict the sample to those born after 1930, and before 1984 (Earnings decomposition estimates are the same to 2 decimals if we include observations through the 1995 birth cohort). With these restrictions in place, in total, our probability sample includes 344,101 unique observations, of which 158,938 are for women and 185,163 are for men. Of these, 62,220 women and 66,313 men have advanced degrees.

Our second primary sample is the one we use for the regressions prior to our decomposition analysis. For this sample, we restrict individuals to be (1) full-time workers, (2) must earn at least $5,000 annually, (3) and they must not be enrolled in any educational program in the year that they were surveyed. We define an individual as a full-time worker if she reported working full-time or if she worked at least 41 weeks per year and at least 35 hours per week. We used 41 weeks to accommodate the employment arrangements of many teachers. With these restrictions, our total sample includes 409,358 observations, 243,156 of which are unique respondents. Our sample contains 159,175 (98,729 unique) women and 250,183 (144,427 unique) men. Among them, 65,602 (40,051 unique) women and 88,861 (51,432 unique) men have advanced degrees.

A.4 Variable definitions

log earnings: log earnings is defined as the annual salaried earnings of the respondent. We deflate all earnings to 2013 dollars, and we drop all respondents with earnings less than $5,000.

Log hourly wage: log hourly wage is constructed by dividing respondent's annual salaried income by the product of their annual weeks worked and their average hours worked per week. As with log earnings, we deflate all earnings to 2013 dollars, and drop all respondents with earnings less than \$5,000.

A.5 ACS Data

The second dataset is the Census/ACS data. We use the Census 5% samples from 1960 to 2000, and the 2001 to 2018 ACS data, restricting the sample to people with 4 or more years of college in the 1960 to 1990 samples, and people with a bachelor's degree or higher in the samples after 1990. We deflate earnings to 2013 dollars. We mirror restrictions imposed on the NSCG earnings sample by further restricting the sample to workers who were not enrolled in school, earn more than \$5,000, and who work 35 or more hours per week and 40 or more weeks per year. Top-coded earnings are adjusted by multiplying the top coded value with 1.5. We remove observations with imputed values for the variables we use.

The 1960-1980 Censuses report years of schooling attended and years of schooling completed as the education measures rather than degree attainment. This data limitation poses challenges in defining the population of individuals with a college degree and with a graduate degree. Based on Park (1996); Frazis et al. (1995); Ureta and Welch (1998); Jaeger (1997); Kominski and Siegel (1993)'s assessments of the correspondence between years of schooling completed and degree attainment as well as our own analyses, a case can be made for setting $G(i)$=1 for individuals with 18 or more years of completed schooling and $G(i) = 0$ for individuals with 16 or 17 years of completed schooling. But in the end we set $G(i) = 1$ for individuals with 17 or more years of completed schooling and to 0 for those with 16 years. For the 1990 and 2000 Census and the ACS we use responses to direct questions about degree attainment.

We use the Census/ACS data to estimate occupational earnings premiums, which serve as the dependent variable in some of our analyses using the NSCG. As we explain in Section 4.3.2, we also use the Census/ACS to estimate the regression coefficients relating earnings to a polynomial in birth cohort and age. We use the coefficients to constrain the interactions between birth cohort and age when we estimate earnings functions in the NSCG.

A.6 Crosswalks

We aggregate 157 postsecondary fields into 19 undergraduate majors and 19 graduate majors. The aggregations are provided in Altonji and Zhong (2021)'s Online Appendix Tables B1 and B2.

We created occupation categories that are consistent across the census and SESTAT. In Altonji and Zhong (2021)'s Online Appendix, Table B3 reports the shares of the 363 disaggregated fields in the 66 consistent categories (from column (3) to column (2)), and the shares of those 66 consistent categories in the 21 aggregated occupations (from column (2) to column (1)). The occupational earnings premiums are constructed using the 1960, 1970, 1980, 1990, and 2000 Census 5% samples in addition to the 2001-2018 waves of the ACS. The construction of the occupational premiums are described in Section 4.5.

A.7 Summary statistics

Tables A.1 and A.2 below provide summary statistics for our regression decomposition sample described above, where Table A.1 provides summary statistics for the women in this sample, while Table A.2 does so similarly for men.

Table A.1: Summary Statistics: Female Sample

	1930s (1)	1940s (2)	1950s (3)	1960s (4)	1970s (5)	1980-1984 (6)	1985 -1995 (7)	Total (Analysis Sample) (8)
Total N	4,081	17,053	35,773	38,317	23,935	17,806	22,210	136,965
Annual earnings	61163 [34141]	64893 [37795]	68656 [46746]	70558 [55964]	77301 [56713]	70267 [49743]	59860 [44085]	70216 [50582]
Age surveyed	55.25 [2.31]	47.81 [4.45]	42.71 [8.24]	39.31 [10.13]	37.67 [5.44]	32.48 [2.96]	28.42 [2.43]	40.56 [8.94]
Year earnings obs.	1991.54 [1.50]	1993.33 [4.10]	1997.64 [8.40]	2003.12 [10.52]	2012.11 [5.14]	2014.66 [2.68]	2015.66 [2.15]	2003.20 [10.60]
Birth year	1936.29 [2.21]	1945.52 [2.71]	1954.93 [2.84]	1963.81 [2.68]	1974.44 [2.88]	1982.18 [1.41]	1987.24 [1.88]	1962.64 [12.41]
Race/Ethnicity								
Asian	378 (4.7%)	1,730 (5.3%)	3,400 (5.3%)	4,236 (6.1%)	3,543 (9.2%)	2,797 (9.8%)	3,678 (9.6%)	16,084 (6.6%)
Black	675 (9.2%)	2,473 (8.2%)	5,186 (9.2%)	4,664 (8.7%)	2,761 (10.4%)	1,938 (7.9%)	2,320 (8.3%)	17,697 (9.0%)
Native American	51 (0.3%)	216 (0.4%)	414 (0.5%)	374 (0.6%)	299 (0.9%)	221 (0.6%)	273 (0.7%)	1,575 (0.6%)
White-H	26 (0.3%)	214 (0.7%)	1,149 (2.2%)	2,110 (3.9%)	2,531 (7.4%)	2,061 (7.4%)	2,673 (10.0%)	8,091 (3.8%)
White-NH	2,919 (85.1%)	12,215 (84.8%)	24,837 (81.7%)	26,001 (79.4%)	13,947 (69.5%)	9,932 (70.9%)	11,936 (67.1%)	89,851 (78.4%)
Other	32 (0.3%)	205 (0.6%)	787 (1.2%)	932 (1.3%)	854 (2.6%)	857 (3.4%)	1,330 (4.3%)	3,667 (1.6%)
Father Education Level								
Less than high school	1,613 (38.5%)	5,297 (28.5%)	7,425 (18.5%)	5,119 (11.8%)	2,185 (8.7%)	1,224 (5.8%)	1,450 (6.4%)	22,863 (15.7%)
High school diploma	960 (23.5%)	4,560 (28.1%)	9,922 (28.7%)	10,093 (28.0%)	5,641 (24.9%)	3,629 (24.2%)	4,507 (22.3%)	34,805 (27.2%)
Some college	659 (17.1%)	3,025 (18.6%)	6,782 (19.7%)	7,809 (20.5%)	5,228 (22.9%)	3,903 (22.0%)	4,694 (20.8%)	27,406 (20.5%)
College degree	432 (10.2%)	2,145 (12.8%)	6,033 (17.4%)	7,773 (20.4%)	5,657 (22.7%)	4,499 (23.4%)	5,800 (26.4%)	26,539 (18.9%)
Graduate education	417 (10.7%)	2,026 (11.9%)	5,611 (15.7%)	7,523 (19.3%)	5,224 (20.8%)	4,551 (24.6%)	5,759 (24.1%)	25,352 (17.8%)
Mother Education Level								
Less than high school	1,447 (32.8%)	4,444 (22.1%)	6,344 (14.2%)	4,595 (9.9%)	2,334 (8.5%)	1,186 (5.3%)	1,447 (5.8%)	20,350 (12.8%)
High school diploma	1,309 (31.8%)	6,261 (39.2%)	13,357 (40.0%)	13,253 (36.6%)	6,414 (29.1%)	3,822 (23.5%)	4,167 (20.7%)	44,416 (35.2%)
Some college	770 (20.5%)	3,487 (21.6%)	8,207 (23.7%)	9,404 (25.1%)	6,217 (26.5%)	4,724 (28.0%)	5,634 (27.1%)	32,809 (24.6%)
College degree	386 (10.6%)	1,917 (11.4%)	5,010 (14.4%)	6,921 (18.2%)	5,430 (22.1%)	4,652 (25.0%)	6,392 (26.9%)	24,316 (17.3%)
Graduate education	169 (4.3%)	944 (5.6%)	2,855 (7.7%)	4,144 (10.2%)	3,540 (13.8%)	3,422 (18.2%)	4,570 (19.6%)	15,074 (10.0%)

Notes: Table reports the summary statistics for all female observations in our regression sample. All continuous variable (Annual Earnings, Age surveyed, Year earnings observed, and Birth year) cells report first the mean and then the standard deviation (in brackets) from the unweighted sample. For all categorical variables (Total N, Race/Ethnicity, and Father/Mother Education Level), the cells report the unweighted N count first, and then the weighted percent of the total decade sample (in parentheses). The Race/Ethnicity category of Black includes both black Hispanics and Black non-Hispanics. Column (8) reports the totals for the sample used in our analysis (1931 - 1984), which includes columns (1) through (6). The education level category, "Graduate Education," includes masters, professional, and doctorate degrees. For mother's education, there was an additional category of "missing," which was absorbed into the "Less than high school" category.

Table A.2: Summary Statistics: Male Sample

Variable	1930s	1940s	1950s	1960s	1970s	1980-1984	1985 -1995	Total (Analysis Sample)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total N	10,813	34,378	62,390	61,325	34,200	20,983	26,094	224,089
Annual earnings	104647 [81634]	98356 [71323]	94420 [71219]	95403 [79881]	99103 [79187]	87434 [63412]	74535 [52459]	95847 [74882]
Age surveyed	55.38 [2.36]	47.72 [4.47]	42.67 [8.36]	40.74 [9.90]	38.27 [5.24]	32.57 [2.98]	28.71 [2.45]	41.91 [8.90]
Year earnings obs.	1991.31 [1.49]	1993.15 [4.13]	1997.58 [8.58]	2004.55 [10.38]	2012.47 [4.95]	2014.79 [2.71]	2015.88 [2.11]	2002.39 [10.67]
Birth year	1935.94 [2.31]	1945.43 [2.73]	1954.91 [2.85]	1963.81 [2.73]	1974.20 [2.88]	1982.22 [1.43]	1987.16 [1.86]	1960.48 [12.48]
Race/Ethnicity								
Asian	820 (3.8%)	2,711 (3.8%)	5,399 (4.3%)	6,779 (6.0%)	5,696 (11.3%)	4,007 (12.2%)	4,719 (12.1%)	25,412 (6.2%)
Black	653 (3.8%)	2,178 (4.0%)	4,273 (4.7%)	3,769 (5.2%)	2,194 (6.0%)	1,157 (5.1%)	1,345 (5.0%)	14,224 (4.9%)
Native American	102 (0.3%)	339 (0.3%)	519 (0.4%)	483 (0.6%)	408 (0.9%)	193 (0.6%)	210 (0.8%)	2,044 (0.5%)
White-H	127 (0.6%)	525 (0.8%)	1,741 (1.5%)	3,494 (4.2%)	3,266 (7.1%)	1,918 (6.8%)	2,349 (8.4%)	11,071 (3.2%)
White-NH	9,049 (91.3%)	28,187 (90.5%)	49,409 (88.1%)	45,571 (82.8%)	21,760 (73.2%)	12,897 (72.5%)	16,340 (70.2%)	166,873 (84.0%)
Other	62 (0.3%)	438 (0.7%)	1,049 (1.0%)	1,229 (1.3%)	876 (1.6%)	811 (2.8%)	1,131 (3.5%)	4,465 (1.2%)
Father Education Level								
Less than High School	4,746 (43.2%)	10,497 (29.2%)	11,142 (16.1%)	6,948 (10.0%)	2,594 (7.1%)	1,075 (4.1%)	1,199 (4.6%)	37,002 (16.1%)
High school diploma	2,628 (24.8%)	10,563 (31.6%)	18,334 (31.3%)	16,059 (26.5%)	7,736 (23.3%)	4,027 (21.6%)	4,724 (19.9%)	59,347 (27.9%)
Some college	1,486 (14.0%)	5,765 (17.4%)	11,074 (17.7%)	11,085 (18.5%)	6,540 (19.2%)	4,069 (19.5%)	4,754 (17.8%)	40,019 (18.0%)
College Degree	1,111 (10.5%)	4,187 (11.8%)	12,087 (19.2%)	14,456 (24.0%)	9,023 (25.6%)	6,156 (30.1%)	7,967 (30.4%)	47,020 (20.4%)
Graduate Education	842 (7.5%)	3,366 (10.0%)	9,753 (15.8%)	12,777 (21.1%)	8,307 (24.8%)	5,656 (24.8%)	7,450 (27.3%)	40,701 (17.7%)
Mother Education Level								
Less than High School	3,907 (33.9%)	8,482 (21.8%)	9,886 (12.8%)	7,037 (9.4%)	3,130 (7.6%)	1,387 (4.5%)	1,226 (4.4%)	33,829 (13.3%)
High school diploma	4,115 (39.3%)	15,260 (46.5%)	26,783 (45.4%)	22,984 (38.5%)	9,709 (29.6%)	4,770 (24.0%)	5,232 (21.1%)	83,621 (39.7%)
Some college	1,561 (15.1%)	5,791 (17.4%)	12,223 (19.8%)	13,311 (22.8%)	7,774 (23.3%)	4,623 (23.7%)	5,790 (22.0%)	45,283 (20.7%)
College Degree	941 (9.2%)	3,423 (10.1%)	9,173 (15.2%)	11,815 (19.4%)	8,182 (24.3%)	5,995 (29.1%)	8,425 (32.1%)	39,529 (17.4%)
Graduate Education	289 (2.5%)	1,422 (4.1%)	4,325 (6.7%)	6,178 (10.0%)	5,405 (15.3%)	4,208 (18.7%)	5,421 (20.4%)	21,827 (9.0%)

Notes: Table reports the summary statistics for all male observations in our regression sample. All continuous variable (Annual Earnings, Age surveyed, Year earnings observed, and Birth year) cells report first the mean and then the standard deviation (in brackets) from the unweighted sample. For all categorical variables (Total N, Race/Ethnicity, and Father/Mother Education Level), the cells report the unweighted N count first, and then the weighted percent of the total decade sample (in parentheses). The Race/Ethnicity category of Black includes both black Hispanics and Black non-Hispanics. Column (8) reports the totals for the sample used in our analysis (1931 - 1984), which includes columns (1) through (6). The education level category, "Graduate Education," includes masters, professional, and doctorate degrees. For mother's education, there was an additional category of "missing," which was absorbed into the "Less than high school" category.

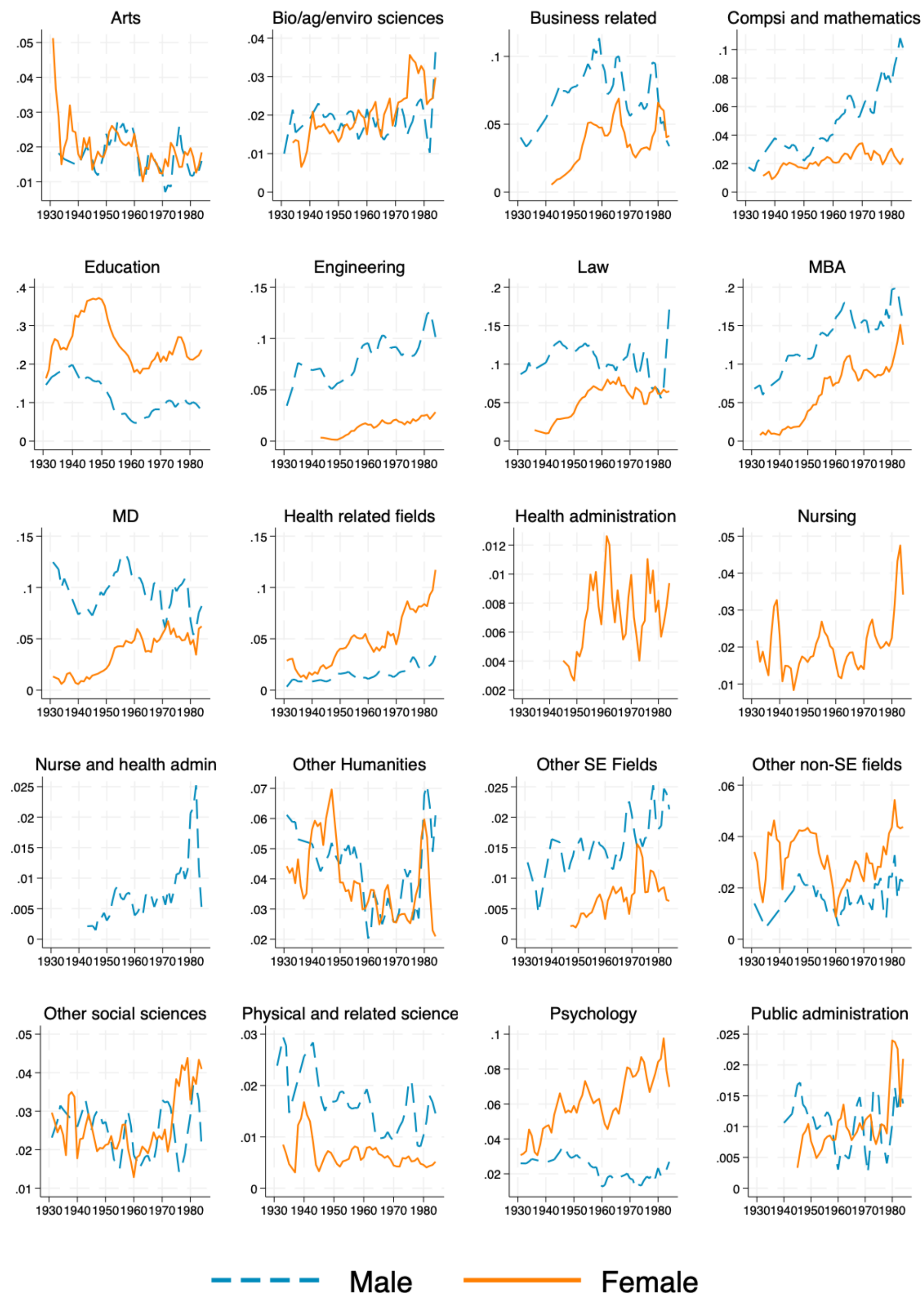
B NSCG Estimates of Aggregate Trends in College and Graduate Majors by Gender

We estimate the probabilities of men and women choosing each college major by birth cohort. Figure B.1 shows the aggregate trends. We also estimate the probabilities of men and women choosing each graduate major conditioning on having a graduate degree by birth cohort and show the results in Figure B.2. The estimates are based on 3 year moving averages.

Figure B.1: Aggregate Trends in College Majors by Gender

Notes: The figure shows the proportion of men and women in specific college majors by birth cohorts from 1931 and 1984. The blue dashed line shows the male proportion and the orange solid line shows the female proportion. The estimates are based on 3 year moving averages. The early birth years for men in Nursing, women in Marketing, and women in Other SE Fields are not reported due to low cell counts.

Figure B.2: Aggregate Trends of Graduate Fields by Gender



Notes: The figure shows the proportion of men and women in specific graduate fields by birth cohorts from 1931 and 1984, conditional on having a graduate degree. The blue dash line shows the male proportion and the orange solid line shows the female proportion. Some observations are omitted due to small cell counts. Although we have 19 graduate degree categories, there are 20 panels because we merge together Nursing and Health Administration fields for men due to small cell counts. The estimates are based on 3 year moving averages.

C Estimates of Trends in BA and Graduate Degrees Based on HEGIS and IPEDS

In our decomposition, we rely on having the marginal distributions of undergraduate and graduate degrees. The NSCG over-samples graduates from STEM fields and relies on graduates to recall their exact degree name, which may have been obtained many years ago. We then use sample weights to create a nationally representative sample. An alternative is to use the HEGIS/IPEDS data to create estimates of the marginal distributions of undergraduate and graduate fields of study. Note that the HEGIS/IPEDS data cannot be used to estimate probabilities of graduate fields conditional on undergraduate field. It has the disadvantage that we must make assumptions about the age at which people obtain undergraduate and graduate degrees. Furthermore, the HEGIS/IPEDS data includes degrees obtained by foreign students who do not remain in the US. This may lead to bias to the extent that field choices of such students differ by gender from those who reside in the US.

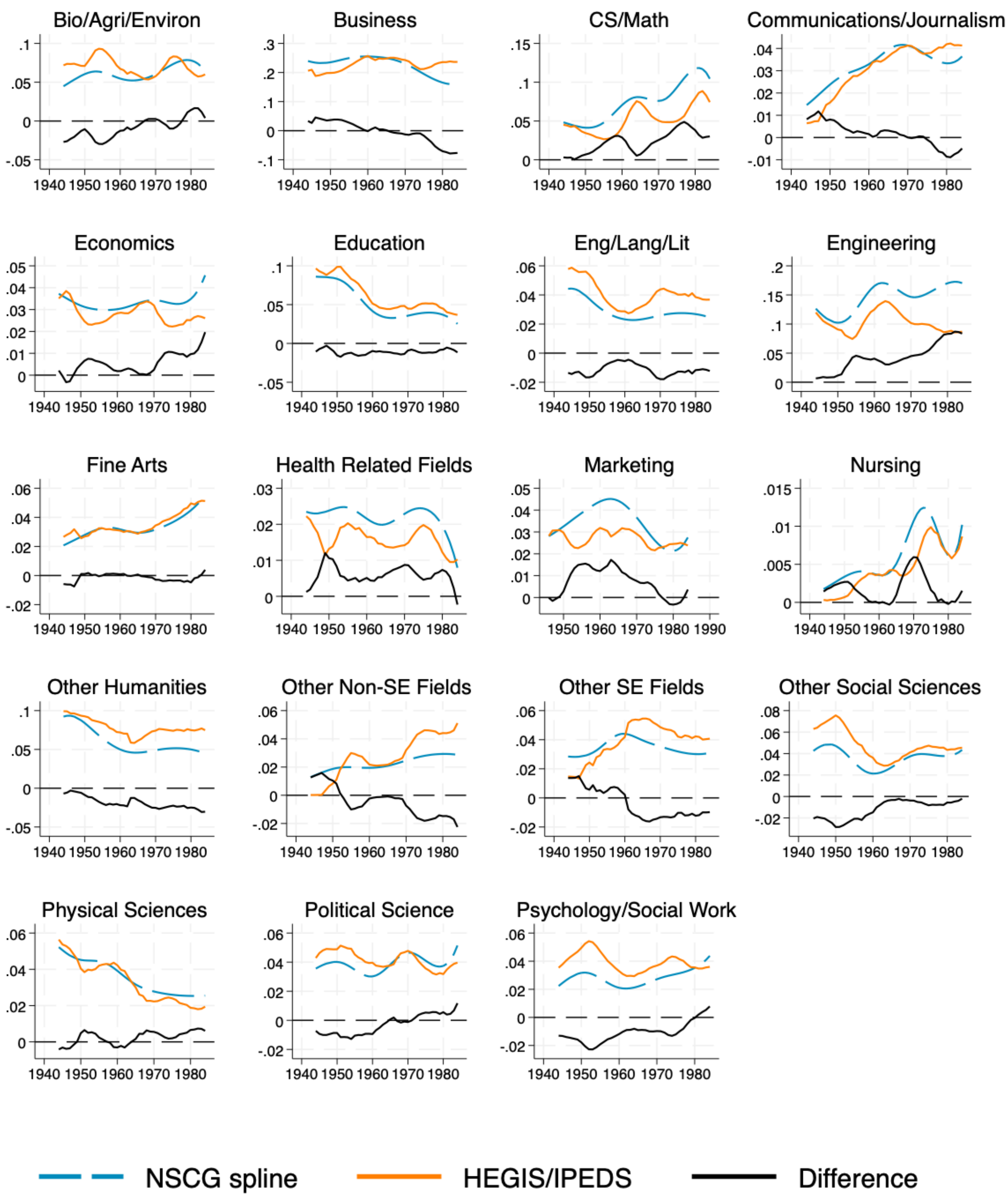
To use HEGIS/IPEDS data, we relied on multiple crosswalks to aggregate the degrees to the 19 undergraduate and graduate classifications used in the paper. The HEGIS data spans from 1966-1985. During this period, it used three different classification systems. The first two classifications, 1966-1969 and 1971-1982, use a classification unique to HEGIS. The taxonomy from 1982-1985 uses an older form of CIP codes that would become the basis of the subject codes used in IPEDS. For each taxonomy, we created a crosswalk between the HEGIS dataset and the NSF degree classification used in the NSCG surveys. We then used our existing crosswalk to aggregate into our 19 undergraduate and graduate degrees. There is no HEGIS data for 1970.

The IPEDS data spans from 1985-2019. All years of the data use the CIP codes to classify degree subjects, though the classifications are updated at the beginning of every decade. We use crosswalks provided by the National Center for Education Statistics (NCES) to convert all CIP codes to the 2010 classification system. We then use the 4-digit CIP code to aggregate into our 19 undergraduate and graduate degrees. In 1985, we observe data from both HEGIS and IPEDS and use the average of the two.

Figures C.1 and C.2 show the marginal distributions of undergraduate degrees in the NSCG (blue), the marginal distributions from HEGIS/IPEDS (red), and their difference (green). The NSCG estimates are based on the b-spline approach discussed in section 4.3.1. Since HEGIS/IPEDS only provides the year the degree was conferred, we assume people are 22 years old when they receive their degree to impute their birth year. This assumption allows us to compare the marginals for birth cohorts going back to 1944. The NSCG shows more graduates with engineering and computer science/math degrees than HEGIS/IPEDS for both men and women. In business, the NSCG based estimates are very similar to the HEGIS/IPEDS for men. For women, the NSCG based estimates for business degrees are above the HEGIS/IPEDS based estimates in early birth cohorts, but the difference turns negative after the 1970 birth cohort. For education degrees, we again see the NSCG and HEGIS/IPEDS show very similar results for men. However, for women, the NSCG probabilities are about 0.1 below the HEGIS/IPEDS values between 1944 and 1950. This difference then quickly moves towards 0 for the rest of the birth cohorts.

