Decomposing Trends in the Gender Gap for Highly Educated Workers*

Joseph G. Altonji*, John Eric Humphries**,
Yagmur Yuksel***, and Ling Zhong****†

October 23, 2024

Abstract

This paper studies the gender gap among full time college educated workers born between 1931 and 1984. Using rich data from the National Survey of College Graduates and other sources on college graduates and their labor market outcomes, we decompose trends in the gender earnings gap across birth cohorts into trends due to differences in the relative returns to undergraduate and graduate degree field combinations, trends in gender differences in undergraduate field, graduate degree attainment, and graduate field, and trends in a cohort and gender specific "residual component" that shifts the gender gap in earnings by the same amount for all college graduates. We have three main sets of findings. First, we find that much of the large gap in earnings between the 1931 and 1950 cohorts is due to the "residual component". Most of the decline is within occupation, especially for the early cohorts. The residual gap varies little from 1951 to the late 70s, after which it resumes its decline. Second, we find that gender differences in the relative return to undergraduate and graduate degree combinations matter for the gender gap, but contribute very little to the decline in the gender gap over the full time period. Third, we study and further decompose the "education gap"--the contribution of college major choice, graduate degree attainment and graduate field to the gap. When evaluated at fixed relative returns to each degree type, the education gap declines substantially and is an important part of the narrowing of the gender gap. But this decline is largely offset by cohort trends in the relative returns to specific fields that worked in favor of men against women. Overall, the education gap varies in a narrow range around 0.2 and accounts for very little of the decline.

*Preliminary and Incomplete. We are grateful to Lauren Harris, Lucas Marron, Hoon Pyo Jeon and Owen Rask for outstanding research assistance. 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. We thank Edwin Lueven, participants in seminars at ANU, Deakin, ITAM, Monash, Northwester, the University of Hong Kong, the University of Melbourne, Zhejiang University, and participants in conferences organized by the AEA, Barcelona School of Economics, CESifo, CReAM UCL, NCER-CCER for helpful comments. All statistics in the paper that are based on that data went through disclosure review. All errors and shortcomings are our responsibility.

†*Yale University, NBER and IZA, **Yale University and NBER, ***Zelus Analytics, ****Cheung Kong Graduate School of Business and IZA

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 and fewer labor force interruptions. 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 past work has 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 education 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.

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

[1] The Decennial Census and the Current Population Survey (CPS), primary data sets 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

earnings observations back to only 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, such as assuming that age and cohort effects are additively separable and that time trends prior to 1990 are the same for men and women. 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 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]

After discussing the data in section 2, in section 3 we set the stage by providing 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% in 1935 and still 26.7% in 1947. For men, Business and Engineering together accounted for more than 44% of college graduates for the 1935 cohort, dropping to 32.3% 1949. During this period men moved into lower paying majors such as Humanities, Political Science, the Social Sciences, and the Arts. Both these trends contribute to the large drop in the earnings gap we document below, not just changes in what women studied.

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 10% in 1950 to 22% in 1962.. These changes reduced the gender gap, but they were partially offset by the fact that men moved from Education and Humanities back toward Engineering and Business, and moved toward Computer Science and Math, which became the third most popular category for men.

From the late 1960s through 1984 birth cohorts, the fraction of women majoring in Business declined to 12.2% at the end of the period, while the percentage of women in Psychology, Biology and other Social Sciences (rel-

2009, but this is too late for the early cohorts. It also lacks information on graduate field.

[2]We investigated two other data sets: 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 base year survey included a subsample from a broader set of occupations.

atively low paying majors) grew to 25% by 1985. The percentage of men majoring in Computer Science and Math rose from 6.3% in 1960 to 10.7% in 1984.

Overall, the trends in college major are much more complicated than a monotonic shift of women toward higher paying fields previously dominated by men. The most popular fields shifted by large amounts for both men and women over the long period that we study, both toward and away from the higher paying fields. But the overall trend is that on net, women shifted toward higher paying majors relative to men, with most of the change occurring prior to the early 60s birth cohorts.

Graduate degree attainment also increased dramatically for women relative to men. The percentage of women who obtained a graduate degree rose from 10% of college graduates for the 1932 birth cohort to almost 40% for the 1984 cohort. Women started 15 percentage points behind men and ended up about 6 points ahead. This 21 point shift underlies our finding