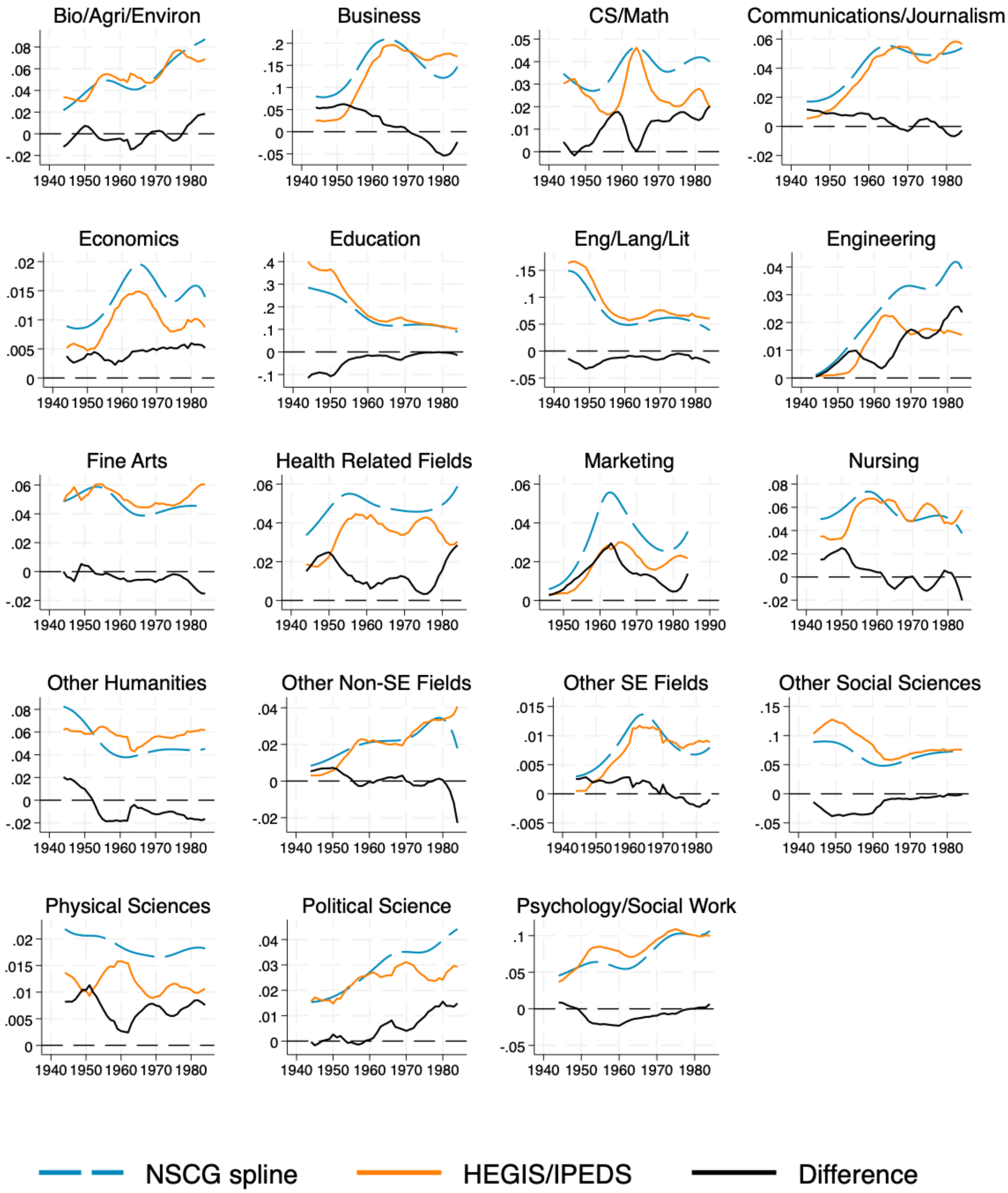
Figures C.3 and C.4 show the marginal distributions for graduate majors for men and women respectively for those who get a graduate degree. Comparisons with the NSCG are harder for graduate degrees because there is a large variation in the age of attainment for graduate degrees, even within subject and gender. We impute the birth cohort for graduate degrees in the HEGIS/IPEDS data by subtracting the average age of attainment given the degree field and gender from the year of conferral. The trends for popular graduate degrees are similar between the HEGIS/IPEDS and the NSCG. In early birth cohorts, we see that the fraction of men who obtained graduate degrees in engineering is larger in HEGIS/IPEDS than in our NSCG sample, hovering around 0.1 male graduate degree earners compared to around 0.07. The NSCG and HEGIS/IPEDS both capture the large growth in women receiving Law, MBA, and MD degrees between the 1940 and 1960 cohorts. The HEGIS/IPEDS data shows a much larger drop in humanities degrees in earlier birth cohorts, especially for women. Between the mid 30s and the 1950 cohorts the HEGIS/IPEDS data also shows a decrease relative to the NSCG in the fraction of women obtaining Non-Science and Engineering (SE). Both data sources show a small increase in the fraction of men receiving these degrees. For SE-related graduate degrees, we see growth for both men and women attaining graduate degrees, with the gap widening over time in favor of men. The HEGIS/IPEDS data show more growth for both men and women than the NSF data does.

Figure C.1: Trends in Undergraduate Field Choice by Gender, NSCG versus HEGIS/IPEDS: Males



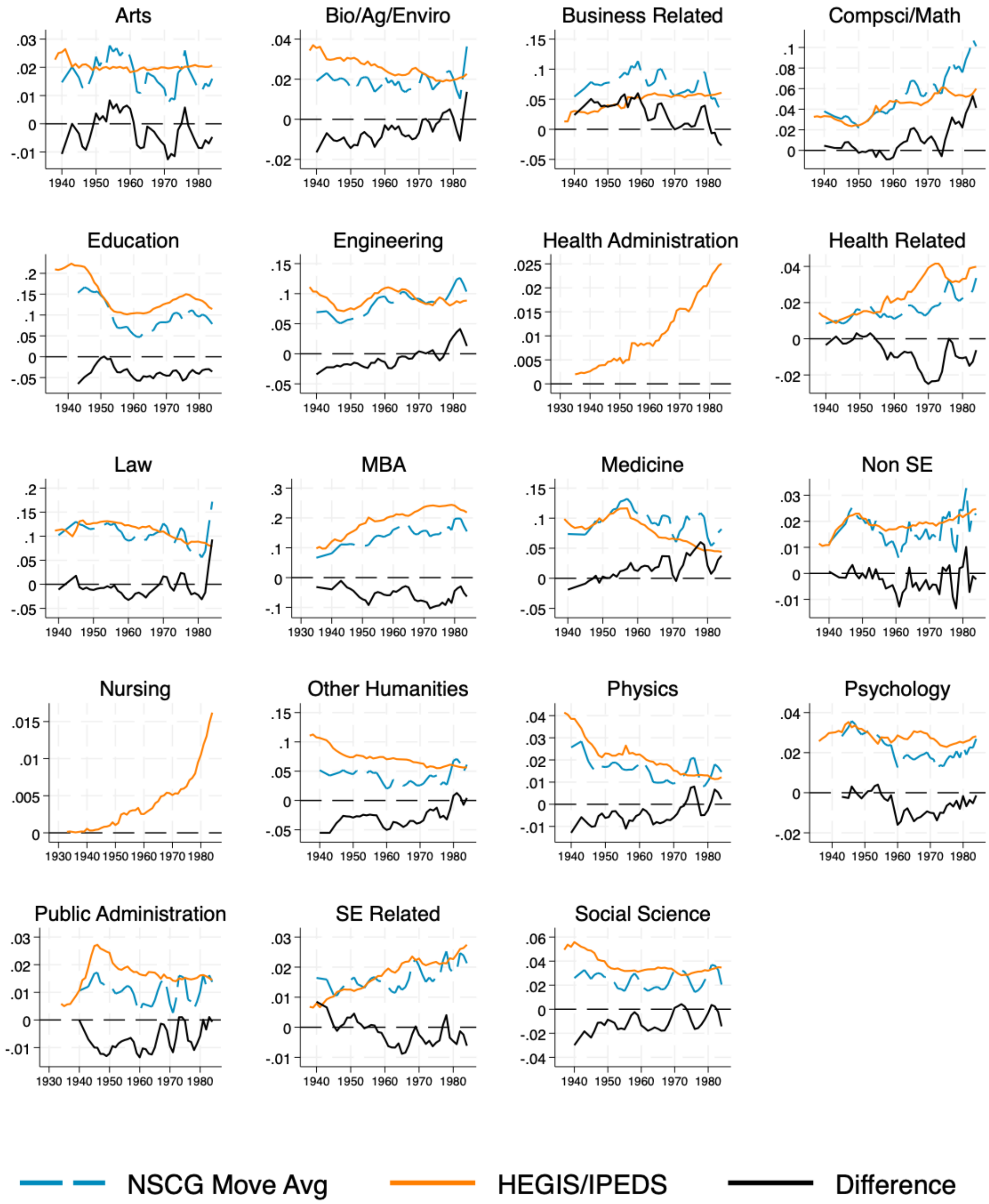
Notes: This figure displays estimates of the fraction of male college graduates in each college major by birth cohort using the NSCG (dashed blue) and HEGIS/IPEDS (solid orange) data. The NSCG estimates are based on a b-spline. The black line shows the difference between the two. We assume the age of obtaining a college degree is 22.

Figure C.2: Trends of Undergraduate Field Choice by Gender, NSCG versus HEGIS/IPEDS: Females



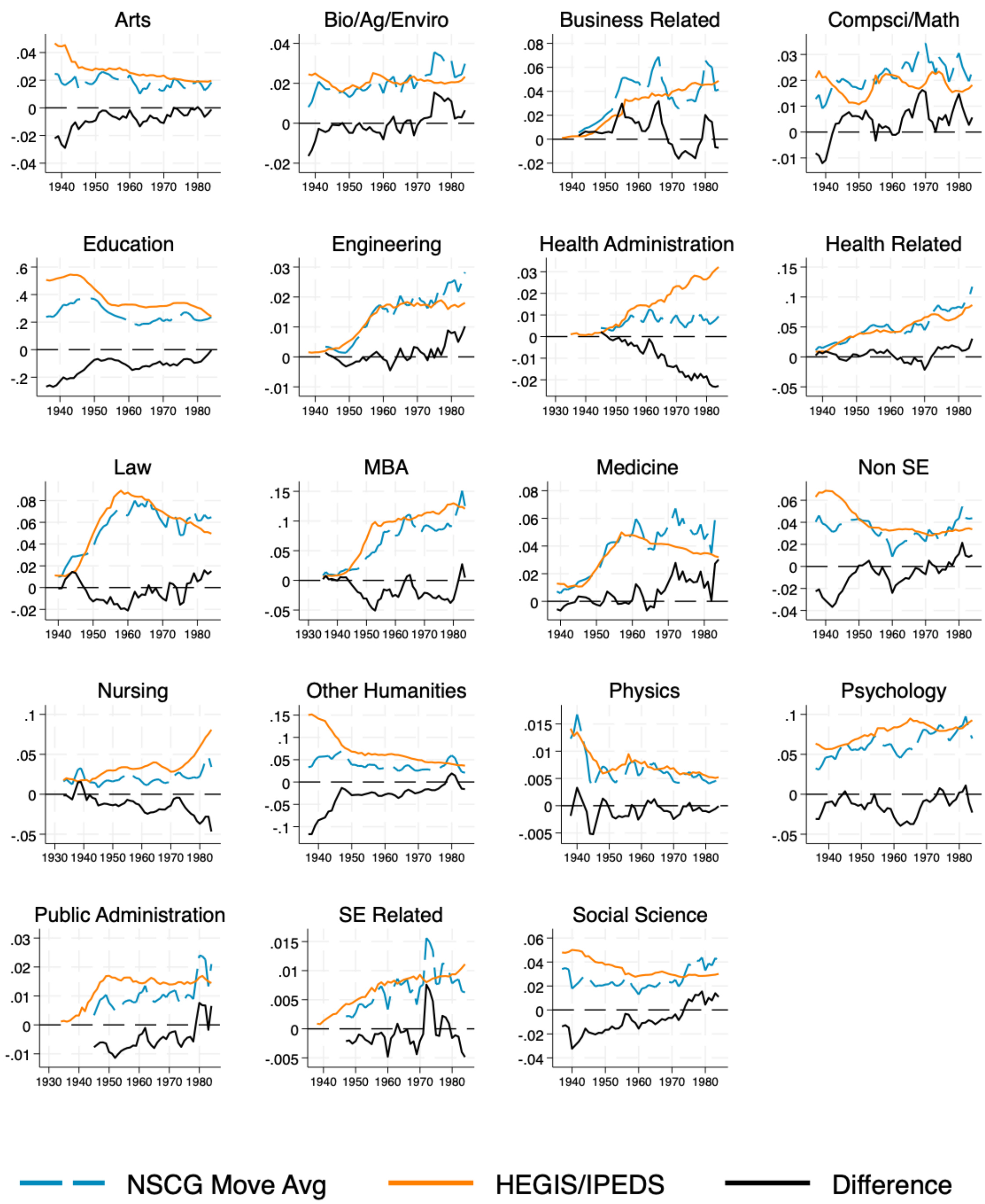
Notes: This figure displays estimates of the fractions of female college graduates in each college major by birth cohort using the NSCG (dashed blue) and HEGIS/IPEDS (solid orange) data. The NSCG estimates are based on a b-spline. The black line shows the difference between the two. We assume the age of obtaining a college degree is 22.

Figure C.3: Trends in Graduate Fields by Gender NSCG versus HEGIS/IPEDS: Males



Notes: The figure displays estimates of the fraction of male advanced degree holders in each graduate field using the NSCG (dashed blue) and the HEGIS/IPEDS (solid orange) data. The NSCG estimates are based on a b-spline. The difference is shown in black. To estimate the birth cohort for HEGIS/IPEDS, we calculated the average age of attainment for each graduate degree in the NSCG and subtracted it from the year of conferral in the HEGIS/IPEDS data.

Figure C.4: Trends of Graduates Fields by Gender, NSCG versus HEGIS/IPEDS: Females



Notes: The figure displays estimates of the fraction of female advanced degree holders in each graduate field using the NSCG (dashed blue) and the HEGIS/IPEDS (solid orange) data. The NSCG estimates are based on a b-spline. The difference is shown in black. To estimate the birth cohort for HEGIS/IPEDS, we calculated the average age of attainment for each graduate degree in the NSCG and subtracted it from the year of conferral in the HEGIS/IPEDS data.

D Formula for the Decomposition of the Education Gap with Cohort Specific Relative Returns

In the cohort specific relative returns decomposition, we can fully decompose the education gap as

$$
\begin{aligned}
&\text{Education Gap } (b) = \\
&\sum_{cg} \alpha_{cg}^{m0} \times \left(P_{g|G,c}^{fb} - P_{g|G,c)}^{fb}\right) \times P_{G|c}^{fb} \times P_{c}^{fb} && \text{grad field gap } (\alpha_{cg}^{m0}) \\
&+ \sum_{cg} \alpha_{cg}^{m0} \times P_{g|G,c}^{fb} \times \left(P_{G|c}^{mb} - P_{G|c}^{fb}\right) \times P_{c}^{fb} && \text{grad enroll gap } (\alpha_{cg}^{m0}) \\
&+ \sum_{cg} \alpha_{cg}^{m0} \times P_{g,G|c}^{fb} \times \left(P_{c}^{mb} - P_{c}^{fb}\right) && \text{BA field gap } (\alpha_{cg}^{m0}) \\
&+ \sum_{cg} \delta_{cg}^{mb} \times \left(P_{g|G,c}^{fb} - P_{g|G,c)}^{fb}\right) \times P_{G|c}^{fb} \times P_{c}^{fb} && \text{grad field gap } (\delta_{cg}^{mb}) \\
&+ \sum_{cg} \delta_{cg}^{mb} \times P_{g|G,c}^{fb} \times \left(P_{G|c}^{mb} - P_{G|c}^{fb}\right) \times P_{c}^{fb} && \text{grad enroll gap } (\delta_{cg}^{mb}) \\
&+ \sum_{cg} \delta_{cg}^{mb} \times P_{g,G|c}^{fb} \times \left(P_{c}^{mb} - P_{c}^{fb}\right) && \text{BA field gap } (\delta_{cg}^{mb}) \\
&+ \Delta ED_{b}^{23} && \text{approx. error.}
\end{aligned}
\tag{12}
$$

Table J.6 reports the estimates of this decomposition of the education gap for earnings for selected years, broken out by within occupation and across occupation.

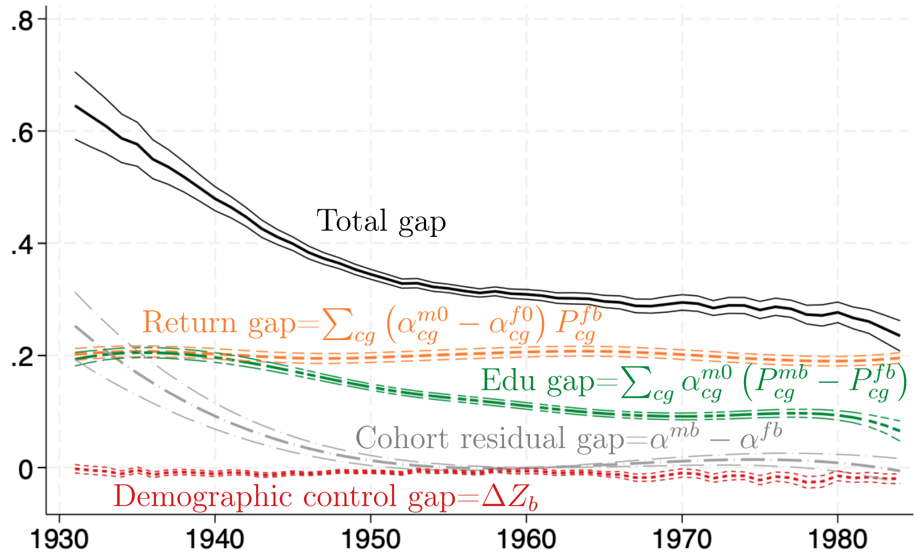
E Gender Gap Decompositions Using Female Earnings Coefficient and Male Degree Probabilities

In this appendix we present gender gap decompositions using female earnings coefficients and male degree probabilities. Overall, the findings are similar. When the results differ from those obtained using the male coefficients and female degree probabilities, we discuss the reasons.

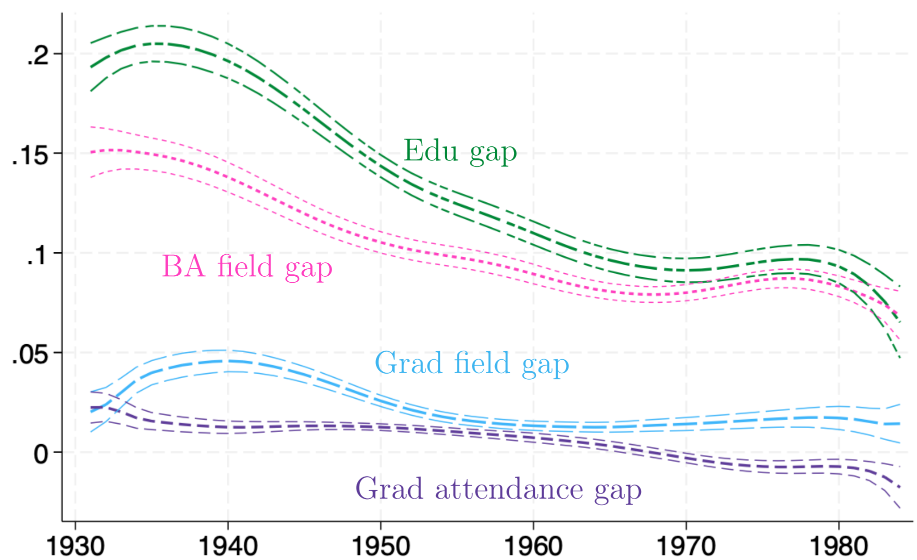
Figure E.1 shows the decompositions of the total gap and the education gap of log earnings using the female coefficients and male probabilities in the decomposition formula instead of the male coefficients and the female probabilities. The estimates are very consistent with the male coefficient version (Figure 2). Figure E.2 shows the decompositions of occupation premium using female coefficients. The estimates of the return gap are, on average, 0.01 higher than the male coefficient version (Figure 6).

Figures E.3 and E.4 are the cohort varying relative return decomposition of log earnings and occupation premium using the female coefficients and male degree probabilities. When using the female coefficients and male probabilities, the education gap in earnings is flatter than the male coefficient version and the return gap is larger in early birth cohorts. Both disparities between the two versions are caused by the varying returns. In the case of the education gap for earnings, one can see this by comparing the graph of $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ in Figure 4 panel C (yellow dashed line) with the graph of $\sum_{cg} \delta_{cg}^{fb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ in Figure E.3 panel C (yellow dashed line). In the case of the return gap, one can compare the graphs of $\sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb}$ and $\sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{mb}$ (yellow dotted line) in panel B of the two figures. See this discussion of the effects of changes across cohorts in relative returns to majors in section 5.4.

Figure E.1: Decomposition of Log Earnings, Constant Returns, Female Coefficients



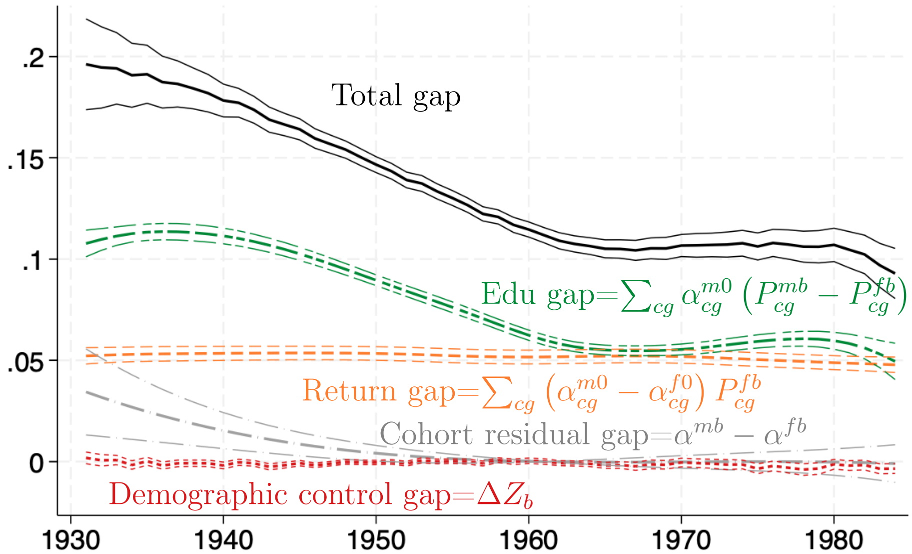
(A) Total gap



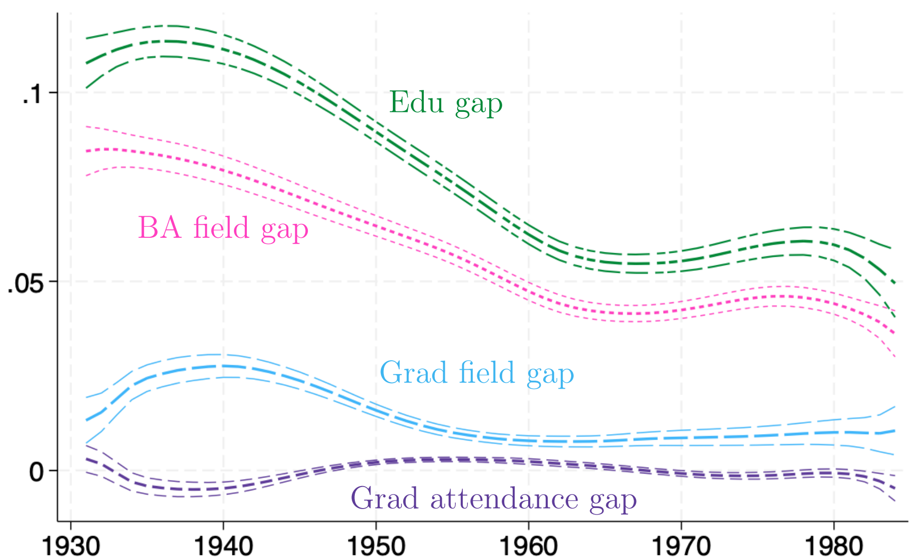
(B) Education gap

Notes: This figure shows the predicted gender gap in log earnings for each birth cohort averaged from age 28 to 52 using women's return to degrees and men's composition over degrees. The specification is the same as Figure 2.

Figure E.2: Decomposition of Occupation Premium, Constant Returns, Female Coefficients



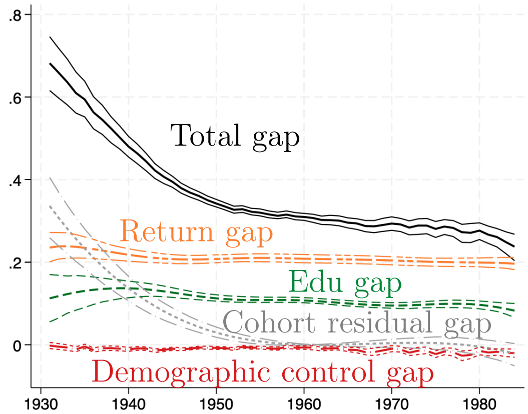
(A) Total gap



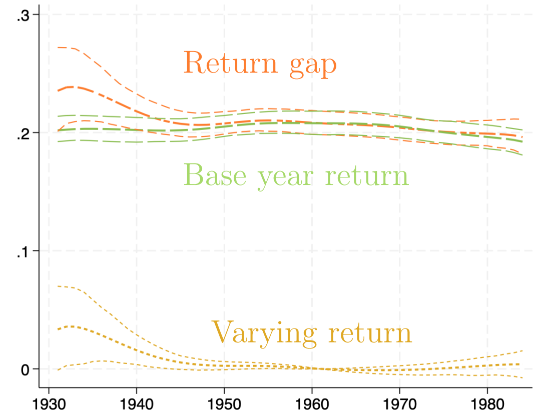
(B) education gap

Notes: This figure shows the predicted gender gap in occupation premium for each birth cohort at the average age distribution using women's return to degrees and men's composition over degrees. The specification is the same as Figure 6.

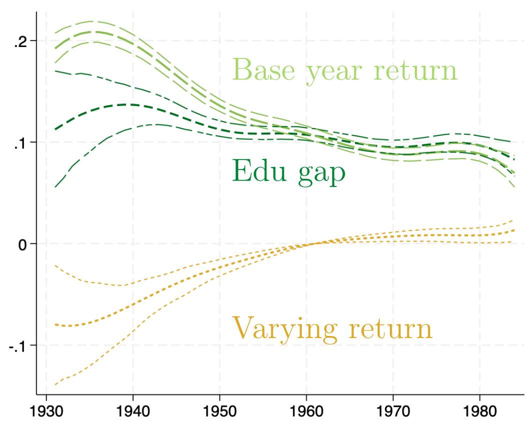
Figure E.3: Decomposition of Log Earnings, Cohort Specific Relative Returns, Female Coefficients



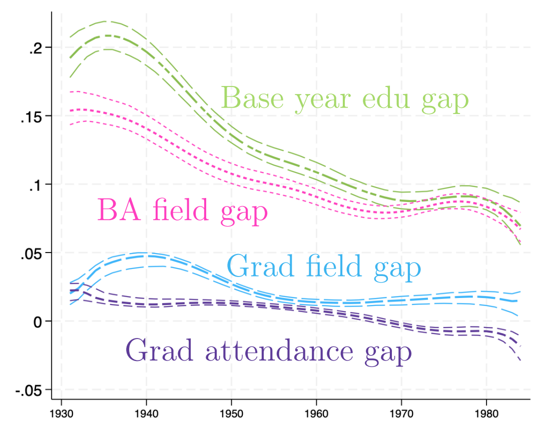
(A) Total gap



(B) The Role in the Return gap of base year and cohort varying relative returns



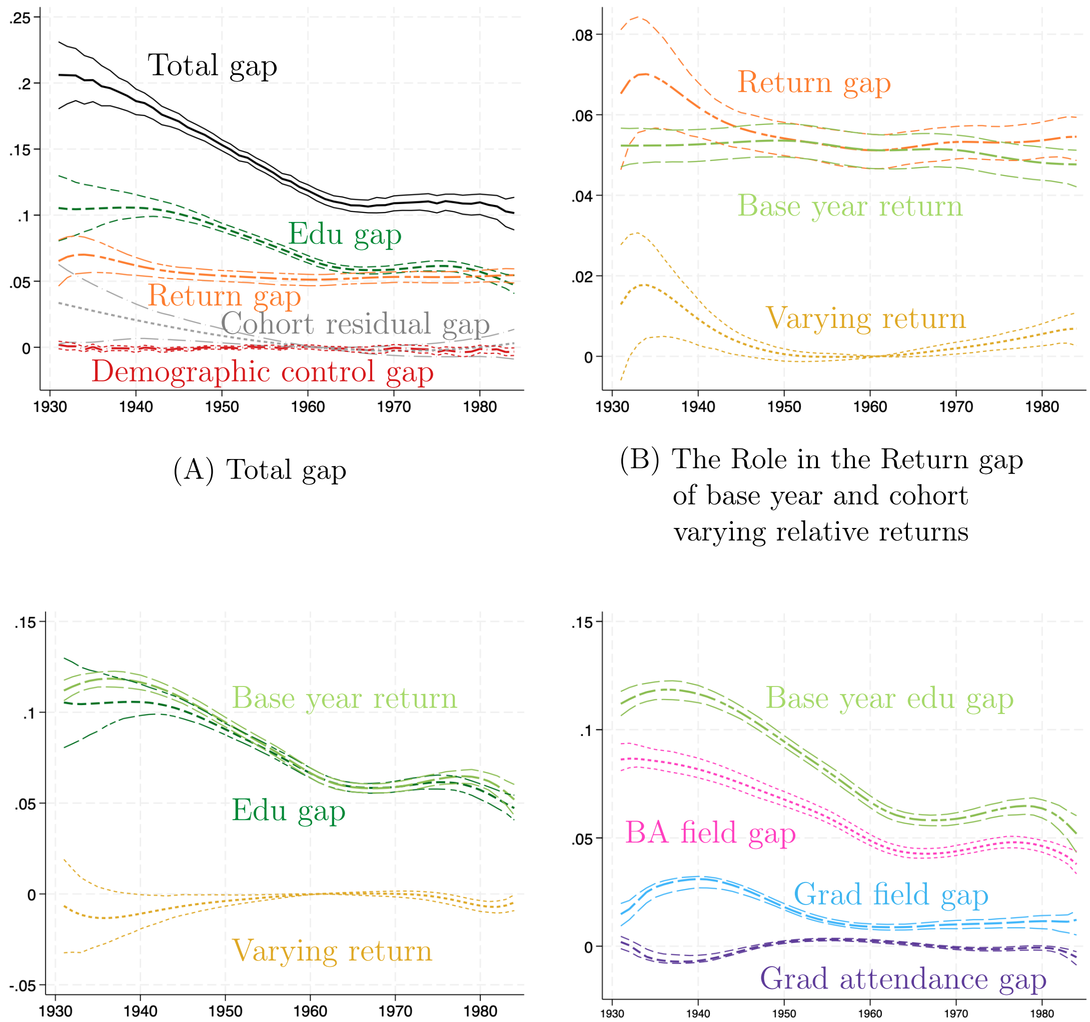
(C) The Role in the education gap of base year and cohort varying relative returns



(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: This figure shows the predicted gender gap in log earnings for each birth cohort averaged from age 28 to 52 using women's return to degrees and men's composition over degrees. The specification is the same as Figure 4.

Figure E.4: Decomposition of Occupation Premium, Cohort Specific Relative Returns, Female Coefficients



(A) Total gap

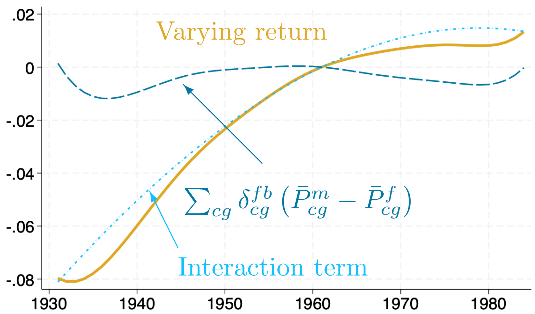
(B) The Role in the Return gap of base year and cohort varying relative returns

(C) The Role in the education gap of base year and cohort varying relative returns

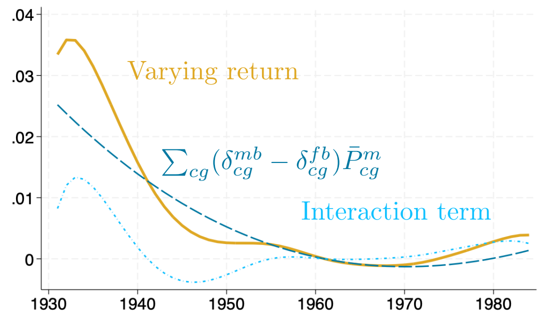
(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: This figure shows the predicted gender gap in the occupation premium for each birth cohort averaged from age 28 to 52 using women's return to degrees and men's composition over degrees. The specification is the same as Figure 8.

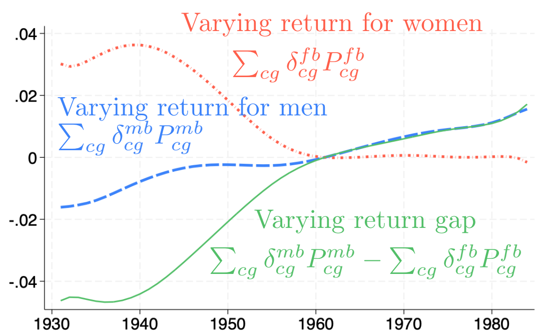
Figure E.5: Decomposition of the Varying Return Component of the Gender Gap in Log Earnings, Female Coefficients



(A) education gap, log earnings



(B) Return gap, log earnings



(C) Varying returns for men and women, log earnings

Notes: Panel A shows the education gap using the gender specific, cohort varying returns in the decomposition of log earnings. The yellow solid line is, $\sum_{cg} \delta^{fb}_{cg} \left(P^{mb}_{cg} - P^{fb}_{cg}\right)$, the same as Figure E.3 panel C. Panel B shows the return gap using the gender-specific, cohort varying relative returns in the decomposition of log earnings (yellow solid line), $\sum_{cg} \left(\delta^{mb}_{cg} - \delta^{fb}_{cg}\right) P^{mb}_{cg}$, and this line is the same as the yellow line in Figure E.3 panel B. These lines are decomposed into two components, following equation (11). For the return gap, the dark blue dashed line use the average probability for men (summand labeled on the figure). The light blue dotted line is the interaction term. Lastly, panel C presents the overall gap using the gender-specific, cohort varying relative returns in the decomposition of log earnings (green solid line) and the raw sums by gender (blue dashed for men and red dotted for women).

F Decomposition using the Census and American Community Survey Data

As a robustness check, we replicate the decomposition using the 1960-2000 decennial Census and the 2001-2018 American Community Survey. These data sources cover a much longer time period and have more balanced coverage across fields of study, but they have less information on graduate degree attainment, no information on graduate field, and no information on undergraduate field prior to the 2009 ACS. To account for the missing field of study data, we modify the earnings regression model (1) for use with the Census and ACS data. The model is

$$Y_{it} = \alpha^s_{G(i)} + X^s_{1it}\beta^s_1 + X^s_{2it}\beta^s_2 + Z^s_i\Gamma^s + u_{it}. \tag{13}$$

In this regression, $\alpha^s_{G(i)}$ is an intercept that depends on gender and whether i has a graduate degree. $G(i)$ is 0 for individuals with only a college degree and 1 for individuals with a graduate degree. The vector X_{1it} contains a gender specific cubic birth cohort polynomial and the triple interaction among gender, the graduate education dummy and a cubic age polynomial. The demographic control Z^s_i contains interactions between gender, race, and Hispanic dummies. The excluded category for men and for women is a non-Hispanic white. We can directly estimate the coefficients on the vector X^s_{2it} of interactions between b_i and age_{it} up to the second order plus $b^3_i \times a_{it}$ and $b_i \times a^3_{it}$, all interacted with $S_{s(i)}$.[39]

The graduate degree probability P^{sb}_G is estimated on the sample of people above 35 years old, and smoothed using a spline basis, then winsorized and normalized to ensure all probabilities are non-negative and sum to 1 for every combination of gender and birth year.

The gender gap can be written as

$$GAP(b) = \sum_{G=0,1} \left(P^{mb}_G \alpha^m_G - P^{fb}_G \alpha^f_G \right) + (\alpha^{mb} - \alpha^{fb}) + \Delta Z_b$$

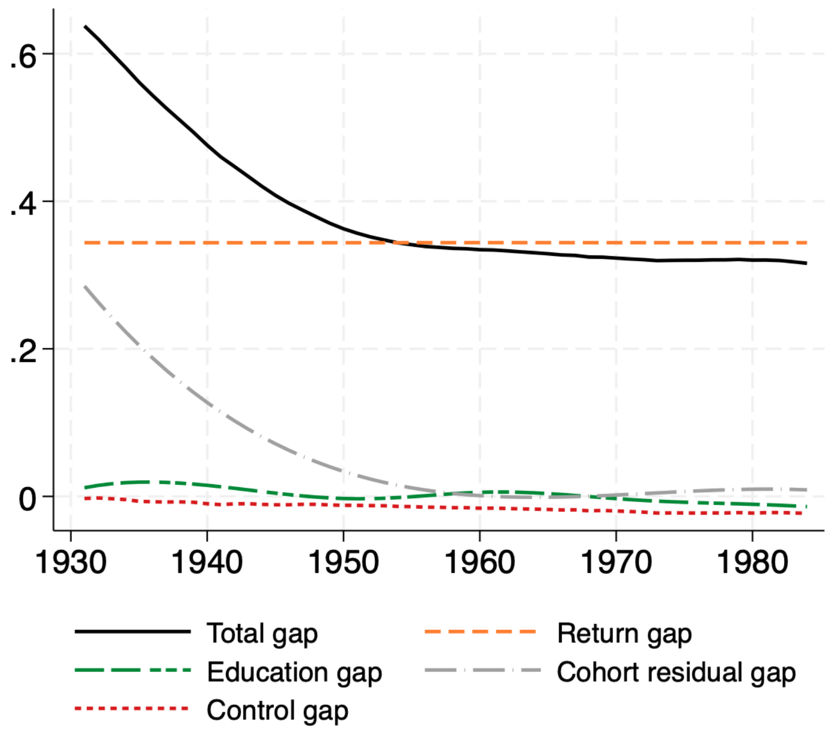
where α^{sb} is restricted to be the gender specific cubic polynomial function of birth cohort, and ΔZ_b is the demographic control gap.

Figure F.1 presents the decomposition results. The Census/ACS estimates of the total gaps for log earnings and occupation premiums are consistent with our main specification, but in the earnings case are about 0.04 higher on average for cohorts after 1974 than the NSCG estimates. The Census/ACS data has limited information on higher education history. As a result, compared with the results for the constant returns specification based on the

[39] Recall that when estimating model (1) using the NSCG we restricted β^s_2 to be proportional to the coefficient vector estimated from the Census/ACS data because of limited range of the NSCG. See Section 4.3.2.

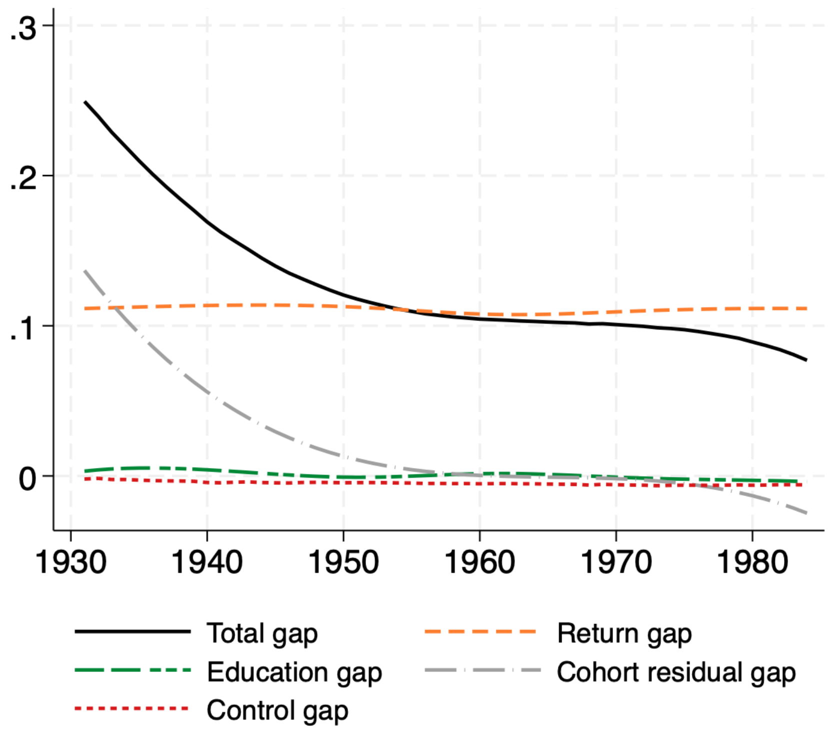
NSF data (Figures 2 and 6), the results here show larger return gaps, much smaller education gaps, and slightly larger cohort residual gaps.

Figure F.1: Decomposition of the Log Earnings Gap Using Census/ACS, Constant Returns



Notes: This Figure shows the decomposition of the predicted gender gap in log earnings for each birth cohort averaged from age 28 to 52. The black line shows the total gender log earnings gap, the orange line shows the portion of the gap at explained by the gender differences in returns to education level (graduate versus college only), the green line shows the education gap, the gray line shows the cohort residual gap that is not related to education level, and the red line shows the demographic control gap. The coefficient estimates are from regression model (13). OLS coefficients were used. The data come from Census/ACS. Ages are restricted to be between 23 and 59. By construction, Total gap = Return gap + education gap + Birth cohort residual gap + Demographic gap. The residual gap is normalized to be zero in 1961.

Figure F.2: Decomposition of the Occupation Premium Gap Using Census/ACS, Constant Returns



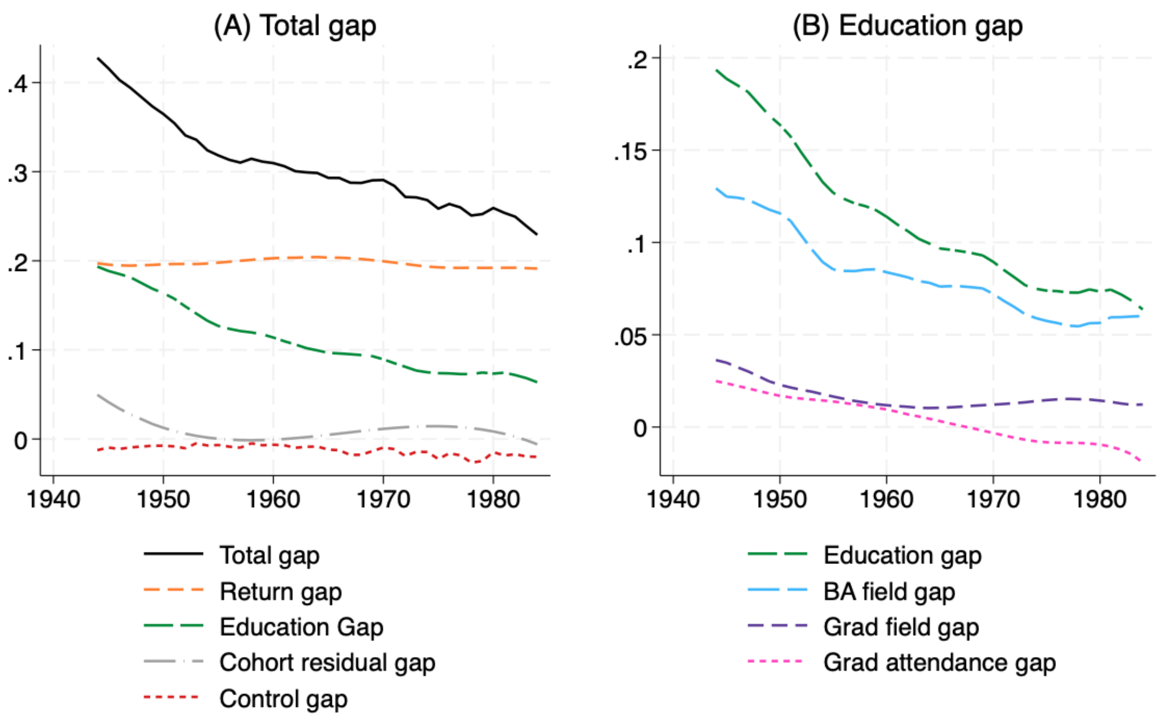
Notes: The figure shows the predicted gender gap in occupation premium for each birth cohort averaged from age 28 to 52. The data are from Census/ACS.

G Decompositions using Alternative Measures of the Undergraduate Degree Probabilities P_c^{fb} and P_c^{mb}

G.1 Using HEGIS/IPEDS to Estimate P_c^{fb} and P_c^{mb}

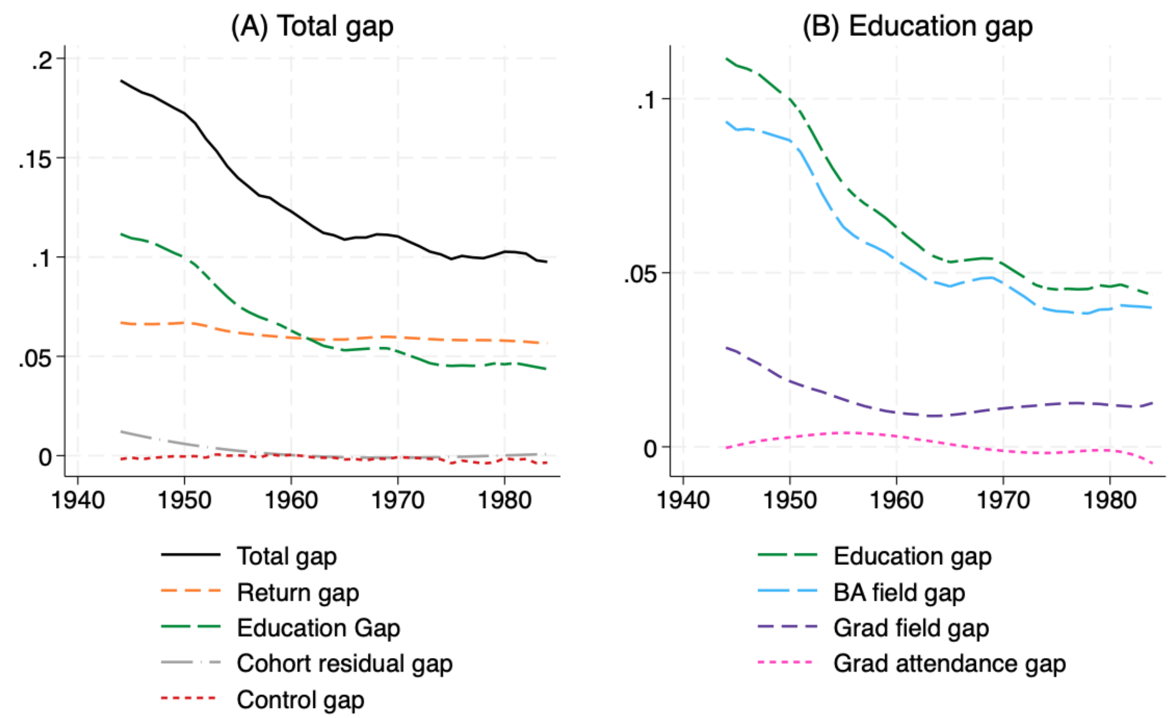
One concern is whether the NSCG data is sufficiently representative to construct the undergraduate degree probabilities P_c^{fb} and P_c^{mb}. As a check, we perform decompositions using probability estimates based on the HEGIS/IPEDS data. We assume that the age of conferral is 22 to impute birth year. Because our HEGIS/IPEDS only goes back to 1966, we are only able to re-create our decomposition back to the 1944 birth cohort. For log earnings, the decomposition, especially in the early birth cohorts, is very similar. We see some deviation from the NSCG-based decompositions starting in 1970, where use of the HEGIS/IPEDS-based estimates of P_c^{fb} and P_c^{mb} leads to an additional decrease in the total gap driven by differences in the BA field portion of the education gap. But this amounts to an average difference of only 0.015 between the two decompositions over 1971 to 1984.

Figure G.1: Decomposition for Log Earnings using $pr(c)$ from HEGIS, Constant Returns



Notes: This figure shows the decomposition results for earnings when we replace the NSCG based estimates of P_c^{fb} and P_c^{mb} with the estimates from the HEGIS/IPEDS data. We assume that the age of attainment of the undergraduate degree is 22. Panels A and B can be compared to panels A and B from Figure 2. Since HEGIS only goes back to 1966, we can only observe birth cohorts after 1944.

Figure G.2: Decomposition for Occupation Premium using P_c from HEGIS, Constant Returns

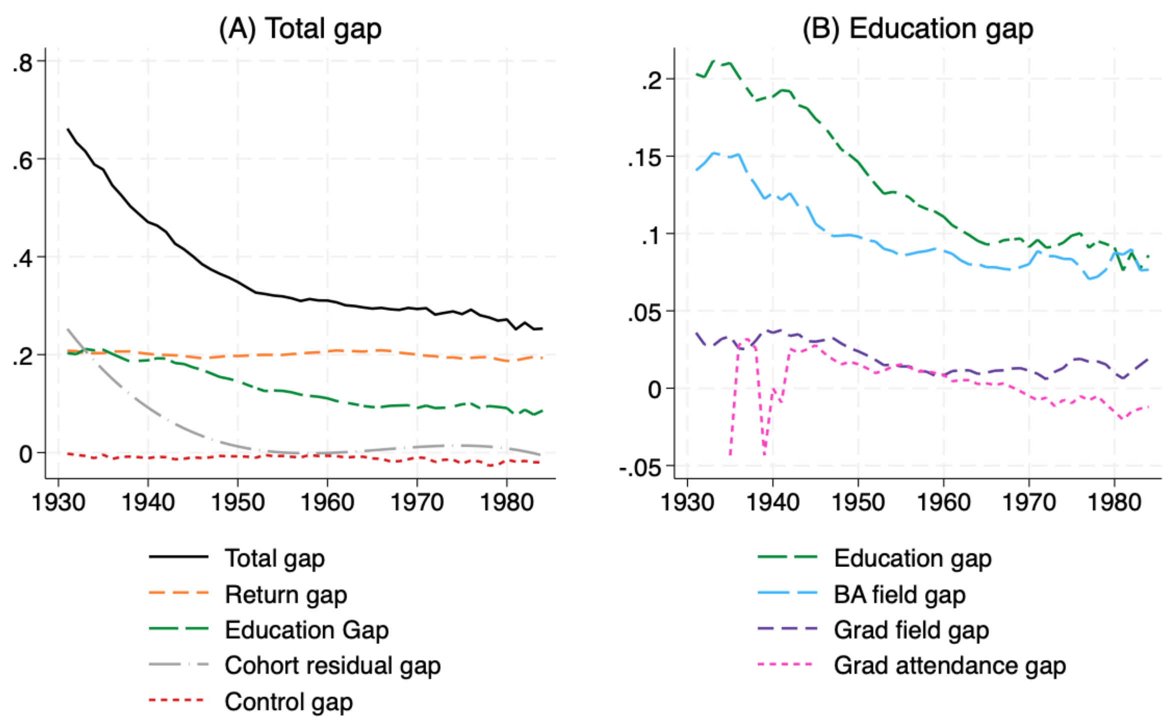


Notes: This figure shows the decomposition results for the education gap when we replace the NSCG based estimates of P_c^{fb} and P_c^{mb} with the estimates from the HEGIS/IPEDS data. We assume that the age of attainment of the undergraduate degree is 22. Panels A and B can be compared to panels A and B from Figure 6. Since HEGIS only goes back to 1966, we can only observe birth cohorts after 1944.

G.2 Using Moving Averages to Estimate Undergraduate Degree Probabilities

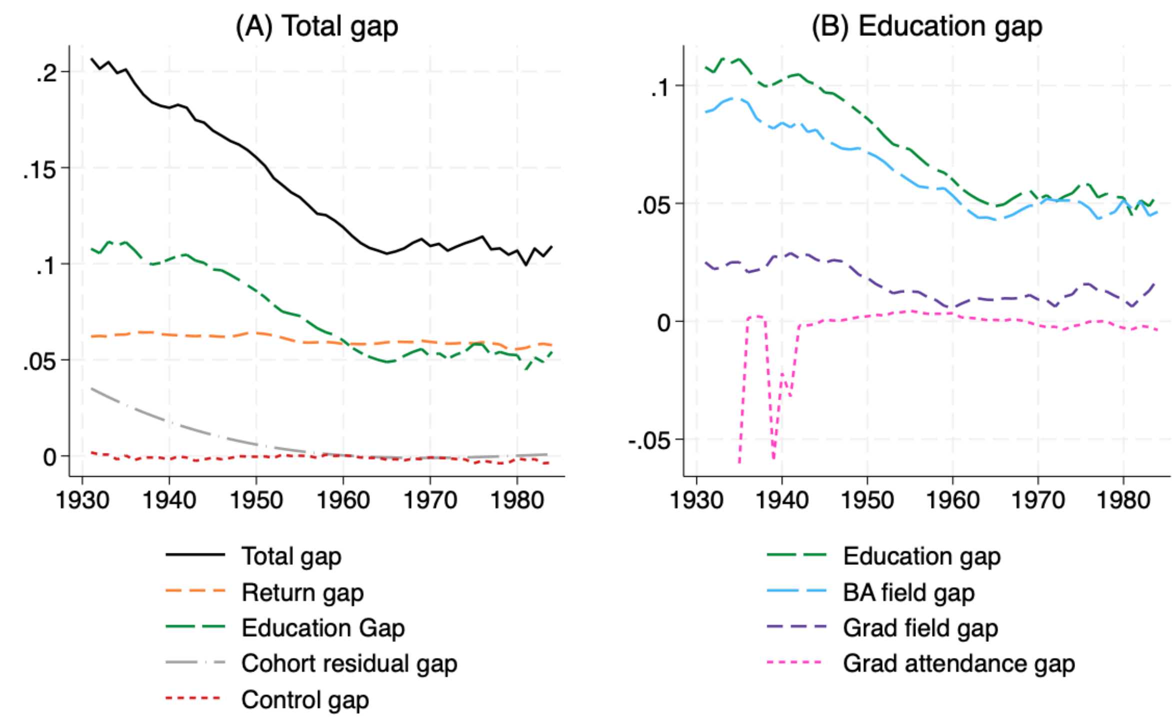
In the earnings and occupation premium decomposition we use b-splines to estimate P_c^{fb} and P_c^{mb}. It is possible that the use of splines leads to overfitting and biases the decompositions. To check on this, we replicate the decomposition using 3 year moving averages of the birth year specific and gender specific distributions of undergraduate degrees. Figures G.3 and G.4 show the decomposition with constant returns with log earnings and occupation premium as the respective dependent variable. Panel A shows the decomposition into the total gap, return gap, education gap, cohort residual gap, and control gap. The decomposition results are similar to our main results although the birth year to birth year variation is somewhat noisy. Panel B shows the decomposition of the education gap. These estimates are also similar our main results, although the estimates are noisy, particularly for the graduate attendance gap prior to 1945.

Figure G.3: Decomposition for Log Earnings using Moving Average Estimates of P_c^{fb} and P_c^{mb}, Constant Returns



Notes: This figure shows the decomposition results when we replace the estimates of P_c^{fb} and P_c^{mb} based on b-splines with the 3 year moving average of the gender and birth year specific probabilities. Panels A and B can be compared to panels A and B from Figure 2.

Figure G.4: Decomposition for Occupation Premium using Moving Average Estimates of P_c^{fb} and P_c^{mb}, Constant Returns



Notes: This figure shows the decomposition results when we replace the P_c^{fb} and P_c^{mb} b-splines with the 3 year moving average of the gender and birth year specific probabilities. Panels A and B can be compared to panels A and B from Figure 6.

H Occupation Premiums

H.1 Details on Construction of the Occupation Premiums

To construct the age specific occupation premiums $\bar{y}^a_{o(it)}$, we use the Census/ACS data to regress earnings Y_{it} on occupation fixed effects, occupation-specific age cubic polynomials, and the interaction between the female indicator F_i and an age cubic. We also include gender specific race ethnicity dummies, a graduate education dummy G_i, a cubic in birth year, and the vector X_{2it} of interactions between age and b. We then construct predicted earnings for every age and occupation, $\bar{y}^a_{o(it)}$, using only the occupation fixed effects and the occupation-specific age cubic. In the regression, we normalize the age profiles so that the intercepts for men and for women refer to the simple average between ages 28 to 52.

We construct age and birth cohort specific occupational premiums $\bar{y}^{ba}_{o(it)}$ using a similar regression procedure. The difference is that we allow occupations premiums to change across cohorts by adding occupation-birth cohort fixed effects to the regression model for Y_{it}. Due to cell size limitations, we impose additive separability between the variation in occupation premiums with b and the variation with a.[40] With this restriction, $\bar{y}^{ba}_{o(it)}$ can be written as

$$\bar{y}^{ba}_o = \bar{y}^{0a}_o + \bar{y}^b_o,$$

where $\bar{y}^{0a}_o$ is the age specific value of the occupation premium in the base year (1961), $\bar{y}^b_o$ is the cohort specific occupation component, and we have suppressed the i and t subscripts.[41]

H.2 Decomposition of the Gender Gap Using Cohort and Age Specific Occupation Premiums

We construct $X^s_{2it}\bar{\beta}^s_2$, the index of interaction terms between a_{it} and b_i in equation (3), using $\bar{y}^{ba}_{o(it)}$ as our new dependent variable (rather than $\bar{y}^a_{o(it)}$). To be specific, we substitute $\bar{y}^{ba}_{o(it)}$ for Y_{it} as the dependent variable in equation (6) and we estimate the equation in the Census/ACS data. Let $\bar{\beta}^{s*}_2$ refer to the estimates of the coefficients on X^s_{2it} in the occupation case (replacing β^{s*}_2). We impose the restriction $X^s_{2it}\bar{\beta}^s_2 = \bar{\beta}^s_3[X^s_{2it}\bar{\beta}^{s*}_2\ ]$. We estimate equation (3) allowing $\bar{\alpha}^{sb}_{cg}$ to vary across cohorts. We write $\bar{\alpha}^{sb}_{cg}$ as

$$\bar{\alpha}^{sb}_{cg} = \bar{\alpha}^{s0}_{cg} + (\bar{\alpha}^{sb} + \bar{\delta}^{sb}_{cg})$$

[40] With enough data one could regress Y_{it} on the controls and fixed effects for the interaction between birth cohort, age and occupation, and then use these fixed effects as dependent variables based on each individual's occupation, age, and birth cohort.

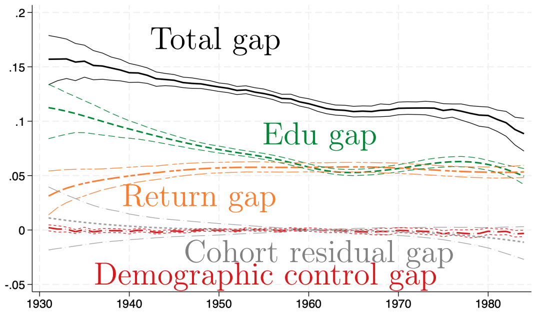
[41] We rescale $\bar{y}^{ba}_{o(it)}$ by 0.816 to address differences in earnings measures and discrepancies across the occupation measures in the Census/ACS and the NSCG. See footnote 24.

and impose normalizations and restrictions on $\bar{\alpha}^{s0}_{cg}$, $\bar{\alpha}^{sb}$ and $\bar{\delta}^{sb}_{cg}$ analogous to those imposed on α^{s0}_{cg}, α^{sb} and δ^{sb}_{cg} in the earnings case (see section 4.4).

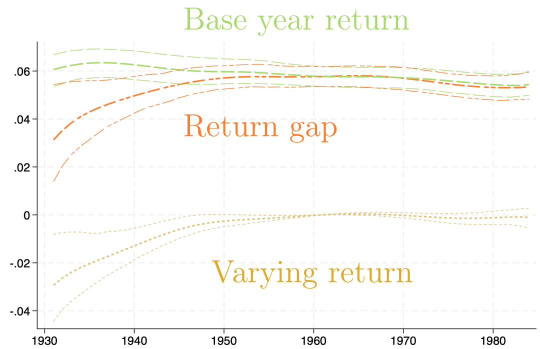
Then we perform the gender gap decompositions. We use the occupation premium counterparts of equation (10) and equation (12) (from Appendix D), replacing α^{s0}_{cg}, α^{sb} and δ^{sb}_{cg} with $\bar{\alpha}^{s0}_{cg}$, $\bar{\alpha}^{sb}$ and $\bar{\delta}^{sb}_{cg}$. The results are reported in Figure H.1. They can be be compared to Figure 8, which is based on $\bar{y}^{a}_{o}$,

Because $\bar{y}^{ba}_{o} = \bar{y}^{0a}_{o} + \bar{y}^{b}_{o}$, we can separate cohort trends in the varying return gap in $\bar{y}^{ba}_{o}$ into trends in $\bar{y}^{0a}_{o}$ and trends in $\bar{y}^{b}_{o}$. To decompose $\bar{y}^{0a}_{o}$ ($\bar{y}^{b}_{o}$), we follow the same steps used to decompose $\bar{y}^{ba}_{o}$ in Figure H.1but replace $\bar{y}^{ba}_{o}$ with $\bar{y}^{0a}_{o}$ ($\bar{y}^{b}_{o}$) as the dependent variable in equation (6) and in (3). Panel A of Figure H.2 reports the return gap decomposition of $\bar{y}^{ba}_{o}$ (long dashed light green line, graph of $\sum_{cg}\left(\bar{\alpha}^{mb}_{cg} - \bar{\alpha}^{fb}_{cg}\right)P^{fb}_{cg}$) into the base year return gap (short and long dashed dark green line, graph of $\sum_{cg}\left(\bar{\alpha}^{m0}_{cg} - \bar{\alpha}^{f0}_{cg}\right)P^{fb}_{cg}$), and the cohort varying return gap (short dashed orange line, graph of $\sum_{cg}(\bar{\delta}^{mb}_{cg}\text{-}\bar{\delta}^{fb}_{cg})P^{fb}_{cg}$). It replicates H.1 Panel B. Panel B of Figure H.2 decomposes the return gap in $\bar{y}^{0a}_{o}$. Panel C reports the gender gap decomposition of $\bar{y}^{b}_{o}$.

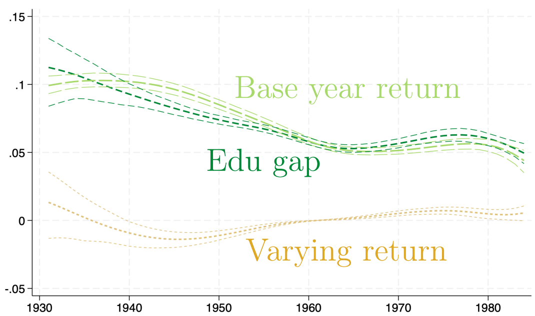
Figure H.1: Decomposition of the Gender Gap in Birth-year-age Specific Occupation Premium, Cohort Specific Relative Returns



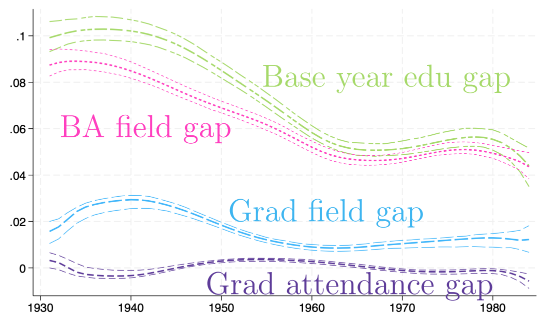
(A) Total gap



(B) Return gap



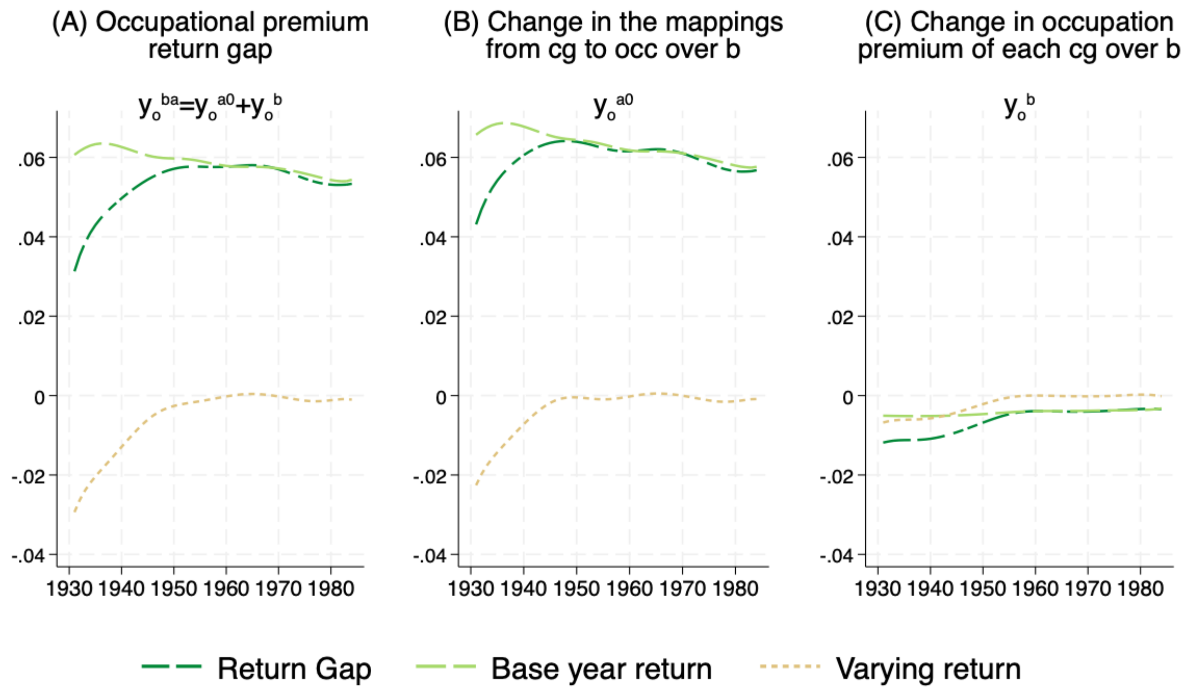
(C) Education gap



(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: This figure shows the predicted gender gap in birth year and age specific occupation premium for each birth cohort averaged from age 28 to 52. The occupation premiums correspond to the variable $\bar{y}^{ba}_{o(it)}$ and are estimated as described in section 4.5. They are used as the dependent variable in equation (3). The gender gap decomposition formulas are discussed in sections 4.4.1 and 4.5. The definitions of the lines are the same as Figure 4, but refer to the occupation premium rather than earnings.

Figure H.2: Decomposing the Varying Return Gap in the Occupation Premium



Notes: This figure decomposes the return component of the occupational premium gap (long dashed light green line) into base year returns (dark short and long dashed green line) and varying returns (short dashed orange line). Panel A is the same as Figure 8 panel B. In panels B and C of this figure, we replace the dependent variable $\bar{y}_o^{ba}$ in the regression model (3) with $\bar{y}_o^{0a}$ and $\bar{y}_o^{b}$, respectively. As defined in section 4.5, $\bar{y}_o^{ba} = \bar{y}_o^{0a} + \bar{y}_o^{b}$. The lines are defined in the same way as panel A. By construction of the dependent variable, the lines in panels B and C sum to the corresponding lines in panel A. In all three panels, the base year return line is the graph of $\sum_{cg}\left(\bar{\alpha}_{cg}^{m0} - \bar{\alpha}_{cg}^{f0}\right) P_{cg}^{fb}$ and the varying relative return line is the graph of $\sum_{cg}\left(\bar{\delta}_{cg}^{mb} - \bar{\delta}_{cg}^{fb}\right) P_{cg}^{fb}$, but the values of $\bar{a}_{cg}^{m0} - \bar{a}_{cg}^{f0}$ and $\bar{\delta}_{cg}^{mb} - \bar{\delta}_{cg}^{fb}$ depend on the dependent variable for the regression model that is used in the panel. The y-axis refers to the gender gap in log points and is consistent across panels.

I Decomposition of the Earnings and Occupation Gender Gaps by Field of Study

In this appendix, we disaggregate the gender gap decompositions by college majors. For all figures in this appendix, the color scheme and the definition of the lines are consistent with the main text. Similar to Figure 3, the education gaps for Figures I.1 to I.8 plot $\sum_g (\alpha_{cg}^{m0} - \alpha^{m*})(P_{cg}^{mb} - P_{cg}^{fb})$ for each major c, where α^{m*} is the average of the earnings intercepts α_{cg}^{m0} weighted by average of the c,g probabilities for men and women. Specifically, $\alpha^{m*} \equiv \sum_{cg} \alpha_{cg}^{m0} \bar{P}_{cg}$, where $\bar{P}_{cg}$ is the average probability for each cg combination across sex and birth cohorts. See the note of Figure 3 for a detailed explanation. The majors are ordered alphabetically. We will now highlight the primary contributing majors for each decomposed gap.

Figure I.1 disaggregates the return gap and the education gap in Figure 2. We do not show the demographic control gap and the cohort residual gap, as they are common components unrelated to college majors. The green solid line is the sum of the education gap for a given college major. The short and long orange dashed line is the sum of the return gap for a given college major. The sums for a given major are over all graduate fields plus the category of never attended graduate school (BA only). The black dashed line is the sum of the education and return gap for a given college major. It does not include the birth cohort fixed effect or the demographic controls. The y-axis of each panel (in log points) presents the size of the education gap and return gap and the sum of the two. Note that the scale varies across majors.

For the sum of the education and return gaps, Education and English, Language and Literature contribute the most to the decreasing trend. Fine Arts, Health, Nursing, Other Social Sciences, and Physical Sciences also contribute. CS/Math and Psychology and Social Work were the most important contributors to an increase in the gap. We already discussed the education gaps by major in Section 5.1.1. One can see that for a few majors the increases or decreases in the return gap across cohorts are substantial. This is true even though Figure 2 shows that the sum across majors of the return gap varies very little across cohorts.

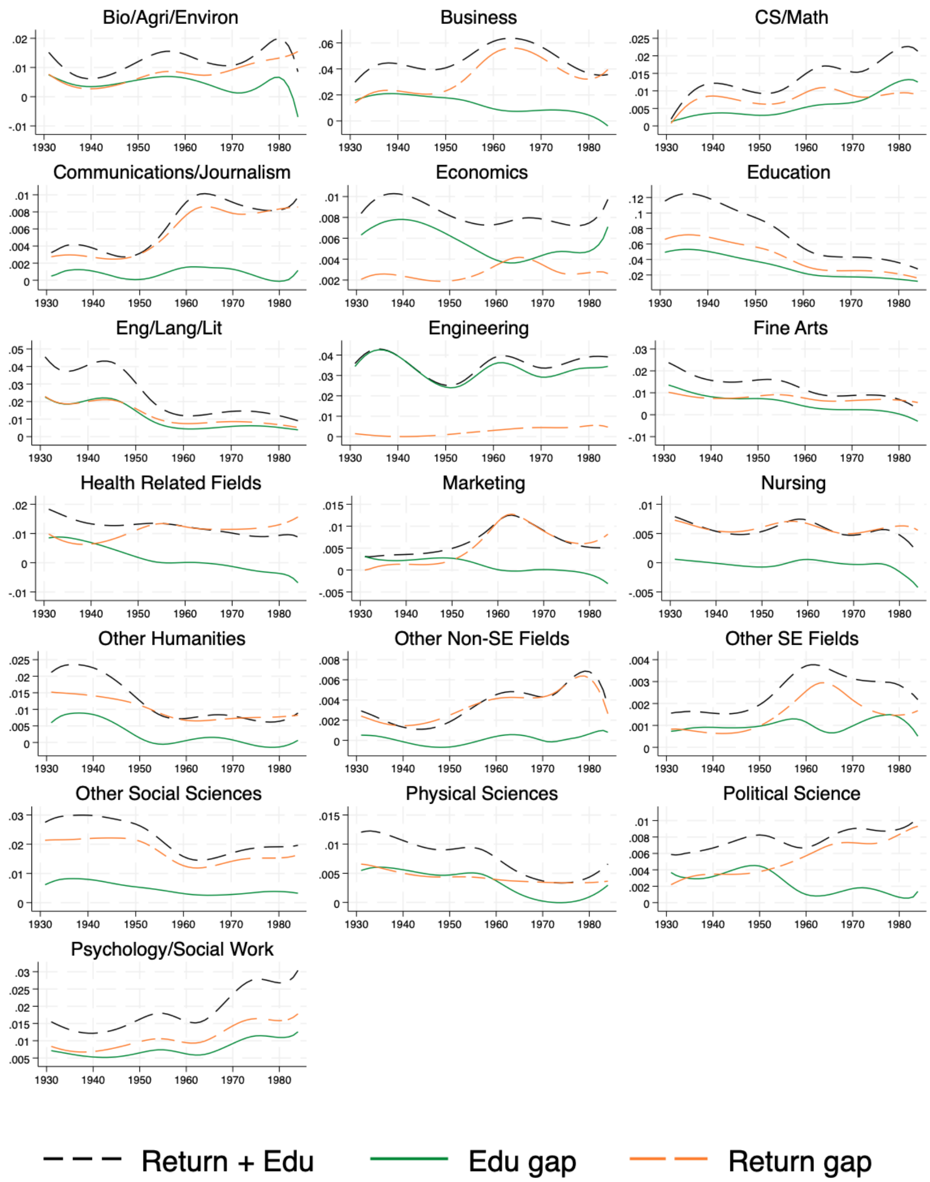
Figure I.2 shows the disaggregation by major using the female coefficients and male composition. It shows the same general patterns as the male coefficient version, although there are some differences. For example, the relative return gap for the business major increases across cohorts when we weight the difference in returns by the female cg probabilities and it declines when we weight using the male cg probabilities (see Figure I.2). Using the female probabilities, the positive gender difference in the relative return to business grows in importance as women move into business after the 1950 birth cohort. Using the male probabilities, the gender difference in relative return to business matters less, because male business enrollments trend down after the late 1950s birth cohorts.

Figure I.3 disaggregates the return gap and the education gap by college major for the constant occupation premium specification in Figure 6. Education, English Language and Literature, Humanities, Fine Arts, and Business contribute the most to the decline in sum of the return and education gaps. CS/Math and Psychology and Social Work are the most important contributors to an increase in the gap.

Figures I.5 to I.8 present the decomposition by college major for our dynamic returns specifications. Figure I.5 disaggregates the dynamic log earnings lines in Figure 4 and Figure I.6 does the same using female coefficients. Figure I.7 disaggregates the dynamic occupation premium results in Figure 8, and Figure I.8 does the same using female coefficients. The contributors are the same as the constant returns specification.

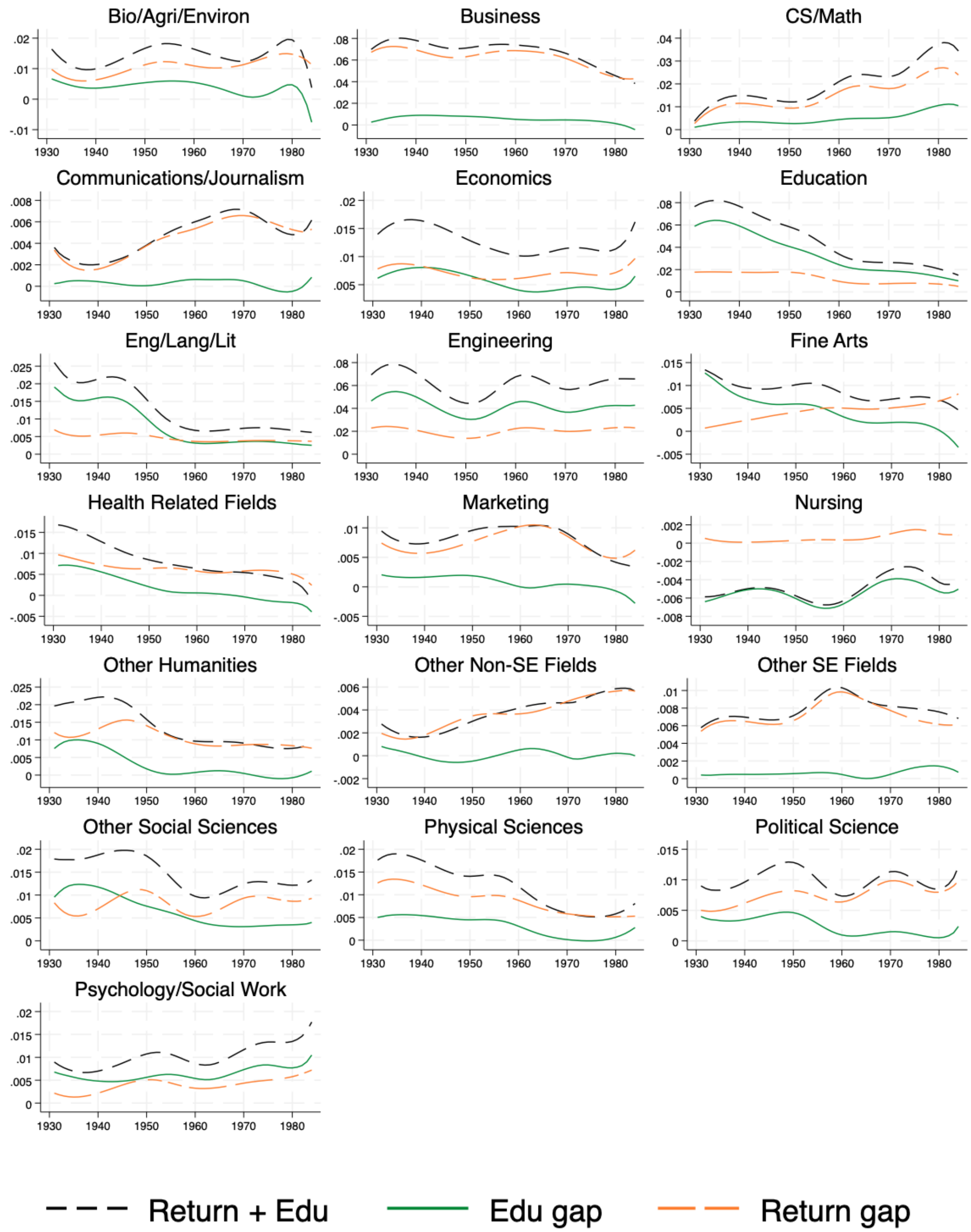
Lastly, Figure I.9 presents the contribution of each college major to the varying returns gaps of the log earnings respectively. Figure I.10 does so for the varying return component of the education gap, $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$. Both figures use the male coefficients.

Figure I.1: Decomposition of Log Earnings Gap by College Major, Constant Returns, Male Coefficients



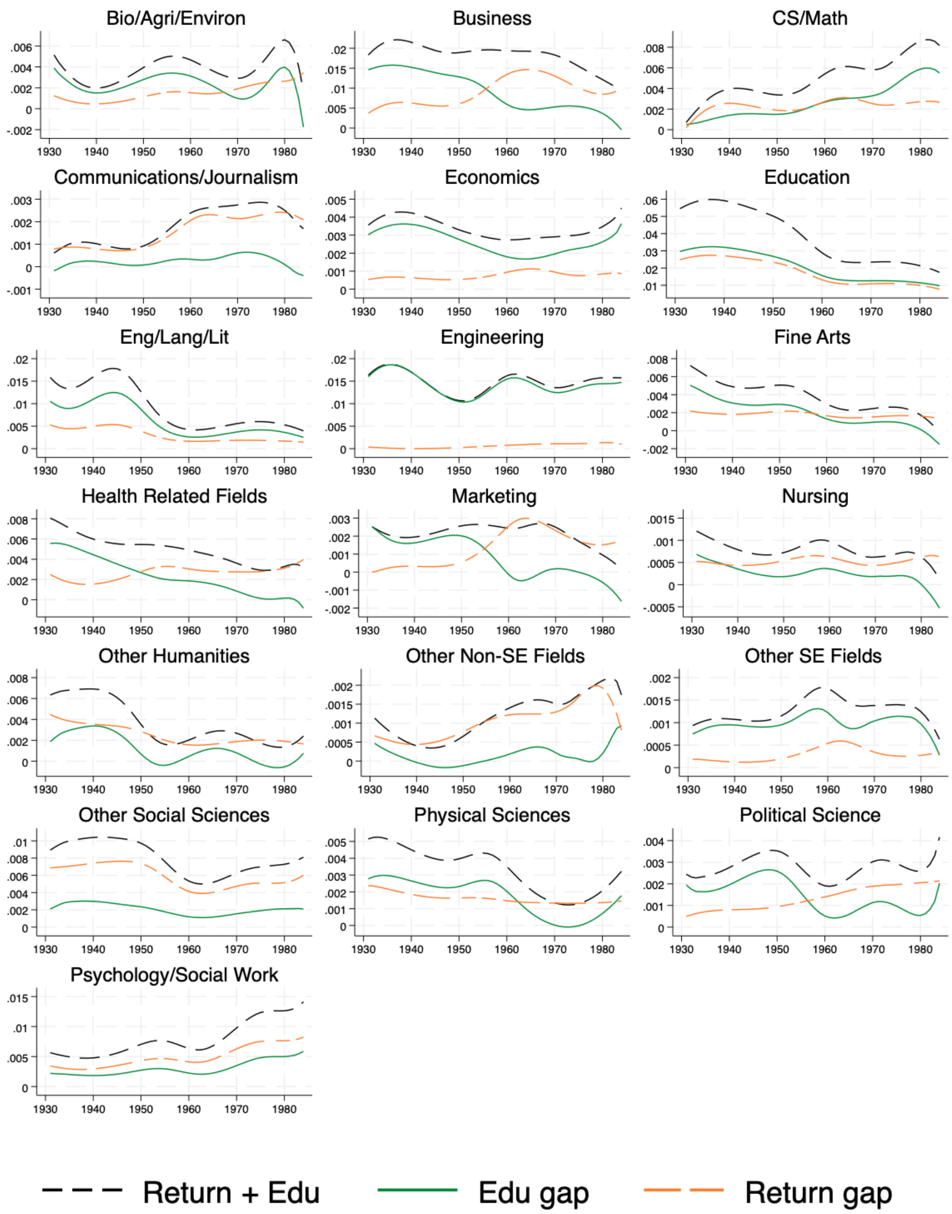
Notes: This figure shows the contribution of each college major towards the gaps shown in Figure 2, which presents the decomposition of log earnings using the constant returns specification. The green line is the contribution to the education gap for a given college major. As in Figure 3, the green line plots $\sum_{g=0}^{G}\left(\alpha_{cg}^{m0}-\alpha^{m*}\right)\left(P_{cg}^{mb}-P_{cg}^{fb}\right)$ for each major c, where α^{m*} is the average of the earnings intercepts α_{cg}^{m0} weighted by average of the c,g probabilities for men and women. The orange short and long dashed line is the sum of the return gap for a given college major. It plots $\sum_{g=0}^{G}\left(\alpha_{cg}^{m0}-\alpha_{cg}^{f0}\right)P_{cg}^{fb}$. The sums are over all graduate fields plus the category of never attended graduate school (BA only). The black dashed line is the sum of the education and return gap. It does not include the birth cohort fixed effect or the demographic controls. The y-axis of each panel (in log points) presents the size of the education gap, the return gap, and the sum of the two. Note that the scale varies across majors.

Figure I.2: Decomposition of Log Earnings Gap by College Major, Constant Returns, Female Coefficients



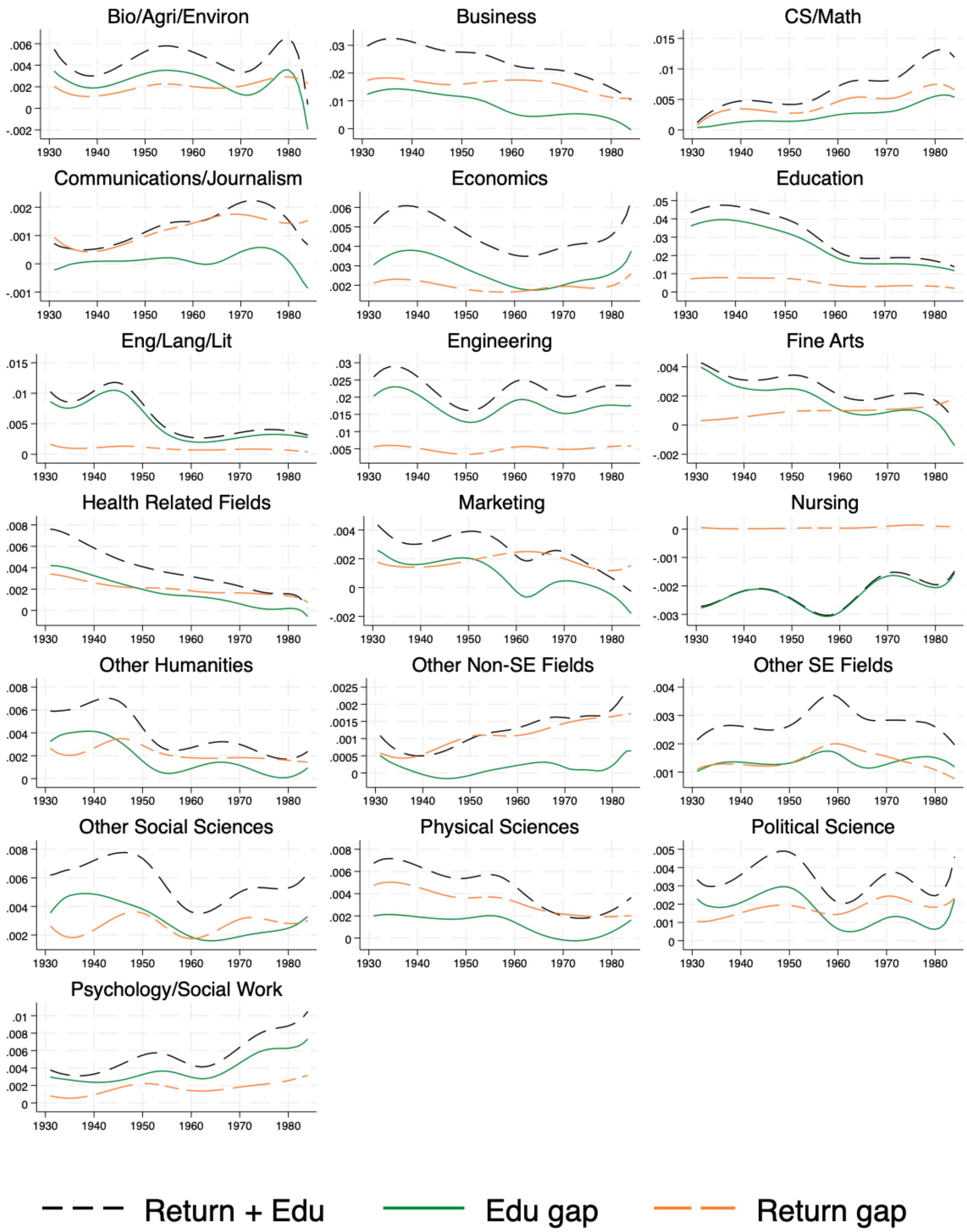
Notes: This figure shows the contribution by birth cohort of each college major towards the gaps shown in Figure E.1. The vertical axis is in log points. For the definition of the lines, see the notes to Figure I.1, except that female coefficients are used for the education gap and male *cg* probabilities are used for the relative return gap.

Figure I.3: Decomposition of Occupation Premium Gap by College Major, Constant Returns, Male Coefficients



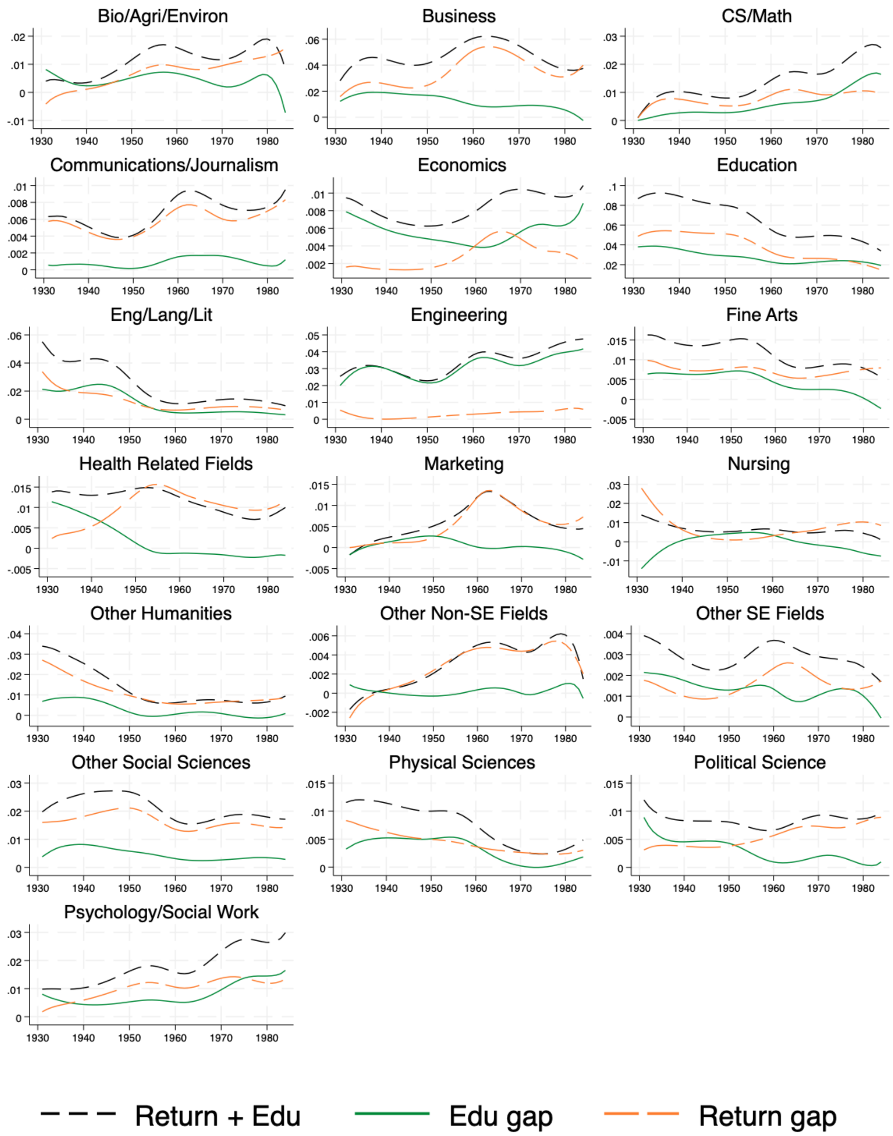
Notes: This figure shows the contribution by birth cohort of each college major towards the occupation premium gaps shown in Figure 6. The vertical axis is in log points. The green line plots $\sum_{g=0}^{G} \left(\bar{\alpha}_{cg}^{m0} - \bar{\alpha}^{m*}\right)\left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ for each major c, where $\bar{\alpha}^{m*}$ is the average of the occupation premium intercepts $\bar{\alpha}_{cg}^{m0}$ weighted by average of the c,g probabilities for men and women. The orange short and long dashed line is $\sum_{g=0}^{G} \left(\bar{\alpha}_{cg}^{m0} - \bar{\alpha}_{cg}^{f0}\right) P_{cg}^{fb}$, the sum of the return gap for a given college major. The sums are over all graduate fields plus the category of never attended graduate school (BA only). The black dashed line is the sum of the education and return gap. It does not include the birth cohort fixed effect or the demographic controls. The y-axis of each panel (in log points) presents the size of the education gap and return gap and the sum of the two. Note that the scale varies across majors.

Figure I.4: Decomposition of Occupation Premium Gap by College Major, Constant Returns, Female Coefficients



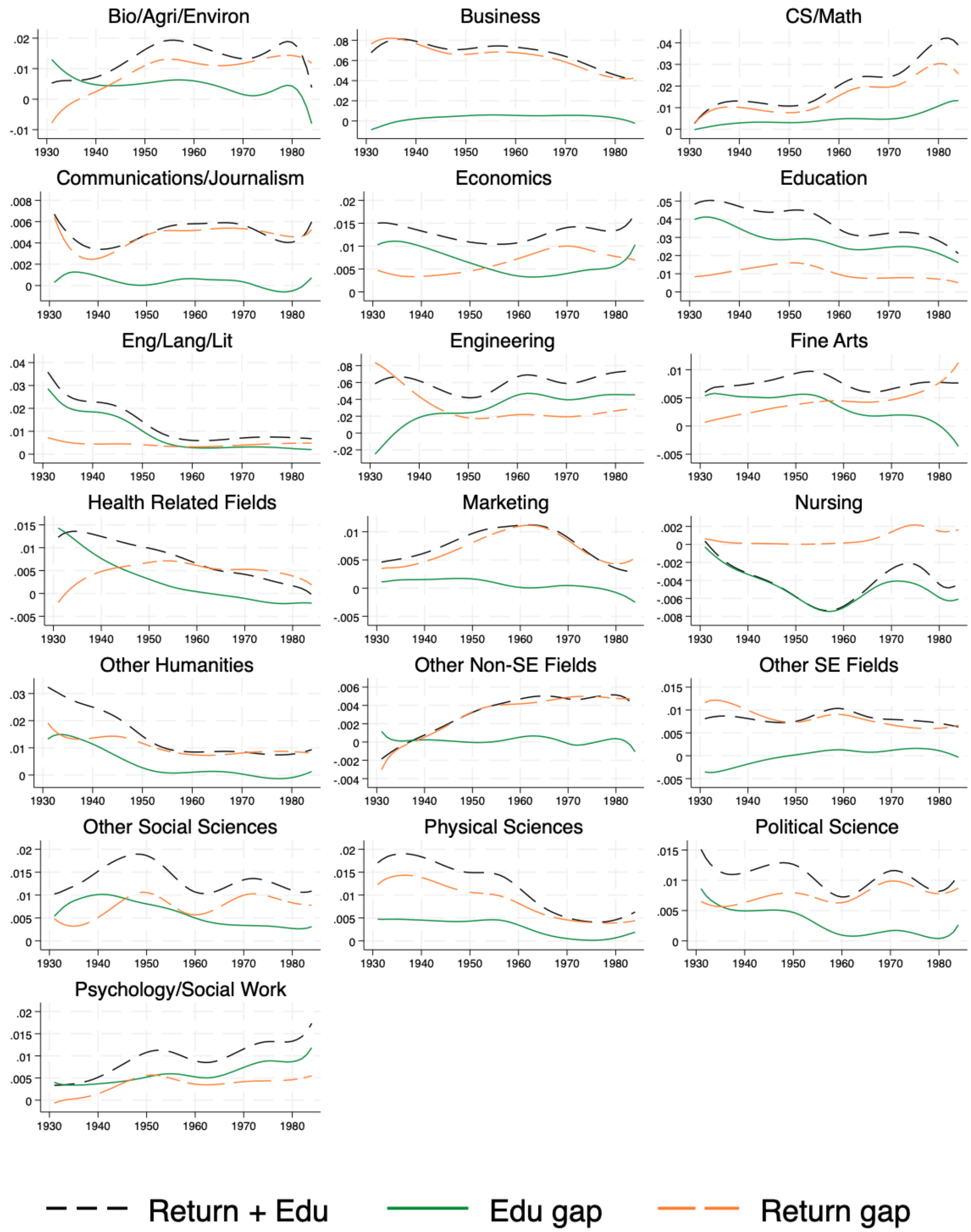
Notes: This figure shows the contribution by birth cohort of each college major to the occupation premium gaps shown in Figure E.2 The vertical axis is in log points. For the definition of the lines, see the notes to Figure I.3, except that we use female coefficients for the education gap and male probabilities are used for the relative return gap.

Figure I.5: Decomposition of Log Earnings Gap by College Major, Dynamic Returns, Male Coefficients



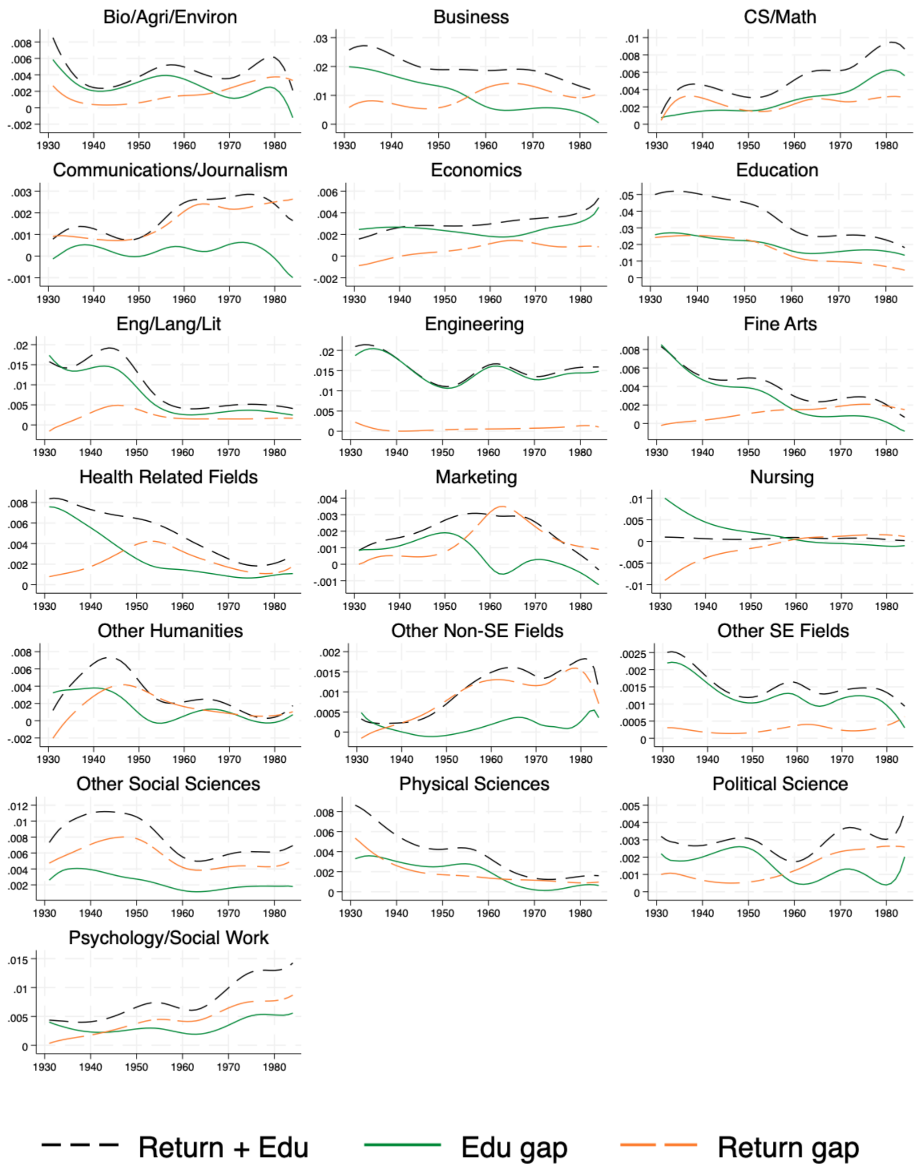
Notes: This figure shows the contribution by birth cohort of each college major towards the gaps shown in Figure 4, which presents the decomposition of log earnings using the cohort varying relative returns specification. The green line is the contribution to the education gap for a given college major. The green line plots $\sum_g \left(\alpha_{cg}^{mb} - \alpha^{mb*}\right)\left(P_{cg}^{mb} - P_{cg}^{fb}\right)$ for each major c, where α^{mb*} is the average of the earnings intercepts α_{cg}^{mb} weighted by $\frac{1}{2}\left(\bar{P}_{cg}^{f} + \bar{P}_{cg}^{m}\right)$, which is the average of the c,g probabilities for men and women. The orange short and long dashed line plots $\sum_g \left(\alpha_{cg}^{mb} - \alpha_{cg}^{fb}\right) P_{cg}^{fb}$. It is the sum of the return gap for a given college major. The sums are over all graduate fields plus the category of never attended graduate school (BA only). The black dashed line is the sum of the education and return gap. It does not include the birth cohort fixed effect or the demographic controls. The y-axis of each panel (in log points) presents the size of the education gap and return gap and the sum of the two. Note that the scale varies across majors.

Figure I.6: Decomposition of Log Earnings Gap by College Major, Dynamic Returns, Female Coefficients



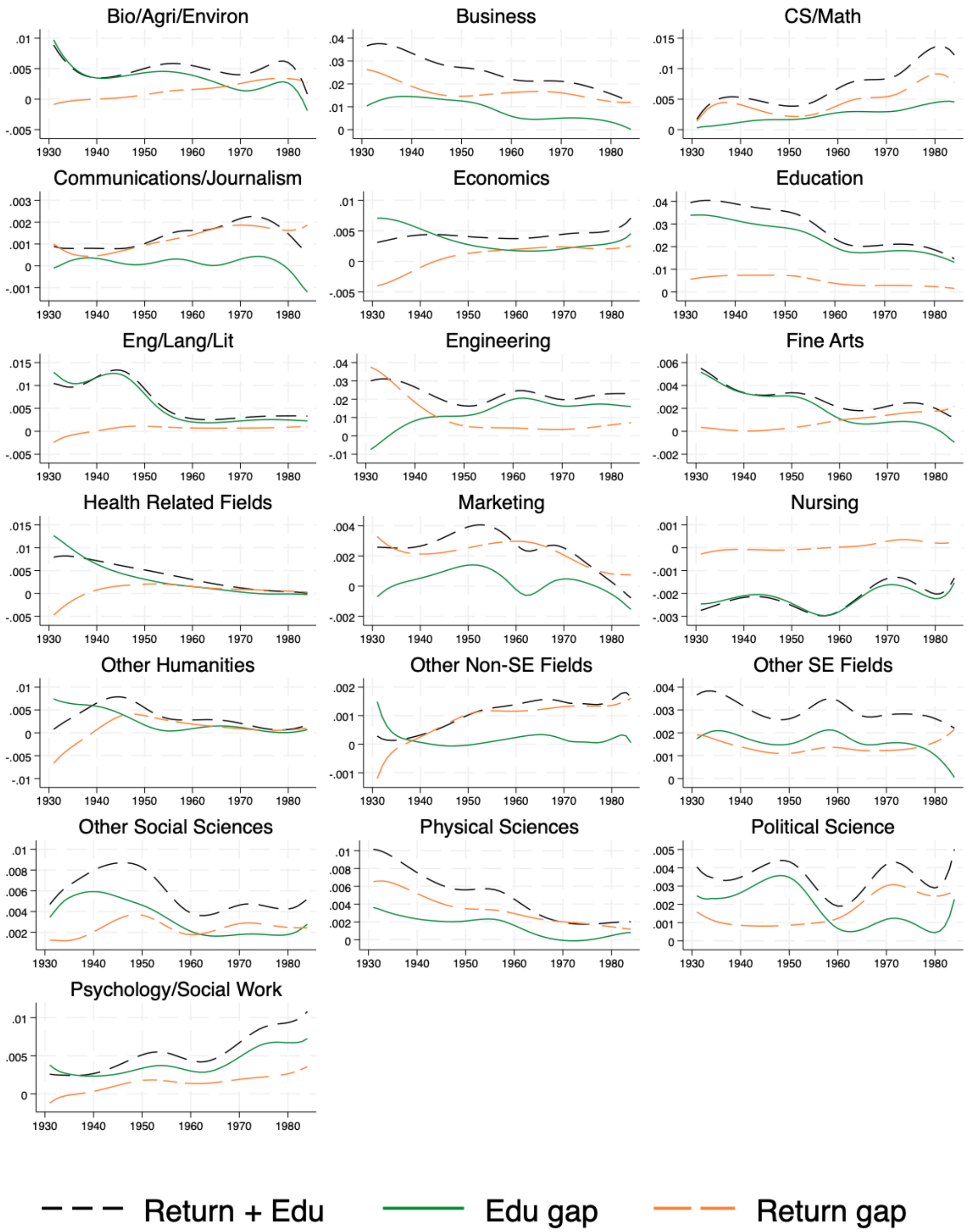
Notes: This figure shows the contribution by birth cohort of each college major towards the gaps shown in Figure E.3. It is based on the cohort varying relative returns specification. The vertical axis is in log points. For the definition of the lines, see the notes to Figure I.5, except that female coefficients are used for the education gap and male probabilities are used for the relative return gap.

Figure I.7: Decomposition of Occupation Premium Gap by College Major, Dynamic Returns, Male Coefficients



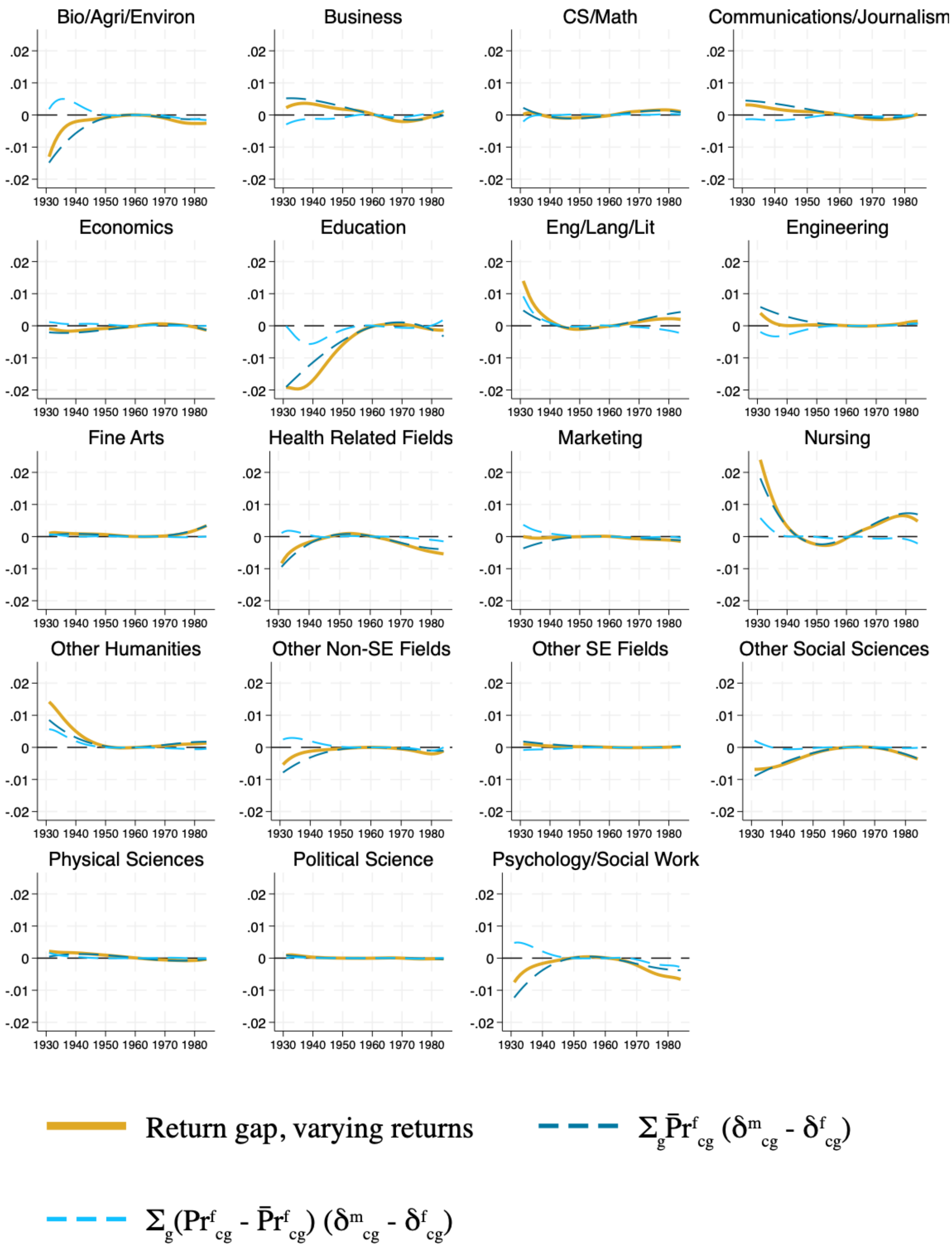
Notes: This figure shows the contribution of each college major towards the gaps shown in Figure 8. The green line is the contribution to the education gap in the occupation premium of a given college major. The green line plots $\sum_g \left(\bar{\alpha}^{mb}_{cg} - \bar{\alpha}^{mb*}\right)\left(P^{mb}_{cg} - P^{fb}_{cg}\right)$ for each major c, where $\bar{\alpha}^{mb*}$ is the average of the earnings intercepts α^{mb}_{cg} weighted by $\frac{1}{2}\left(\bar{P}^{f}_{cg} + \bar{P}^{m}_{cg}\right)$, which is the average of the c,g probabilities for men and women. The orange short and long dashed line plots $\sum_g \left(\bar{\alpha}^{mb}_{cg} - \bar{\alpha}^{fb}_{cg}\right) P^{fb}_{cg}$ for each major c, where $\bar{\alpha}^{mb*}$. It is the sum of the return gap for a given college major. The sums are over all graduate fields plus the category of never attended graduate school (BA only). The black dashed line is the sum of the education and return gap. It does not include the birth cohort fixed effect or the demographic controls. The y-axis of each panel (in log points) presents the size of the education gap and return gap and the sum of the two. Note that the scale varies across majors.

Figure I.8: Decomposition of Occupation Premium by College Major, Dynamic Returns, Female Coefficients



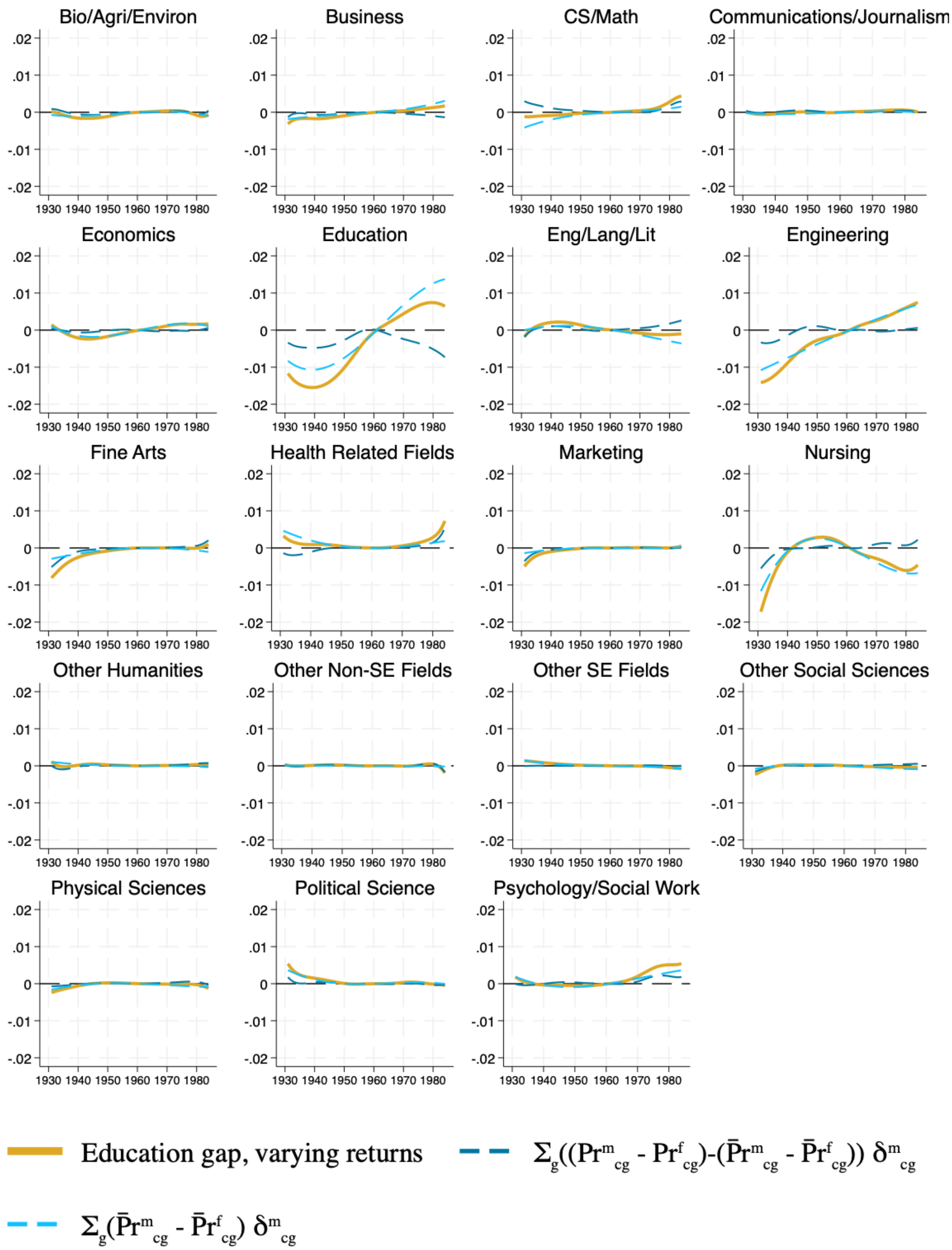
Notes: This figure shows the contribution of each college major towards the gaps shown in Figure E.4. It is the same as Figure I.7 except that female coefficients are used for the education gap and male probabilities are used for the relative return gap.

Figure I.9: Decomposing the Varying Return Gap for Log Earnings by College Major, Male Coefficient



Notes: This figure shows the contribution of each college major towards the gaps shown in Figure 5 panel B. The solid orange line is the sum of the light blue and dark blue dashed lines.

Figure I.10: Decomposing the Varying Return Component of the Education Gap for Log Earnings by College Major, Male Coefficient



Notes: This figure shows the contribution of each college major towards the gaps in the decomposition of the varying return component of the education gap shown in shown in Figure 5 panel A.

J Decomposition Tables

Table J.1: Constant Decomposition: Log Earnings

Birth Cohort	Total Gap	Cohort Residual Gap	Return Gap	Education Gap	Demo Control Gap
1931	0.645 (0.031)	0.253 (0.031)	0.202 (0.006)	0.193 (0.006)	-0.002 (0.004)
1932	0.627 (0.028)	0.230 (0.028)	0.204 (0.006)	0.198 (0.006)	-0.005 (0.004)
1933	0.609 (0.025)	0.208 (0.025)	0.206 (0.006)	0.202 (0.005)	-0.007 (0.004)
1934	0.587 (0.022)	0.188 (0.023)	0.207 (0.005)	0.204 (0.005)	-0.011 (0.003)
1935	0.577 (0.020)	0.169 (0.020)	0.207 (0.005)	0.205 (0.005)	-0.004 (0.003)
1936	0.550 (0.018)	0.151 (0.018)	0.207 (0.005)	0.205 (0.004)	-0.013 (0.003)
1937	0.536 (0.016)	0.134 (0.017)	0.206 (0.005)	0.204 (0.004)	-0.009 (0.002)
1938	0.518 (0.014)	0.119 (0.015)	0.205 (0.005)	0.202 (0.004)	-0.008 (0.002)
1939	0.499 (0.012)	0.105 (0.013)	0.204 (0.005)	0.199 (0.004)	-0.009 (0.002)
1940	0.479 (0.011)	0.092 (0.012)	0.203 (0.005)	0.196 (0.004)	-0.011 (0.002)
1941	0.465 (0.010)	0.080 (0.011)	0.201 (0.005)	0.192 (0.004)	-0.009 (0.002)
1942	0.447 (0.009)	0.069 (0.010)	0.199 (0.005)	0.188 (0.004)	-0.009 (0.002)
1943	0.426 (0.008)	0.059 (0.009)	0.198 (0.005)	0.183 (0.004)	-0.014 (0.002)
1944	0.411 (0.007)	0.050 (0.008)	0.197 (0.005)	0.177 (0.004)	-0.013 (0.002)
1945	0.399 (0.006)	0.041 (0.007)	0.195 (0.005)	0.172 (0.003)	-0.010 (0.002)
1946	0.384 (0.006)	0.034 (0.007)	0.195 (0.005)	0.166 (0.003)	-0.011 (0.002)
1947	0.373 (0.005)	0.028 (0.006)	0.195 (0.005)	0.160 (0.003)	-0.009 (0.002)
1948	0.364 (0.005)	0.022 (0.006)	0.195 (0.005)	0.154 (0.003)	-0.007 (0.002)
1949	0.353 (0.005)	0.017 (0.005)	0.195 (0.005)	0.149 (0.003)	-0.008 (0.002)
1950	0.345 (0.005)	0.013 (0.005)	0.196 (0.005)	0.144 (0.003)	-0.008 (0.002)
1951	0.337 (0.004)	0.009 (0.004)	0.197 (0.005)	0.139 (0.003)	-0.008 (0.002)
1952	0.328 (0.004)	0.006 (0.004)	0.198 (0.005)	0.135 (0.003)	-0.010 (0.002)
1953	0.329 (0.004)	0.004 (0.003)	0.199 (0.004)	0.131 (0.003)	-0.005 (0.002)
1954	0.322 (0.004)	0.002 (0.003)	0.200 (0.004)	0.127 (0.003)	-0.007 (0.002)
1955	0.319 (0.004)	0.000 (0.003)	0.201 (0.004)	0.125 (0.003)	-0.007 (0.002)
1956	0.315 (0.004)	0.000 (0.002)	0.202 (0.004)	0.122 (0.003)	-0.009 (0.002)
1957	0.311 (0.004)	-0.001 (0.002)	0.203 (0.004)	0.119 (0.003)	-0.010 (0.002)
1958	0.314 (0.004)	-0.001 (0.001)	0.204 (0.004)	0.116 (0.003)	-0.005 (0.002)
1959	0.310 (0.004)	-0.001 (0.000)	0.205 (0.004)	0.113 (0.003)	-0.007 (0.002)
1960	0.309 (0.004)	0.000 (0.000)	0.206 (0.004)	0.110 (0.003)	-0.006 (0.002)
1961	0.307 (0.005)	0 (0.000)	0.207 (0.004)	0.107 (0.003)	-0.007 (0.002)
1962	0.302 (0.005)	0.000 (0.000)	0.207 (0.004)	0.104 (0.002)	-0.010 (0.002)
1963	0.302 (0.005)	0.002 (0.000)	0.207 (0.004)	0.101 (0.002)	-0.009 (0.002)
1964	0.301 (0.005)	0.003 (0.001)	0.207 (0.004)	0.098 (0.002)	-0.008 (0.002)
1965	0.296 (0.006)	0.005 (0.002)	0.207 (0.004)	0.096 (0.002)	-0.012 (0.002)
1966	0.295 (0.006)	0.006 (0.002)	0.207 (0.004)	0.095 (0.002)	-0.013 (0.003)
1967	0.288 (0.006)	0.007 (0.003)	0.206 (0.004)	0.093 (0.002)	-0.018 (0.003)
1968	0.288 (0.006)	0.009 (0.003)	0.204 (0.004)	0.092 (0.002)	-0.018 (0.004)
1969	0.291 (0.007)	0.010 (0.003)	0.203 (0.004)	0.091 (0.003)	-0.014 (0.004)
1970	0.294 (0.007)	0.011 (0.004)	0.201 (0.004)	0.091 (0.003)	-0.010 (0.004)
1971	0.292 (0.007)	0.012 (0.004)	0.200 (0.004)	0.091 (0.003)	-0.011 (0.004)
1972	0.284 (0.007)	0.013 (0.005)	0.198 (0.004)	0.092 (0.003)	-0.019 (0.004)
1973	0.289 (0.007)	0.014 (0.005)	0.196 (0.004)	0.093 (0.003)	-0.014 (0.004)
1974	0.289 (0.007)	0.014 (0.005)	0.195 (0.004)	0.094 (0.003)	-0.015 (0.003)
1975	0.281 (0.008)	0.014 (0.006)	0.193 (0.004)	0.096 (0.003)	-0.022 (0.004)
1976	0.287 (0.008)	0.014 (0.006)	0.192 (0.004)	0.096 (0.003)	-0.016 (0.004)
1977	0.283 (0.008)	0.013 (0.006)	0.191 (0.004)	0.097 (0.004)	-0.018 (0.005)
1978	0.273 (0.009)	0.012 (0.007)	0.190 (0.004)	0.097 (0.004)	-0.026 (0.005)
1979	0.271 (0.010)	0.011 (0.007)	0.190 (0.004)	0.096 (0.004)	-0.025 (0.005)
1980	0.277 (0.010)	0.008 (0.008)	0.190 (0.004)	0.093 (0.004)	-0.015 (0.004)
1981	0.267 (0.010)	0.006 (0.009)	0.190 (0.004)	0.089 (0.005)	-0.019 (0.004)
1982	0.260 (0.011)	0.003 (0.009)	0.191 (0.004)	0.083 (0.006)	-0.017 (0.004)
1983	0.248 (0.012)	-0.001 (0.010)	0.193 (0.004)	0.075 (0.007)	-0.020 (0.004)
1984	0.235 (0.014)	-0.006 (0.011)	0.195 (0.005)	0.065 (0.009)	-0.020 (0.004)

Notes: This table shows the predicted gender gap in log earnings for each birth cohort at the average age distribution shown in panel A of Figure 2. The table shows birth year specific coefficients for the total log earnings gap and the portion of the gap explained by gender differences in returns to degrees, education, cohort contribution that is not related to education fields, and demographic controls. The coefficient estimates are from regression model (4). OLS coefficients were used. Bootstrapped standard errors are reported in parentheses. The standard errors are estimated from 200 bootstrap iterations. The methodology is explained in section 4.5. The NSCG base year samples are used with cross sectional weights. Ages are restricted to be between 23 and 59. By construction, Total gap = Return gap + Education gap + Birth cohort Residual gap + Demographic control gap.

Table J.2: Constant Decomposition: Occupation Premium

Birth Cohort	Total Gap	Cohort Residual Gap	Return Gap	Education Gap	Demo Control Gap
1931	0.202 (0.011)	0.035 (0.011)	0.057 (0.003)	0.108 (0.003)	0.002 (0.001)
1932	0.200 (0.010)	0.033 (0.010)	0.057 (0.003)	0.110 (0.003)	0.000 (0.001)
1933	0.200 (0.009)	0.031 (0.009)	0.057 (0.003)	0.112 (0.003)	0.000 (0.001)
1934	0.197 (0.008)	0.029 (0.008)	0.057 (0.002)	0.113 (0.002)	-0.002 (0.001)
1935	0.198 (0.007)	0.027 (0.007)	0.057 (0.002)	0.114 (0.002)	0.000 (0.001)
1936	0.194 (0.007)	0.025 (0.007)	0.057 (0.002)	0.114 (0.002)	-0.002 (0.000)
1937	0.193 (0.006)	0.023 (0.006)	0.057 (0.002)	0.114 (0.002)	0.000 (0.000)
1938	0.191 (0.005)	0.021 (0.006)	0.057 (0.002)	0.114 (0.002)	0.000 (0.000)
1939	0.189 (0.005)	0.019 (0.005)	0.057 (0.002)	0.113 (0.002)	0.000 (0.000)
1940	0.185 (0.004)	0.018 (0.005)	0.057 (0.002)	0.112 (0.002)	-0.002 (0.000)
1941	0.184 (0.004)	0.016 (0.004)	0.057 (0.002)	0.111 (0.002)	0.000 (0.000)
1942	0.181 (0.003)	0.015 (0.004)	0.057 (0.002)	0.109 (0.002)	0.000 (0.000)
1943	0.176 (0.003)	0.013 (0.003)	0.057 (0.002)	0.108 (0.002)	-0.003 (0.000)
1944	0.173 (0.003)	0.012 (0.003)	0.057 (0.002)	0.106 (0.002)	-0.002 (0.000)
1945	0.171 (0.003)	0.011 (0.003)	0.057 (0.002)	0.103 (0.002)	0.000 (0.000)
1946	0.166 (0.002)	0.010 (0.003)	0.057 (0.002)	0.101 (0.002)	-0.002 (0.000)
1947	0.163 (0.002)	0.009 (0.003)	0.057 (0.002)	0.098 (0.002)	-0.001 (0.000)
1948	0.161 (0.002)	0.008 (0.002)	0.057 (0.002)	0.096 (0.002)	0.000 (0.000)
1949	0.157 (0.002)	0.007 (0.002)	0.057 (0.002)	0.093 (0.002)	0.000 (0.000)
1950	0.153 (0.002)	0.006 (0.002)	0.057 (0.002)	0.090 (0.001)	0.000 (0.000)
1951	0.149 (0.002)	0.005 (0.002)	0.057 (0.002)	0.087 (0.001)	0.000 (0.000)
1952	0.145 (0.002)	0.004 (0.002)	0.057 (0.002)	0.085 (0.001)	-0.001 (0.000)
1953	0.143 (0.002)	0.004 (0.001)	0.057 (0.002)	0.082 (0.001)	0.000 (0.000)
1954	0.139 (0.002)	0.003 (0.001)	0.056 (0.002)	0.079 (0.001)	0.000 (0.000)
1955	0.135 (0.002)	0.002 (0.001)	0.056 (0.002)	0.077 (0.001)	0.000 (0.000)
1956	0.131 (0.002)	0.002 (0.000)	0.056 (0.002)	0.074 (0.001)	0.000 (0.000)
1957	0.127 (0.002)	0.001 (0.000)	0.056 (0.002)	0.071 (0.001)	-0.001 (0.000)
1958	0.125 (0.002)	0.001 (0.000)	0.056 (0.002)	0.068 (0.001)	0.000 (0.000)
1959	0.121 (0.002)	0.000 (0.000)	0.055 (0.002)	0.065 (0.001)	0.000 (0.000)
1960	0.118 (0.002)	0.000 (0.000)	0.055 (0.002)	0.062 (0.001)	0.000 (0.000)
1961	0.115 (0.002)	0 (0.000)	0.055 (0.002)	0.060 (0.001)	0.000 (0.000)
1962	0.112 (0.002)	0.000 (0.000)	0.055 (0.002)	0.058 (0.001)	0.000 (0.000)
1963	0.110 (0.002)	0.000 (0.000)	0.055 (0.002)	0.057 (0.001)	-0.001 (0.000)
1964	0.109 (0.002)	0.000 (0.000)	0.055 (0.002)	0.056 (0.001)	-0.001 (0.000)
1965	0.108 (0.002)	0.000 (0.000)	0.055 (0.002)	0.055 (0.001)	-0.002 (0.000)
1966	0.107 (0.002)	0.000 (0.000)	0.055 (0.002)	0.055 (0.001)	-0.002 (0.000)
1967	0.107 (0.003)	0.000 (0.001)	0.055 (0.002)	0.055 (0.001)	-0.002 (0.001)
1968	0.108 (0.003)	-0.001 (0.001)	0.055 (0.002)	0.055 (0.001)	-0.001 (0.001)
1969	0.107 (0.003)	-0.001 (0.001)	0.055 (0.002)	0.055 (0.001)	-0.002 (0.001)
1970	0.109 (0.003)	-0.001 (0.002)	0.055 (0.002)	0.056 (0.001)	0.000 (0.001)
1971	0.109 (0.003)	0.000 (0.002)	0.055 (0.002)	0.056 (0.002)	-0.001 (0.001)
1972	0.109 (0.003)	0.000 (0.002)	0.054 (0.002)	0.057 (0.002)	-0.001 (0.001)
1973	0.110 (0.003)	0.000 (0.002)	0.054 (0.002)	0.058 (0.002)	-0.002 (0.001)
1974	0.110 (0.003)	0.000 (0.002)	0.054 (0.002)	0.059 (0.002)	-0.002 (0.001)
1975	0.109 (0.003)	0.000 (0.002)	0.053 (0.002)	0.060 (0.002)	-0.004 (0.001)
1976	0.110 (0.003)	0.000 (0.002)	0.053 (0.002)	0.060 (0.002)	-0.002 (0.001)
1977	0.110 (0.004)	0.000 (0.003)	0.053 (0.002)	0.061 (0.002)	-0.003 (0.001)
1978	0.109 (0.004)	0.000 (0.003)	0.053 (0.002)	0.061 (0.002)	-0.004 (0.001)
1979	0.109 (0.004)	0.000 (0.003)	0.052 (0.002)	0.060 (0.002)	-0.003 (0.001)
1980	0.110 (0.004)	0.000 (0.003)	0.052 (0.002)	0.060 (0.002)	-0.001 (0.001)
1981	0.108 (0.005)	0.000 (0.004)	0.052 (0.002)	0.058 (0.003)	-0.002 (0.001)
1982	0.106 (0.005)	0.000 (0.004)	0.052 (0.002)	0.056 (0.003)	-0.002 (0.001)
1983	0.101 (0.006)	0.000 (0.004)	0.051 (0.002)	0.053 (0.004)	-0.004 (0.001)
1984	0.098 (0.006)	0.000 (0.005)	0.051 (0.002)	0.050 (0.005)	-0.004 (0.001)

Notes: This table shows the predicted gender gap in occupational premiums for each birth cohort at the average age distribution shown in panel A of Figure 6. The table shows birth year specific coefficients for the total occupational premium gap and the portion of the gap explained by gender differences in returns to degrees, education, cohort contribution that is not related to education fields, and demographic controls. The coefficient estimates are from regression model (4). OLS coefficients were used. Bootstrapped standard errors are reported in parentheses. The standard errors are estimated from 200 bootstrap iterations using the methodology explained in section 4.5. The occupation premiums are estimated as described in section 4.5, and are used as the dependent variable in equation (4) to estimate the gender gap. The NSCG base year samples are used with cross sectional weights. Ages are restricted to be between 23 and 59. By construction, Total gap = Return gap + Education gap + Birth cohort residual gap + Demographic gap.

Table J.3: Dynamic Decomposition: Earnings Birth Year Average

Birth Cohort	Total Gap	Cohort Residual Gap	Return: Base Year Return	Return: Varying Year Return	Edu: Base Year Return	Edu: Varying Year Return	Demo Control Gap
1931	0.682 (0.039)	0.336 (0.047)	0.198 (0.009)	0.006 (0.031)	0.196 (0.008)	-0.052 (0.031)	-0.002 (0.004)
1932	0.659 (0.036)	0.308 (0.042)	0.200 (0.008)	0.005 (0.027)	0.201 (0.007)	-0.050 (0.028)	-0.005 (0.005)
1933	0.636 (0.032)	0.282 (0.038)	0.202 (0.008)	0.003 (0.024)	0.205 (0.007)	-0.048 (0.025)	-0.007 (0.004)
1934	0.610 (0.029)	0.257 (0.035)	0.203 (0.008)	0.000 (0.021)	0.207 (0.006)	-0.046 (0.023)	-0.011 (0.004)
1935	0.594 (0.025)	0.233 (0.031)	0.203 (0.008)	-0.003 (0.019)	0.208 (0.006)	-0.043 (0.020)	-0.004 (0.004)
1936	0.563 (0.023)	0.211 (0.028)	0.203 (0.008)	-0.006 (0.016)	0.208 (0.006)	-0.041 (0.018)	-0.013 (0.003)
1937	0.545 (0.021)	0.190 (0.025)	0.203 (0.008)	-0.008 (0.015)	0.207 (0.006)	-0.038 (0.016)	-0.009 (0.003)
1938	0.524 (0.018)	0.171 (0.023)	0.202 (0.008)	-0.010 (0.013)	0.206 (0.006)	-0.036 (0.014)	-0.008 (0.003)
1939	0.502 (0.017)	0.153 (0.021)	0.200 (0.008)	-0.012 (0.012)	0.203 (0.006)	-0.033 (0.013)	-0.009 (0.003)
1940	0.480 (0.015)	0.136 (0.018)	0.199 (0.008)	-0.013 (0.011)	0.200 (0.005)	-0.031 (0.011)	-0.011 (0.003)
1941	0.463 (0.013)	0.121 (0.017)	0.197 (0.008)	-0.014 (0.010)	0.197 (0.005)	-0.029 (0.010)	-0.009 (0.003)
1942	0.444 (0.012)	0.107 (0.015)	0.195 (0.008)	-0.014 (0.009)	0.192 (0.005)	-0.027 (0.009)	-0.009 (0.002)
1943	0.422 (0.011)	0.093 (0.014)	0.194 (0.008)	-0.014 (0.008)	0.187 (0.005)	-0.025 (0.008)	-0.014 (0.003)
1944	0.407 (0.009)	0.081 (0.012)	0.193 (0.008)	-0.014 (0.007)	0.182 (0.005)	-0.022 (0.007)	-0.012 (0.002)
1945	0.395 (0.008)	0.070 (0.011)	0.192 (0.008)	-0.013 (0.007)	0.176 (0.005)	-0.021 (0.007)	-0.010 (0.002)
1946	0.379 (0.007)	0.060 (0.010)	0.191 (0.007)	-0.012 (0.006)	0.170 (0.005)	-0.019 (0.006)	-0.011 (0.002)
1947	0.369 (0.007)	0.051 (0.010)	0.191 (0.007)	-0.012 (0.005)	0.164 (0.004)	-0.017 (0.006)	-0.009 (0.002)
1948	0.360 (0.006)	0.043 (0.009)	0.191 (0.007)	-0.011 (0.005)	0.158 (0.004)	-0.016 (0.005)	-0.007 (0.002)
1949	0.350 (0.006)	0.036 (0.008)	0.192 (0.007)	-0.009 (0.004)	0.153 (0.004)	-0.014 (0.005)	-0.008 (0.002)
1950	0.342 (0.006)	0.030 (0.007)	0.193 (0.007)	-0.008 (0.004)	0.148 (0.004)	-0.012 (0.004)	-0.008 (0.002)
1951	0.335 (0.005)	0.024 (0.007)	0.195 (0.007)	-0.007 (0.003)	0.143 (0.004)	-0.011 (0.004)	-0.008 (0.002)
1952	0.327 (0.005)	0.019 (0.006)	0.196 (0.007)	-0.006 (0.003)	0.138 (0.004)	-0.010 (0.003)	-0.010 (0.002)
1953	0.328 (0.005)	0.015 (0.006)	0.197 (0.007)	-0.005 (0.002)	0.134 (0.004)	-0.008 (0.003)	-0.005 (0.002)
1954	0.322 (0.005)	0.011 (0.005)	0.199 (0.007)	-0.004 (0.002)	0.131 (0.004)	-0.007 (0.003)	-0.007 (0.002)
1955	0.320 (0.005)	0.008 (0.004)	0.200 (0.006)	-0.003 (0.002)	0.128 (0.004)	-0.006 (0.002)	-0.007 (0.002)
1956	0.316 (0.005)	0.006 (0.004)	0.201 (0.006)	-0.002 (0.001)	0.125 (0.004)	-0.004 (0.002)	-0.009 (0.002)
1957	0.312 (0.005)	0.004 (0.003)	0.202 (0.006)	-0.002 (0.000)	0.122 (0.004)	-0.003 (0.001)	-0.010 (0.002)
1958	0.315 (0.005)	0.002 (0.002)	0.203 (0.006)	-0.001 (0.000)	0.118 (0.004)	-0.002 (0.001)	-0.005 (0.002)
1959	0.311 (0.005)	0.001 (0.001)	0.204 (0.006)	0.000 (0.000)	0.115 (0.004)	-0.002 (0.000)	-0.007 (0.002)
1960	0.310 (0.005)	0.000 (0.000)	0.205 (0.006)	0.000 (0.000)	0.112 (0.004)	0.000 (0.000)	-0.006 (0.003)
1961	0.307 (0.006)	0 (0.000)	0.205 (0.006)	0.000 (0.000)	0.109 (0.004)	0.000 (0.000)	-0.007 (0.003)
1962	0.302 (0.006)	0.000 (0.000)	0.206 (0.006)	0.000 (0.000)	0.106 (0.004)	0.000 (0.000)	-0.010 (0.003)
1963	0.302 (0.006)	0.000 (0.001)	0.206 (0.006)	0.000 (0.000)	0.103 (0.003)	0.001 (0.000)	-0.009 (0.003)
1964	0.301 (0.007)	0.000 (0.002)	0.206 (0.006)	0.000 (0.000)	0.100 (0.003)	0.002 (0.000)	-0.008 (0.003)
1965	0.296 (0.007)	0.000 (0.003)	0.206 (0.006)	0.000 (0.000)	0.098 (0.003)	0.003 (0.001)	-0.012 (0.003)
1966	0.294 (0.008)	0.002 (0.004)	0.206 (0.006)	0.000 (0.001)	0.096 (0.003)	0.004 (0.001)	-0.013 (0.004)
1967	0.288 (0.008)	0.002 (0.004)	0.205 (0.006)	0.000 (0.001)	0.094 (0.003)	0.005 (0.002)	-0.018 (0.004)
1968	0.287 (0.009)	0.003 (0.005)	0.204 (0.006)	-0.001 (0.001)	0.093 (0.003)	0.006 (0.002)	-0.018 (0.006)
1969	0.290 (0.010)	0.003 (0.006)	0.202 (0.006)	-0.002 (0.002)	0.092 (0.003)	0.007 (0.002)	-0.014 (0.005)
1970	0.294 (0.010)	0.004 (0.006)	0.201 (0.006)	-0.003 (0.002)	0.092 (0.003)	0.009 (0.002)	-0.010 (0.005)
1971	0.291 (0.009)	0.004 (0.007)	0.200 (0.006)	-0.003 (0.002)	0.092 (0.003)	0.010 (0.002)	-0.011 (0.004)
1972	0.284 (0.010)	0.005 (0.007)	0.198 (0.006)	-0.004 (0.002)	0.093 (0.004)	0.011 (0.003)	-0.019 (0.005)
1973	0.289 (0.011)	0.005 (0.008)	0.197 (0.006)	-0.005 (0.002)	0.094 (0.004)	0.013 (0.003)	-0.014 (0.005)
1974	0.289 (0.011)	0.005 (0.008)	0.195 (0.006)	-0.005 (0.003)	0.095 (0.004)	0.014 (0.003)	-0.015 (0.004)
1975	0.281 (0.011)	0.004 (0.009)	0.194 (0.006)	-0.006 (0.003)	0.096 (0.004)	0.015 (0.004)	-0.023 (0.005)
1976	0.286 (0.012)	0.003 (0.009)	0.193 (0.006)	-0.007 (0.003)	0.097 (0.004)	0.016 (0.004)	-0.016 (0.005)
1977	0.283 (0.012)	0.002 (0.010)	0.192 (0.006)	-0.007 (0.003)	0.097 (0.004)	0.017 (0.004)	-0.018 (0.006)
1978	0.273 (0.013)	0.000 (0.010)	0.192 (0.006)	-0.007 (0.004)	0.097 (0.005)	0.017 (0.004)	-0.027 (0.007)
1979	0.271 (0.013)	-0.002 (0.011)	0.191 (0.006)	-0.008 (0.004)	0.096 (0.005)	0.018 (0.005)	-0.025 (0.006)
1980	0.277 (0.012)	-0.004 (0.012)	0.191 (0.006)	-0.008 (0.004)	0.093 (0.005)	0.019 (0.005)	-0.015 (0.006)
1981	0.267 (0.013)	-0.008 (0.013)	0.192 (0.006)	-0.008 (0.004)	0.089 (0.006)	0.020 (0.005)	-0.019 (0.005)
1982	0.261 (0.014)	-0.012 (0.014)	0.193 (0.006)	-0.008 (0.005)	0.083 (0.006)	0.021 (0.005)	-0.017 (0.005)
1983	0.249 (0.016)	-0.016 (0.015)	0.195 (0.006)	-0.008 (0.005)	0.075 (0.008)	0.023 (0.006)	-0.020 (0.006)
1984	0.238 (0.019)	-0.021 (0.016)	0.198 (0.006)	-0.008 (0.006)	0.064 (0.010)	0.025 (0.007)	-0.020 (0.006)

Notes: This table shows the predicted gender gap in log earnings for each birth cohort at the average age distribution shown in panel A, B, and C of Figure 4. The table shows birth year specific coefficients for the total log earnings gap and the portion of the gap explained by gender differences in returns to degrees, education, cohort contribution that is not related to education fields, demographic controls, and the base year component and cohort varying component of the returns and education gaps respectively. The coefficient estimates are from regression model (10). Bootstrapped standard errors, reported in parentheses, are estimated from 200 iterations using the methodology explained in section 4.5. By construction, Total Gap = Cohort Residual gap + Return gap (Base year) + Return gap (Varying year) + Education gap (Base year) + Education gap (Varying year) + Demo control gap.

Table J.4: Dynamic Decomposition: Occupation Premium Birth Year Average

Birth Cohort	Total Gap	Cohort Residual Gap	Return: Base Year Return	Return: Varying Year Return	Edu: Base Year Return	Edu: Varying Year Return	Demo Control Gap
1931	0.206 (0.014)	0.034 (0.017)	0.061 (0.004)	-0.026 (0.012)	0.103 (0.004)	0.033 (0.014)	0.002 (0.002)
1932	0.206 (0.013)	0.032 (0.016)	0.062 (0.004)	-0.023 (0.011)	0.104 (0.004)	0.030 (0.013)	0.000 (0.002)
1933	0.206 (0.012)	0.031 (0.015)	0.063 (0.004)	-0.021 (0.009)	0.105 (0.003)	0.027 (0.011)	0.000 (0.001)
1934	0.202 (0.010)	0.029 (0.013)	0.064 (0.004)	-0.019 (0.008)	0.106 (0.003)	0.024 (0.010)	-0.002 (0.001)
1935	0.202 (0.009)	0.028 (0.012)	0.064 (0.004)	-0.017 (0.007)	0.107 (0.003)	0.021 (0.009)	0.000 (0.001)
1936	0.198 (0.008)	0.026 (0.011)	0.064 (0.004)	-0.015 (0.007)	0.107 (0.003)	0.018 (0.008)	-0.002 (0.001)
1937	0.196 (0.008)	0.025 (0.010)	0.064 (0.004)	-0.014 (0.006)	0.107 (0.003)	0.015 (0.007)	0.000 (0.001)
1938	0.193 (0.007)	0.023 (0.009)	0.064 (0.004)	-0.012 (0.005)	0.107 (0.003)	0.012 (0.007)	0.000 (0.001)
1939	0.190 (0.006)	0.022 (0.008)	0.064 (0.004)	-0.011 (0.005)	0.106 (0.003)	0.010 (0.006)	0.000 (0.001)
1940	0.186 (0.006)	0.021 (0.008)	0.063 (0.004)	-0.009 (0.004)	0.106 (0.003)	0.008 (0.005)	-0.002 (0.001)
1941	0.185 (0.005)	0.019 (0.007)	0.063 (0.004)	-0.008 (0.004)	0.105 (0.003)	0.006 (0.005)	0.000 (0.001)
1942	0.181 (0.005)	0.018 (0.006)	0.062 (0.004)	-0.007 (0.004)	0.104 (0.003)	0.004 (0.004)	0.000 (0.000)
1943	0.176 (0.004)	0.017 (0.006)	0.062 (0.004)	-0.005 (0.003)	0.103 (0.003)	0.002 (0.004)	-0.002 (0.000)
1944	0.174 (0.004)	0.015 (0.005)	0.062 (0.004)	-0.004 (0.003)	0.102 (0.003)	0.000 (0.004)	-0.002 (0.000)
1945	0.171 (0.004)	0.014 (0.005)	0.061 (0.004)	-0.003 (0.003)	0.100 (0.002)	0.000 (0.003)	0.000 (0.000)
1946	0.167 (0.003)	0.013 (0.005)	0.061 (0.004)	-0.003 (0.003)	0.098 (0.002)	0.000 (0.003)	-0.002 (0.000)
1947	0.164 (0.003)	0.012 (0.004)	0.061 (0.004)	-0.002 (0.002)	0.095 (0.002)	-0.002 (0.003)	-0.001 (0.000)
1948	0.161 (0.003)	0.011 (0.004)	0.061 (0.003)	-0.002 (0.002)	0.093 (0.002)	-0.002 (0.003)	0.000 (0.000)
1949	0.157 (0.003)	0.010 (0.004)	0.060 (0.003)	-0.001 (0.002)	0.090 (0.002)	-0.002 (0.002)	0.000 (0.000)
1950	0.153 (0.003)	0.009 (0.003)	0.060 (0.003)	-0.001 (0.002)	0.088 (0.002)	-0.002 (0.002)	0.000 (0.000)
1951	0.150 (0.003)	0.008 (0.003)	0.060 (0.003)	-0.001 (0.001)	0.085 (0.002)	-0.002 (0.002)	0.000 (0.000)
1952	0.145 (0.002)	0.007 (0.003)	0.060 (0.003)	-0.001 (0.001)	0.082 (0.002)	-0.002 (0.002)	0.000 (0.000)
1953	0.143 (0.002)	0.006 (0.002)	0.060 (0.003)	-0.001 (0.001)	0.080 (0.002)	-0.002 (0.001)	0.000 (0.000)
1954	0.139 (0.002)	0.005 (0.002)	0.060 (0.003)	-0.001 (0.000)	0.077 (0.002)	-0.001 (0.001)	0.000 (0.000)
1955	0.136 (0.002)	0.004 (0.002)	0.059 (0.003)	-0.001 (0.000)	0.074 (0.002)	0.000 (0.001)	0.000 (0.000)
1956	0.132 (0.002)	0.003 (0.001)	0.059 (0.003)	0.000 (0.000)	0.071 (0.002)	0.000 (0.000)	0.000 (0.000)
1957	0.127 (0.002)	0.003 (0.001)	0.059 (0.003)	0.000 (0.000)	0.069 (0.002)	0.000 (0.000)	0.000 (0.000)
1958	0.126 (0.002)	0.002 (0.000)	0.058 (0.003)	0.000 (0.000)	0.066 (0.002)	0.000 (0.000)	0.000 (0.000)
1959	0.121 (0.002)	0.001 (0.000)	0.058 (0.003)	0.000 (0.000)	0.063 (0.002)	0.000 (0.000)	0.000 (0.000)
1960	0.119 (0.002)	0.000 (0.000)	0.058 (0.003)	0.000 (0.000)	0.060 (0.002)	0.000 (0.000)	0.000 (0.000)
1961	0.115 (0.003)	0 (0.000)	0.057 (0.003)	0.000 (0.000)	0.058 (0.002)	0.000 (0.000)	0.000 (0.001)
1962	0.112 (0.003)	0.000 (0.000)	0.057 (0.003)	0.000 (0.000)	0.056 (0.002)	0.000 (0.000)	0.000 (0.000)
1963	0.110 (0.003)	0.000 (0.000)	0.057 (0.003)	0.000 (0.000)	0.055 (0.002)	0.000 (0.000)	0.000 (0.000)
1964	0.109 (0.003)	-0.001 (0.000)	0.057 (0.003)	0.000 (0.000)	0.054 (0.002)	0.000 (0.000)	0.000 (0.000)
1965	0.107 (0.003)	-0.002 (0.001)	0.057 (0.003)	0.000 (0.000)	0.053 (0.002)	0.000 (0.000)	-0.002 (0.001)
1966	0.107 (0.003)	-0.002 (0.001)	0.057 (0.003)	0.000 (0.000)	0.053 (0.002)	0.000 (0.000)	-0.002 (0.001)
1967	0.106 (0.003)	-0.002 (0.002)	0.057 (0.003)	0.000 (0.000)	0.052 (0.002)	0.001 (0.000)	-0.002 (0.001)
1968	0.107 (0.003)	-0.002 (0.002)	0.057 (0.003)	0.000 (0.000)	0.052 (0.002)	0.001 (0.000)	-0.001 (0.002)
1969	0.108 (0.003)	-0.003 (0.002)	0.057 (0.003)	0.000 (0.000)	0.053 (0.002)	0.002 (0.000)	-0.002 (0.002)
1970	0.109 (0.003)	-0.003 (0.002)	0.057 (0.003)	0.000 (0.000)	0.053 (0.002)	0.002 (0.001)	0.000 (0.001)
1971	0.109 (0.004)	-0.003 (0.002)	0.057 (0.003)	0.000 (0.000)	0.054 (0.002)	0.003 (0.001)	0.000 (0.001)
1972	0.110 (0.004)	-0.003 (0.003)	0.057 (0.003)	0.000 (0.000)	0.054 (0.002)	0.003 (0.001)	-0.001 (0.001)
1973	0.110 (0.004)	-0.003 (0.003)	0.056 (0.003)	0.000 (0.000)	0.055 (0.002)	0.003 (0.001)	-0.001 (0.002)
1974	0.110 (0.004)	-0.002 (0.003)	0.056 (0.003)	-0.001 (0.001)	0.056 (0.002)	0.003 (0.002)	-0.002 (0.002)
1975	0.109 (0.004)	-0.002 (0.003)	0.056 (0.003)	-0.001 (0.001)	0.057 (0.002)	0.003 (0.002)	-0.004 (0.002)
1976	0.111 (0.004)	-0.002 (0.003)	0.056 (0.003)	-0.001 (0.001)	0.057 (0.002)	0.003 (0.002)	-0.002 (0.001)
1977	0.110 (0.004)	-0.001 (0.004)	0.055 (0.003)	-0.002 (0.001)	0.058 (0.002)	0.002 (0.002)	-0.003 (0.002)
1978	0.109 (0.004)	-0.001 (0.004)	0.055 (0.003)	-0.002 (0.002)	0.058 (0.002)	0.002 (0.002)	-0.004 (0.002)
1979	0.109 (0.005)	0.000 (0.004)	0.055 (0.003)	-0.002 (0.002)	0.058 (0.002)	0.001 (0.002)	-0.003 (0.002)
1980	0.110 (0.005)	0.000 (0.005)	0.055 (0.003)	-0.002 (0.002)	0.057 (0.003)	0.000 (0.002)	-0.001 (0.002)
1981	0.108 (0.005)	0.000 (0.005)	0.054 (0.003)	-0.002 (0.002)	0.056 (0.003)	0.000 (0.002)	-0.002 (0.002)
1982	0.107 (0.006)	0.001 (0.005)	0.054 (0.003)	-0.001 (0.002)	0.053 (0.003)	0.000 (0.002)	-0.001 (0.002)
1983	0.103 (0.006)	0.002 (0.006)	0.054 (0.003)	-0.001 (0.002)	0.049 (0.004)	0.002 (0.003)	-0.004 (0.002)
1984	0.102 (0.007)	0.003 (0.007)	0.055 (0.003)	-0.002 (0.003)	0.045 (0.005)	0.004 (0.003)	-0.003 (0.002)

Notes: This table shows the predicted gender gap in occupational premiums for each birth cohort at the average age distribution shown in panel A, B, and C of Figure 8. The table shows birth year specific coefficients for the total occupational premium gap and the portion of the gap explained by gender differences in returns to degrees, education, cohort contribution that is not related to education fields, demographic controls, and the base year and cohort varying components of the return and education gaps respectively. The coefficient estimates are from regression model (10). Bootstrapped standard errors are reported in parentheses and are estimated from 200 iterations explained in section 4.5. By construction, Total Gap = Cohort Residual gap + Return gap (Base year) + Return gap (Varying year) + Education gap (Base year) + Education gap (Varying year) + Demo control gap.

Table J.5: Dynamic Decomposition: Total Gap, Within and Across Occupation

Birth Cohort	1932	1940	1948	1964	1975	1982
Edu gap, within occ, base year	0.097	0.094	0.065	0.047	0.039	0.030
Edu gap, across occ, base year	0.104	0.106	0.093	0.054	0.057	0.053
Edu gap, within occ, varying return	-0.080	-0.039	-0.014	0.002	0.012	0.020
Edu gap, across occ, varying return	0.030	0.008	-0.002	0.000	0.003	0.001
Return gap, within occ, base year	0.138	0.135	0.131	0.149	0.138	0.139
Return gap, across occ, base year	0.062	0.063	0.061	0.057	0.056	0.054
Return gap, within occ, varying return	0.028	-0.004	-0.009	-0.000	-0.005	-0.007
Return gap, across occ, varying return	-0.023	-0.009	-0.002	0.000	-0.001	-0.001
Cohort residual gap, within occ	0.276	0.116	0.032	0.002	0.006	-0.013
Cohort residual gap, across occ	0.032	0.021	0.011	-0.001	-0.002	0.001
Demographic control gap	-0.005	-0.011	-0.007	-0.008	-0.023	-0.017
Overall gap, within occ	0.459	0.303	0.206	0.199	0.191	0.170
Overall gap, across occ	0.205	0.188	0.161	0.110	0.112	0.108
Overall gap, total	0.659	0.480	0.360	0.301	0.281	0.261

Notes: This table shows the predicted gender gap in earnings for selected birth cohorts by the within and across occupation effects. The coefficient estimates are from regression model (10). The gender gap of log earning is the sum of the within occupation effect and across occupation effect. The across occupation effect is estimated by decomposing the gender gap in occupation premium. So we can calculate the within occupation gap by taking the difference between the gap estimated for log earnings and the gap estimated for occupation premium, for each component in the decomposition. Specifically, the total overall gap is the left-hand side of equation (10) when we decompose the gender gap in log earnings, and is the same as the black line in Figure 4 panel A. The overall gap across occupation is the left-hand side of equation (10) when we decompose the gender gap in occupation premium (the black line in Figure 8 panel A). The overall gap within occupation is the difference between the total gap and the across occupation gap. Similarly, the sum of within and across occupation gaps is the total gap for the base year return education gap (light green line in Figure 4 panel C), the varying return education gap (yellow line in Figure 4 panel C), the base year return gap (light green line in Figure 4 panel B), and the varying return gap (yellow line in Figure 4 panel B), respectively.

Table J.6: Dynamic Decomposition: Education Gap, Within and Across Occupation

Birth Cohort	1932	1940	1948	1964	1975	1982
BA field, within occ, base year	0.068	0.059	0.042	0.039	0.039	0.034
BA field, across occ, base year	0.087	0.082	0.071	0.043	0.048	0.043
BA field, within occ, varying return	-0.081	-0.035	-0.011	0.002	0.010	0.018
BA field, across occ, varying return	0.021	0.003	-0.004	0.000	0.002	0.001
Grad attendance, within occ, base year	0.022	0.019	0.013	0.003	-0.006	-0.008
Grad attendance, across occ, base year	0.000	-0.007	0.001	0.001	-0.001	-0.002
Grad attendance, within occ, varying return	-0.015	-0.005	-0.001	0.000	-0.000	0.000
Grad attendance, across occ, varying return	0.008	0.007	0.002	-0.000	0.001	0.002
Grad field, within occ, base year	0.007	0.016	0.011	0.004	0.006	0.004
Grad field, across occ, base year	0.017	0.031	0.022	0.009	0.011	0.011
Grad field, within occ, varying return	0.016	0.001	-0.002	0.000	0.002	0.002
Grad field, across occ, varying return	0.001	-0.002	-0.001	-0.000	-0.001	-0.002
Edu gap, within occ	0.017	0.056	0.052	0.049	0.051	0.051
Edu gap, across occ	0.134	0.114	0.091	0.054	0.060	0.054
Edu gap, total	0.151	0.169	0.143	0.102	0.111	0.104

Notes: This table shows the predicted education gender gap in earnings for selected birth cohorts by the within and across occupation effects. The coefficient estimates are from regression model 12. The within occupation gap is calculated by taking the difference between the gap in earnings and the gap in occupation premium (i.e. across occupation gap). The total education gap is the green line in Figure 4 panel A. The across occupation education gap (second to last row) is the green line in Figure 8 panel A. Panel D of Figures 4 and 8 decompose the base year return education gap into the contributions of BA field, Grad attendance, and Grad field. For example, the pink line in Figure 4 panel D is the sum of "BA field, within occ, base year" and "BA field, across occ, base year". This decomposition table includes both the base year return and the varying return gaps.

Table J.7: Constant Decomposition: Log Earnings

Birth Cohort	Total Gap	Cohort Residual Gap	Return Gap	Education Gap	Demo Gap
1931	0.662	0.253	0.208	0.203	-0.002
1932	0.633	0.230	0.207	0.201	-0.005
1933	0.615	0.208	0.202	0.212	-0.007
1934	0.588	0.188	0.203	0.209	-0.011
1935	0.578	0.169	0.203	0.210	-0.004
1936	0.546	0.151	0.206	0.202	-0.013
1937	0.526	0.134	0.207	0.194	-0.009
1938	0.503	0.119	0.207	0.186	-0.008
1939	0.487	0.105	0.204	0.188	-0.009
1940	0.471	0.092	0.202	0.188	-0.011
1941	0.464	0.080	0.200	0.193	-0.009
1942	0.451	0.069	0.199	0.192	-0.009
1943	0.427	0.059	0.199	0.183	-0.014
1944	0.415	0.050	0.197	0.181	-0.013
1945	0.401	0.041	0.195	0.174	-0.010
1946	0.385	0.034	0.193	0.169	-0.011
1947	0.374	0.028	0.194	0.162	-0.009
1948	0.365	0.022	0.196	0.154	-0.007
1949	0.357	0.017	0.198	0.151	-0.008
1950	0.348	0.013	0.197	0.146	-0.008
1951	0.337	0.009	0.198	0.139	-0.008
1952	0.327	0.006	0.199	0.132	-0.010
1953	0.324	0.004	0.200	0.126	-0.005
1954	0.321	0.002	0.200	0.127	-0.007
1955	0.319	0.000	0.199	0.126	-0.007
1956	0.315	0.000	0.201	0.123	-0.009
1957	0.310	-0.001	0.202	0.118	-0.010
1958	0.314	-0.001	0.204	0.116	-0.005
1959	0.311	-0.001	0.205	0.114	-0.007
1960	0.311	0.000	0.207	0.111	-0.006
1961	0.307	0.000	0.209	0.105	-0.007
1962	0.301	0.000	0.208	0.102	-0.010
1963	0.299	0.002	0.207	0.099	-0.009
1964	0.296	0.003	0.206	0.095	-0.008
1965	0.294	0.005	0.208	0.093	-0.012
1966	0.295	0.006	0.209	0.093	-0.013
1967	0.293	0.007	0.208	0.096	-0.018
1968	0.291	0.009	0.204	0.096	-0.018
1969	0.295	0.010	0.202	0.097	-0.014
1970	0.293	0.011	0.200	0.091	-0.010
1971	0.295	0.012	0.198	0.096	-0.011
1972	0.282	0.013	0.197	0.091	-0.019
1973	0.285	0.014	0.194	0.092	-0.014
1974	0.288	0.014	0.195	0.094	-0.015
1975	0.283	0.014	0.192	0.099	-0.022
1976	0.292	0.014	0.194	0.100	-0.016
1977	0.280	0.013	0.195	0.091	-0.018
1978	0.276	0.012	0.195	0.095	-0.026
1979	0.269	0.011	0.190	0.093	-0.025
1980	0.272	0.008	0.187	0.091	-0.015
1981	0.252	0.006	0.189	0.076	-0.019
1982	0.265	0.003	0.192	0.088	-0.017
1983	0.252	-0.001	0.196	0.077	-0.020
1984	0.253	-0.006	0.193	0.086	-0.020
1985	*0.246*	*-0.006*	*0.190*	*0.082*	*-0.020*
1986	*0.246*	*-0.006*	*0.190*	*0.082*	*-0.020*
1987	*0.247*	*-0.006*	*0.190*	*0.083*	*-0.020*
1988	*0.248*	*-0.006*	*0.189*	*0.085*	*-0.020*
1989	*0.249*	*-0.006*	*0.189*	*0.086*	*-0.020*
1990	*0.251*	*-0.006*	*0.189*	*0.088*	*-0.020*
1991	*0.253*	*-0.006*	*0.188*	*0.091*	*-0.020*
1992	*0.256*	*-0.006*	*0.188*	*0.094*	*-0.020*
1993	*0.259*	*-0.006*	*0.187*	*0.098*	*-0.020*

Notes: This table shows the predicted gender gap in log earnings for each birth cohort at the average age distribution using the moving average specification (see Appendix G) shown in panel A of Figure G.3. Additionally, the table provides estimates for the extrapolated gender gap from 1985-1993 (italicized) which similarly use the moving average specification and are discussed in the Conclusion.

Table J.8: Dynamic Decomposition: Earnings Birth Year Average

Birth Cohort	Total Gap	Cohort Residual Gap	Return: Base Year Return	Return: Varying Year Return	Edu: Base Year Return	Edu: Varying Year Return	Demo Control Gap
1931	0.752	0.451	0.198	0.025	0.211	-0.127	-0.007
1932	0.735	0.412	0.197	0.026	0.210	-0.124	0.013
1933	0.673	0.376	0.197	0.024	0.218	-0.127	-0.015
1934	0.648	0.342	0.195	0.024	0.218	-0.120	-0.012
1935	0.626	0.310	0.194	0.019	0.219	-0.111	-0.006
1936	0.594	0.280	0.194	0.016	0.214	-0.096	-0.014
1937	0.566	0.252	0.194	0.012	0.205	-0.086	-0.011
1938	0.546	0.226	0.195	0.011	0.196	-0.079	-0.004
1939	0.522	0.202	0.196	0.013	0.194	-0.076	-0.007
1940	0.502	0.179	0.197	0.011	0.192	-0.072	-0.007
1941	0.492	0.159	0.200	0.007	0.194	-0.062	-0.005
1942	0.470	0.140	0.199	0.003	0.194	-0.056	-0.010
1943	0.448	0.122	0.198	0.003	0.185	-0.048	-0.012
1944	0.436	0.107	0.198	0.003	0.183	-0.045	-0.009
1945	0.418	0.092	0.200	0.001	0.172	-0.039	-0.009
1946	0.401	0.079	0.202	0.002	0.163	-0.036	-0.009
1947	0.388	0.067	0.202	0.002	0.157	-0.032	-0.008
1948	0.378	0.057	0.202	0.003	0.151	-0.030	-0.006
1949	0.369	0.047	0.201	0.002	0.151	-0.027	-0.006
1950	0.359	0.039	0.201	0.002	0.146	-0.023	-0.006
1951	0.348	0.032	0.202	0.002	0.139	-0.020	-0.007
1952	0.341	0.026	0.204	0.002	0.132	-0.017	-0.006
1953	0.333	0.020	0.205	0.002	0.124	-0.014	-0.005
1954	0.334	0.016	0.205	0.002	0.126	-0.012	-0.003
1955	0.327	0.012	0.206	0.002	0.123	-0.009	-0.007
1956	0.326	0.009	0.205	0.002	0.122	-0.007	-0.004
1957	0.319	0.006	0.204	0.001	0.118	-0.005	-0.006
1958	0.320	0.004	0.204	0.000	0.118	-0.003	-0.003
1959	0.314	0.002	0.204	0.000	0.116	-0.002	-0.006
1960	0.309	0.000	0.204	0.000	0.113	0.000	-0.008
1961	0.310	0.000	0.203	0.000	0.109	0.000	-0.003
1962	0.301	0.000	0.203	0.000	0.106	0.000	-0.007
1963	0.298	-0.001	0.202	0.000	0.103	0.001	-0.007
1964	0.295	-0.002	0.202	0.000	0.099	0.002	-0.005
1965	0.290	-0.002	0.201	0.000	0.099	0.002	-0.010
1966	0.292	-0.003	0.203	0.000	0.099	0.003	-0.009
1967	0.291	-0.003	0.205	0.000	0.099	0.003	-0.011
1968	0.284	-0.004	0.205	-0.001	0.096	0.004	-0.015
1969	0.291	-0.005	0.203	-0.001	0.095	0.006	-0.008
1970	0.287	-0.006	0.199	0.000	0.092	0.007	-0.005
1971	0.287	-0.008	0.198	0.002	0.097	0.007	-0.008
1972	0.275	-0.010	0.197	0.001	0.092	0.008	-0.013
1973	0.274	-0.013	0.197	0.001	0.091	0.008	-0.011
1974	0.273	-0.016	0.196	0.000	0.097	0.008	-0.011
1975	0.263	-0.020	0.196	0.000	0.100	0.008	-0.020
1976	0.266	-0.025	0.197	-0.001	0.099	0.008	-0.012
1977	0.250	-0.031	0.199	0.000	0.087	0.007	-0.012
1978	0.237	-0.037	0.198	0.003	0.092	0.008	-0.026
1979	0.235	-0.045	0.197	0.002	0.090	0.009	-0.019
1980	0.240	-0.053	0.193	0.003	0.090	0.015	-0.009
1981	0.215	-0.063	0.194	0.001	0.078	0.016	-0.011
1982	0.218	-0.074	0.192	0.001	0.094	0.018	-0.014
1983	0.200	-0.086	0.192	0.000	0.086	0.017	-0.010
1984	0.193	-0.099	0.189	0.001	0.095	0.023	-0.015
1985	*0.246*	*-0.022*	*0.193*	*-0.005*	*0.079*	*0.020*	*-0.020*
1986	*0.246*	*-0.022*	*0.193*	*-0.005*	*0.080*	*0.020*	*-0.020*
1987	*0.247*	*-0.022*	*0.193*	*-0.005*	*0.080*	*0.020*	*-0.020*
1988	*0.247*	*-0.022*	*0.193*	*-0.005*	*0.081*	*0.020*	*-0.020*
1989	*0.248*	*-0.022*	*0.193*	*-0.006*	*0.083*	*0.020*	*-0.020*
1990	*0.249*	*-0.022*	*0.193*	*-0.006*	*0.084*	*0.020*	*-0.020*
1991	*0.250*	*-0.022*	*0.193*	*-0.007*	*0.086*	*0.020*	*-0.020*
1992	*0.252*	*-0.022*	*0.194*	*-0.008*	*0.088*	*0.020*	*-0.020*
1993	*0.253*	*-0.022*	*0.194*	*-0.010*	*0.091*	*0.021*	*-0.020*

Notes: This table shows the predicted gender gap in log earnings for each birth cohort at the average age distribution using the moving average specification (see Appendix G). Additionally, the table provides estimates for the extrapolated gender gap from 1985-1993 (italicized), which similarly use the moving average specification and are discussed in the Conclusion.

K Decomposition Using Log Hourly Wage

Our decompositions of gender differences in log annual earnings of full time workers capture differences in both annual hours conditional on full time work and wage rates. Data from the 1960-2000 decennial Census and 2001-2018 ACS show that, across all birth cohorts, men work more hours than women, with the peak being between 3 and 4.5 hours around age 35 (Appendix Figure L.4 panel A). However, in the most recent cohorts, the hours gap declines with age.

In this appendix, we isolate gender differences in wage rates by re-running our decomposition analysis, replacing log annual earnings with log hourly wage earnings. For the log hourly earnings constant returns specification, the total gap starts at 0.488 for the 1931 birth cohort, declining rapidly by an average of 0.0114 per year to 0.306 in 1947. After 1947, the rate of decline then slows to an average of 0.007 per year until 1954 when the gap reaches 0.255. From 1955 onward the gap declines by only about 0.001 per year, ending at 0.212 for the 1984 birth cohort.

Breaking down the total gap, the relative return gap varies little. It begins at 0.146 for the 1931 birth cohort, peaks at 0.156 in 1960, and ends at 0.144 in 1984. The cohort residual gap drops quickly for the early cohorts, starting at 0.180 in 1931 and drops by an average of 0.010, 0.006, and 0.003 for the 1930s, 1940s, and 1950s birth cohorts until it is 0 in 1961. It then remains close to zero until 1971, after which it begins to gradually rise by an average of 0.002 per year, reaching 0.022 by the 1984 birth cohort.

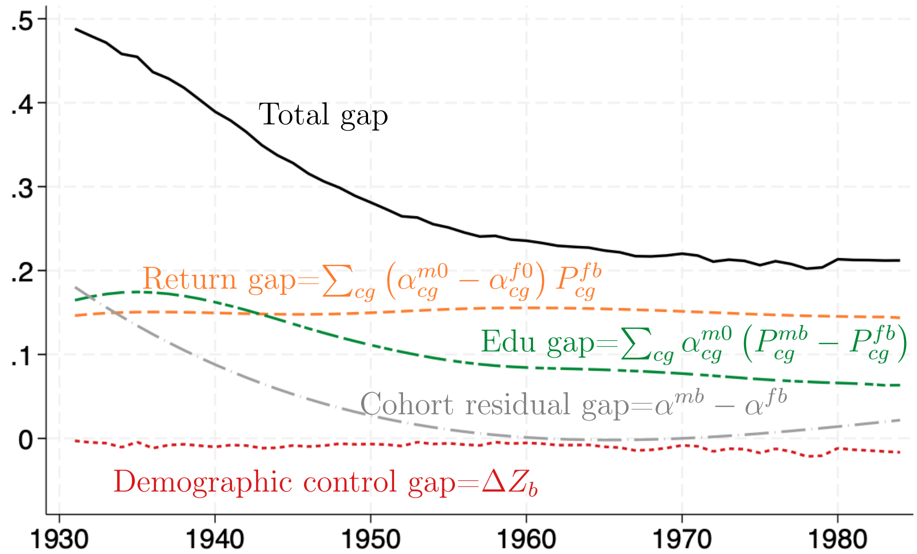
Finally, the education gap starts at 0.165 in 1931 and peaks at 0.174 for the 1935 birth cohort. It then decreases quickly by an average of 0.004 per year from 1936 to 1950, reaching 0.107. After 1950, the decline slows considerably, falling by around 0.001 per year till it ends at 0.063 for the 1984 birth cohort.

Comparing the constant log hourly earnings decomposition to the log annual earnings decomposition, the hourly wage gap is 0.157 smaller in 1931 and 0.023 smaller in 1984. However, the make up of the decreasing trends in the gaps are similar. For log annual earnings, the total gap, relative gap, and education gap decline by 63%, 9%, and 67% respectively from 1931; for log hourly wage the corresponding declines are 57%, 8%, and 64%. The main difference is that the decline in the cohort residual gap between 1931 and 1984 in the log wage case is about 0.11 smaller than the decline in the log annual earnings case. Most of the difference is between 1931 and 1961 (the year in which the cohort residual gap is normalized to zero in all of our analyses). The gap in hours per week among full time workers in the early cohorts documented in Figure L.4 is likely a key factor, both because of the mechanical link between hours worked and earnings at given wage rate and because earnings are a convex function of hours (Gicheva (2013), Goldin (2014)). For the log annual earnings decomposition, the cohort residual gap rises from 1961 to 1975, peaking at 0.014, before falling back to around zero by the 1984 cohort. In contrast, for the log hourly wage decomposition, the residual cohort gap continues to rise after 1975, reaching 0.022 by 1984.

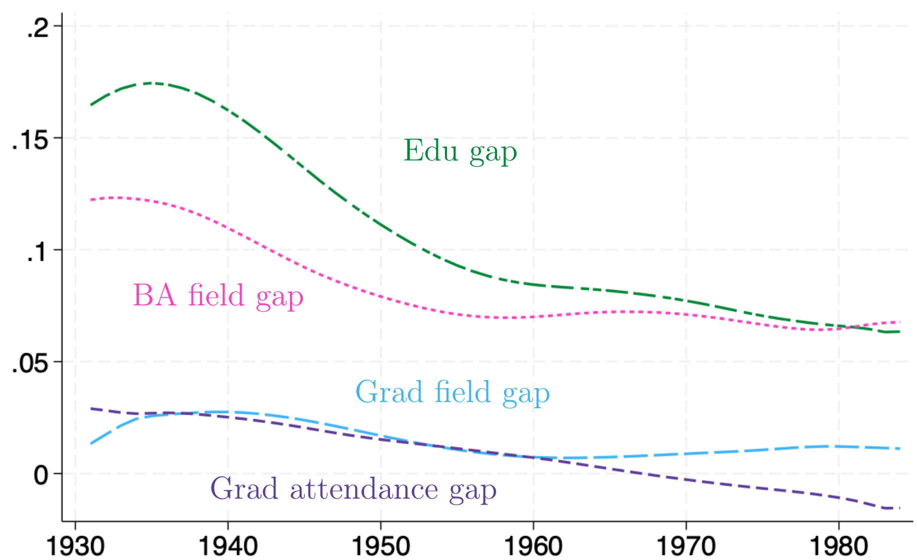
When allowing relative returns to vary across birth cohorts for log hourly wage earnings, the total gap starts at 0.531 and declines rapidly by an average of 0.013 per year, until reaching 0.278 in 1950. This steep decline is again largely attributed to the decline in the cohort residual gap. After 1950, the decline becomes more gradual, dropping to 0.217 by 1969 (around 0.004 a year from 1951). The gap then oscillates around 0.21 until it ends at 0.211 in 1984.

Looking at the components of the total gap, we see similar patterns compared to the dynamic decomposition using log annual earnings. The cohort residual gap declines quickly, dropping from 0.235 to zero by 1961. The education gap and the roles of undergraduate field, graduate attendance, and graduate field are very similar in magnitude and trends across birth cohorts. The most noticeable difference is in the return gap. Panel B of Figure K.2 shows the varying portion of the return gap starts at 0.026 dropping quickly to its lowest point at -0.017 by 1943. After 1943 it gradually climbs back to zero by 1961, where it remains zero. This is in contrast to the varying returns of log annual earnings, which oscillates at zero from 1931 to 1984.

Figure K.1: Decomposition of Log Hourly Wage Earnings, Constant Returns



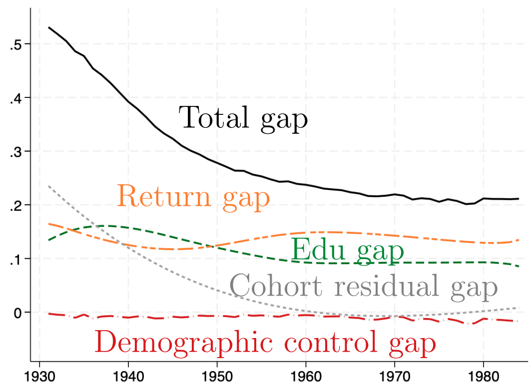
(A) Total gap



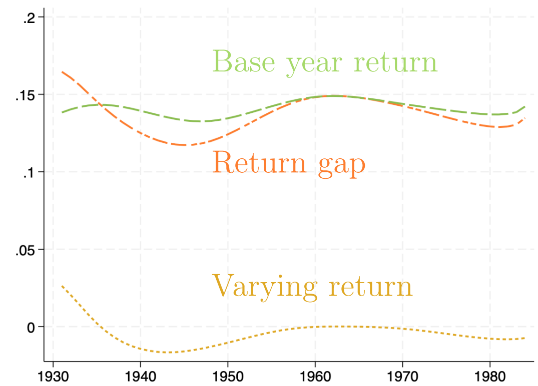
(B) Education gap

Notes: Panel A shows the decomposition of the predicted gender gap in log hourly wage earnings for each birth cohort averaged from age 28 to 52. The construction, description, and meaning of the lines are identical to those in Figure 2, with log hourly wage earnings as the dependent variable instead of log annual earnings. The estimates are constructed using OLS estimates of equation (4). The NSCG base year samples are used with cross sectional weights. Ages are restricted to be between 23 and 59. Panel B shows the decomposition of the education gap based on equation (5). The green line is the education gap, copied from panel A. The pink line shows the contribution of college majors, the purple line shows the contribution of graduate attendance, and the blue line shows the contribution of graduate degree field conditional on college major.

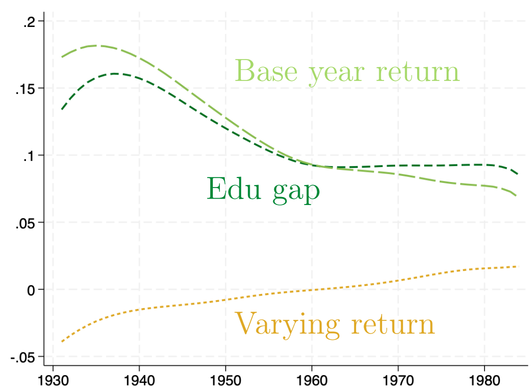
Figure K.2: Decomposition of Log Hourly Wage Earnings, Cohort Specific Relative Returns



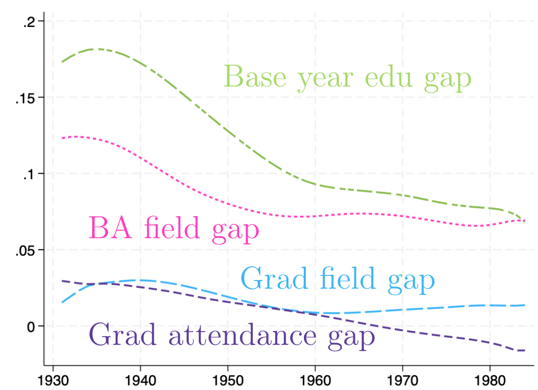
(A) Total gap



(B) The role in the return gap of base year and cohort varying relative returns



(C) The role in the education gap of base year and cohort varying relative returns



(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: Panel A shows the predicted gender gap in log hourly wage earnings for each birth cohort averaged from age 28 to 52. The description of each line and panel is identical to Figure 4. The coefficient estimates are from regression model (10). Panel B decomposes the return gap, $\sum_{cg}\left(\alpha_{cg}^{m0}+\delta_{cg}^{mb}-\alpha_{cg}^{f0}-\delta_{cg}^{fb}\right)P_{cg}^{mb}$, into two components. The light green line uses the base year return α_{cg}^{m0}, so it is comparable with the orange in Figure K.1. The yellow uses the gender specific, cohort varying relative return δ_{cg}^{mb}. Panel C decomposes the education gap, $\sum_{cg}\alpha_{cg}^{mb}\left(P_{cg}^{mb}-P_{cg}^{fb}\right)$, into two components in the same way as panel B. They sum up to the green line, the education gap. Panel D decomposes the base year return education gap into three components, the contributions of undergrad field in pink, grad attendance in purple, and grad field in blue. They sum up to the education gap with base year return.

L The Role of Selection

The NSF data we use does not contain measures of skills or abilities, such as standardized tests. As such, our statistical decompositions of the gender earnings gap do not account for changes in the composition of college graduates working full-time. Since we study the gap, differential changes in selection between genders are particularly important. Here, we focus on three specific margins of selection: (1) selection into college graduation, (2) selection into full-time work among college graduates, and (3) selection into graduate degree attainment among college graduates. Specifically, we focus on selection based on standardized test scores. While test scores do not capture the complete set of skills and capabilities individuals sort on, they are the only feasible measure of skill or ability we can compare across birth cohorts.

We use five different sources of data. First, we use average quantitative reasoning and verbal reasoning scores for college-bound seniors over time, as reported by the College Entrance Exam Board. Our other four datasets are longitudinal studies that include standardized achievement or cognitive ability tests and follow individuals into adulthood. Our earliest longitudinal study is Project Talent (PT), a nationally representative survey of 5% of high school students in 1960. We use test scores administered as part of the survey, and outcomes are measured 11 years after graduation with a modal age of 29. Next, we use the National Longitudinal Study of 1972 (NLS72), a survey of high school seniors in 1972 with periodic follow-up surveys through 1986, when the modal age of respondents was 32. Our third longitudinal survey is the National Longitudinal Survey of Youth 1979 (NLSY79), a survey of 14 to 22-year-olds in 1979 with ongoing follow-up surveys. Lastly, we use the National Longitudinal Survey of Youth 1997 (NLSY97), a survey of individuals who were 12-16 years old at the start of 1997.

Starting with SAT scores, Figure L.1 plots verbal and quantitative reasoning SAT scores over time compiled by the NCES from annual College Entrance Examination Board reports for college-bound seniors. The top panel shows average quantitative and verbal scores for men and women, while the bottom panel shows the female difference in average verbal and quantitative scores. The x-axis on both plots is birth cohort, assuming exams are taken at age 17. The y-axes are test score and score difference, respectively. The gap in quantitative scores is about 40 points for the 1949 birth cohort, which rises to 45 points over the 1957-59 birth cohorts. After 1959, there was a gradual decline in the gap to 35 points for the 1998 birth cohort. The gap in verbal scores started at -5 for the 1949 birth cohort, increasing over time to 12 points in 1970, after which there was a gradual decline to 2 points for the 1998 birth cohort. These trends are modest, given that the standard deviation in test scores in most years is around 100 points. Yet, the trends suggest gaps increase early on (until the 1959 birth cohort for math and the 1970s birth cohort for verbal), followed by gradual declines (15 points for math and 10 points for verbal).

Next, using our four longitudinal surveys, we consider how selection into college graduation, full-time work, and graduate degree attainment has changed by gender based on test scores across cohorts. Figure L.2 plots the average test score percentile among these three groups, where our full-time work and graduate degree samples are restricted to college graduates. For all the panels of the figure, the x-axis is birth cohort and the y-axis is the average test score percentile. The percentiles are calculated using survey weights and should be nationally representative for the full population. Panels A, C, and E plot the average percentile score for male (blue line) and female (orange line) for each group, while panels B, D, and F plot the corresponding group difference between males and females. Each line has four data points positioned on the x-axis corresponding to the approximate birth cohort of each of the four data sets we study. Additionally, Table L.1 reports all graphed averages and differences plotted in Figure L.2.

Beginning with college graduates, Figure L.2 panel A shows that college graduates had an average test score percentile between 65 and 75 in all four datasets.[42] In the PT data, the gap slightly favors women by around three percentiles, but this gap is flipped in favor of men in the NLS72 data and persists with gaps favoring men of 7 percentiles in the NLSY79 and 5 percentiles in the NLSY97. The estimates suggest there may have been a small widening in the gender gap in test scores among college graduates from the mid-1950s to the early 1960s. This widening gap broadly corresponds to a closing and reversal of the gender gap in college graduation, with the fraction of women going to college rising much more rapidly for women than men from the 1950 to 1980 birth cohort (Patnaik et al., 2020). The top two panels of Figure L.3 plot the share of males and females with a college degree (left panel) and the difference between males and females (right panel) by birth cohort, visualizing the more rapid growth of female college graduates.

Looking at men and women working full time, Panels C and D of Figure L.2 show that the gender gap in test scores among college graduates working full time increased substantially from the PT to the NLSY79 and then shrunk some from the NLSY79 to the NLSY97. The difference between test scores increases approximately linearly by ten percentiles from PT (1943 birth cohort) to the NLSY79 (1961 birth cohort), corresponding to a period when the share of women working full-time was increasing.[43] Figure L.4 panel

[42] One limitation of our data is that we rely on different cognitive or achievement tests in each survey, which may measure different combinations of skills and abilities.

[43] Note that we do not observe labor supply at the same age in every dataset. In particular, we observe individuals in the PT data around age 29 and individuals in the NLS around age 32. These small differences

A documents this trend, showing that the gender gap in full time work between male and female college graduates nearly halved between the 1945-1949 and the 1965-1969 birth cohorts. From the NLSY79 to the NLSY97 (1982 birth cohort) the gender gap in average test score percentiles among college graduates shrinks by three percentiles.

Another concern is that men working full time may still work more hours than women, and this difference may have changed over time. Figure L.4 panel B plots the difference in average hours worked per week between men and women among those who report working full time by age (x-axis) and cohort (line color). We see that women who report working full time work fewer hours across all cohorts and ages, with the difference peaking around age 35, where the gap is between 3 and 4.5 hours depending on the cohort. The gaps tend to be smaller at all ages for the more recent 1975-1979 and 1985-1989 cohorts.

Moving onto the last group, panels E and F of Figure L.2 plot selection into earning a graduate degree by gender over time. Test score percentiles are similar for the PT data, which grows to a three-percentile gap in the NLS72 data. The gap then grows to 11 percentiles in the NLSY79 and then shrinks to 6 percentiles in the NLSY97. The large increase in the gap between the 1954 (NLS72) and 1961 (NLSY79) birth cohorts comes almost entirely from a decrease among women. This decrease is consistent with the large increase in graduate degree attainment (conditional on having a BA) among women relative to men in this period, which is plotted in the bottom two panels of Figure L.3.

Overall, the figures above show that there are moderate changes in the relative test scores of male and female (1) college graduates, (2) college graduates working full time, and (3) graduate degree holders over time. The most notable changes are increases in the gender gap between the 1942 (PT) and 1961 (NLSY79) birth cohorts, which also broadly corresponds to a time when there was a rapid increase in the share of women graduating from college, earning graduate degrees, and working full time. Together, the evidence suggests the average test scores of female college graduates working full time fell between the late 1940s and early 1960s birth cohorts as the share of women graduating from college, earning graduate degrees, and working full time all increased. Since we cannot control for test scores in our primary analysis, we miss this trend, which may explain part of the residual gender gap in earnings.

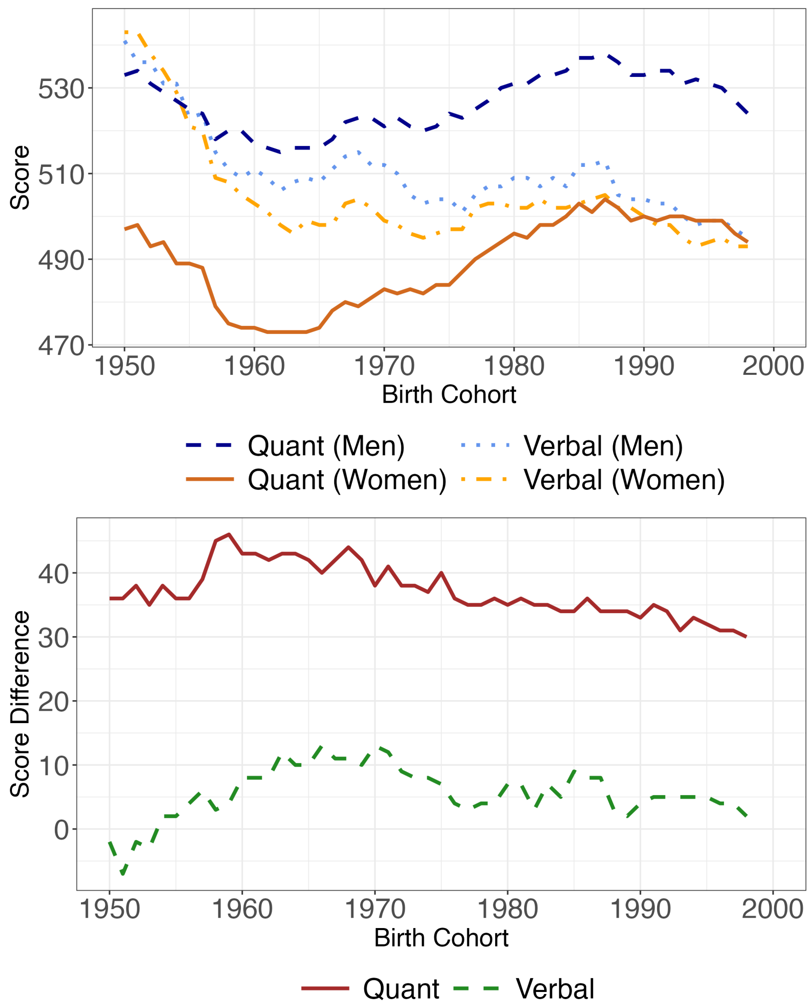
may affect differences between datasets, in particular for women who may be more likely to work part time when there is a young child in the household.

Table L.1: Selection into Undergraduate and Graduate Degree Attainment on Test Scores by Gender

Birth	With BA			With MA			Working FT		
Cohort	Male	Female	Diff	Male	Female	Diff	Male	Female	Diff
1943	76.614	79.598	-2.984	81.822	82.331	-0.509	77.216	79.217	-2.001
1954	74.426	71.196	3.230	79.868	77.073	2.795	74.529	70.095	4.435
1961	72.882	65.559	7.323	79.824	68.667	11.157	73.927	65.547	8.380
1982	70.967	65.893	5.073	77.516	71.378	6.138	71.596	66.504	5.092

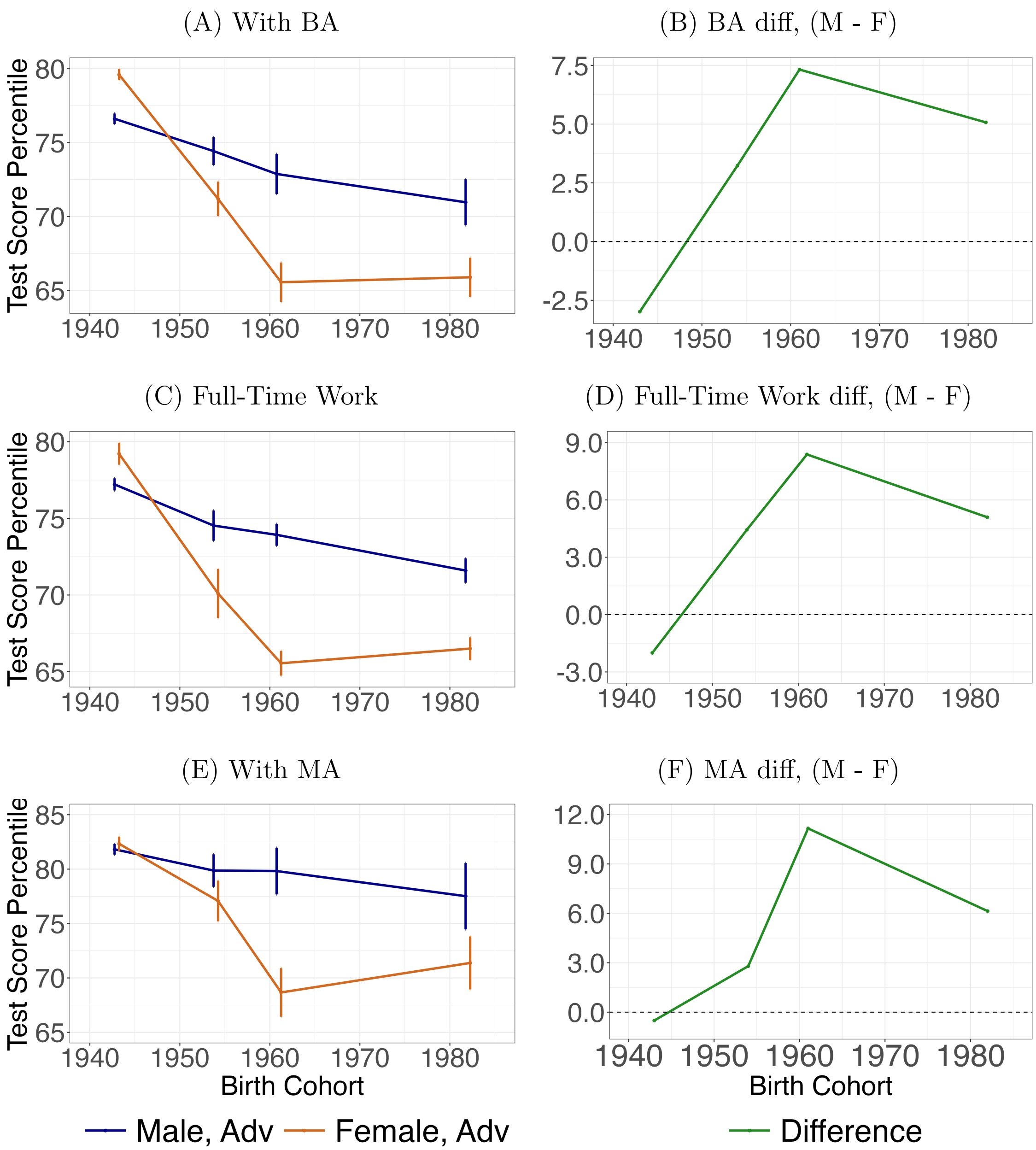
Notes: This table shows the average test score percentile for males, females, and their difference for those who selected into: college graduation (BA), full-time work (working FT), and graduate degrees (MA). Those who selected into full-time work and graduate degrees are restricted to college graduates. The table uses data from Project Talent, The National Longitudinal Study 1972, the National Longitudinal Survey of Youth 1979, and the National Longitudinal Survey of Youth 1997. Birth cohort is assigned based on the average age of respondents in the surveys; from top to bottom, the data comes from Project Talent (1943 birth cohort), NLS72 (1954 birth cohort), NLSY79 (1961 birth cohort), and the NSLY97 (1982 birth cohort). Test score percentiles are the weighted percentile of achievement or cognitive tests given in each survey.

Figure L.1: SAT Scores for Men and Women Over Time



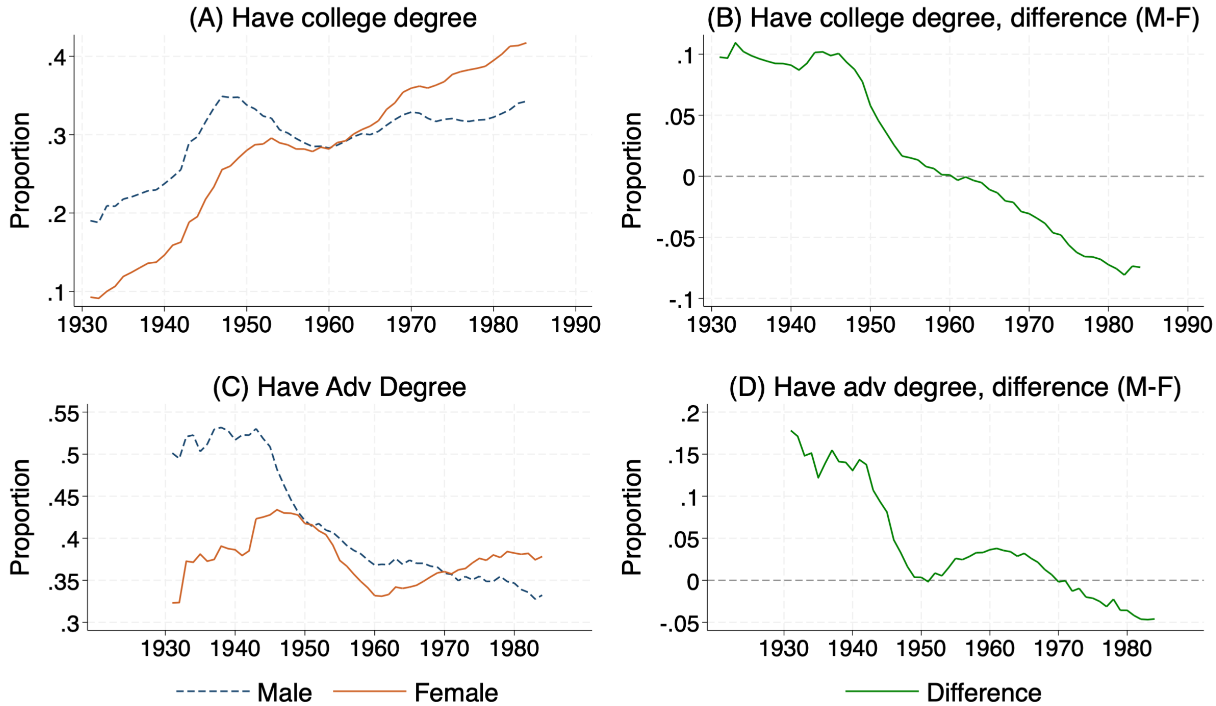
Notes: This figure plots the average SAT scores on the quantitative and verbal sections for college-bound high school seniors who took the SAT. Data is from the College Entrance Examination Board annual reports, which was then collected and harmonized and reported by the NCES in the Digest of Education Statistics (2019, Table 226.20).

Figure L.2: Selection into College Graduation, Full-Time Work, and Graduate Degree Attainment on Test Scores by Gender



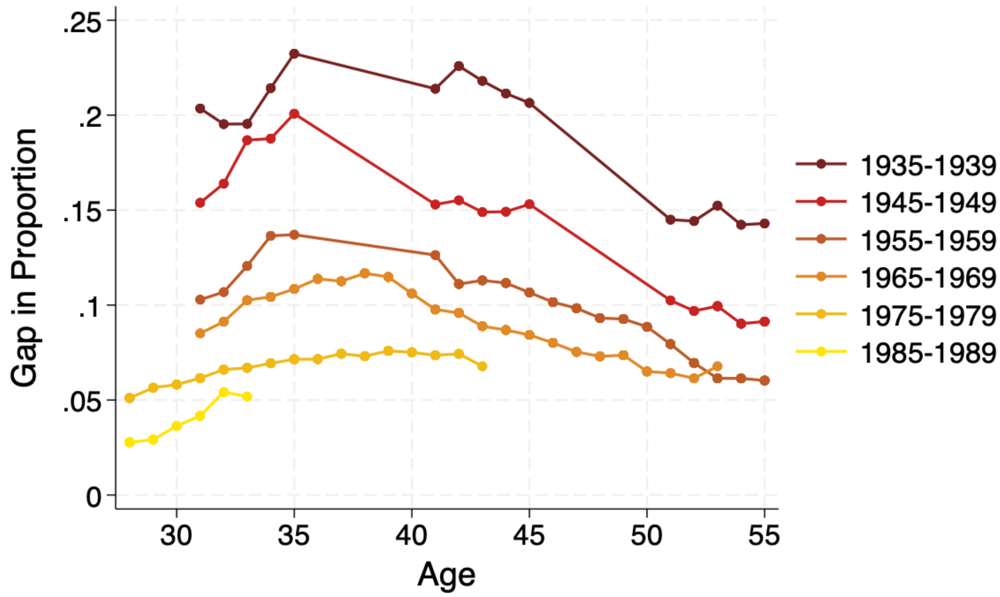
Notes: Panels A, C, and E show the selection into college graduation, graduate degree attainment, and full-time work on test scores by gender, respectively. Panels B, D, and E show the difference between the two genders for each respective left-hand panel. The graduate degree attainment and full-time work estimates are conditional on receiving a BA. These figures use data from Project Talent, The National Longitudinal Study 1972, the National Longitudinal Survey of Youth 1979, and the National Longitudinal Survey of Youth 1997. Birth cohort is assigned based on the average age of respondents in the surveys; from left to right the data points on each graph come from PT, NLS72, NLSY79, and the NLSY97. Test score percentiles are the weighted percentile of achievement or cognitive tests given in each survey.

Figure L.3: Proportion of Birth Cohort by Gender that have College and Advanced Degrees

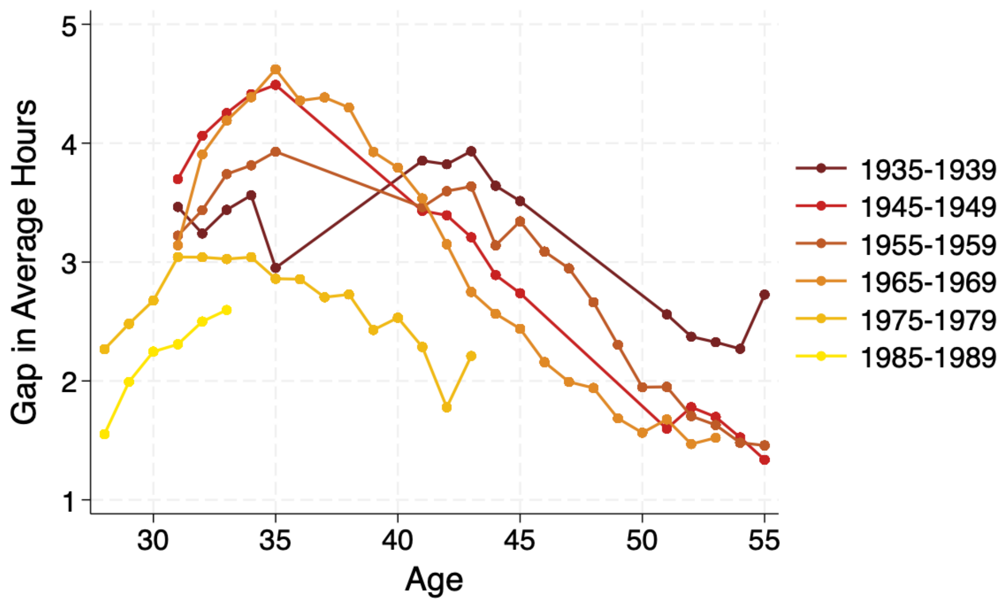


Notes: This figure uses data from the ACS (2001-2018) and the decennial Census (1960, 1970, 1980, 1990, 2000). Panel A shows the proportion of the birth cohorts that have an undergraduate degree disaggregated by gender. Panel B shows the difference (male - female) of panel A. Panel C shows the proportion of the birth cohorts that have advanced degrees disaggregated by gender. Panel D shows the difference (male - female) of panel C. The blue lines show the proportions for men, the orange lines show the proportions for women, and the green lines show the difference in the proportions. All panels show information for the birth cohorts from 1931-1984.

Figure L.4: Gender Gap in the Proportion and Average Working Hours of Full-Time Workers and Among Those with a BA Degree



(A) Gender gap of the proportion of full-time workers among people with BA degrees



(B) Gender gap of the average working hours of full-time workers with BA degrees

Notes: Panel A shows the difference in the proportion of the population, conditional on gender, that work full time, conditioned on holding a BA degree. Panel B shows the difference in the number of hours worked conditioned on working full time and holding a BA degree. The data comes from 1960-2000 decennial Census and the 2001-2018 ACS.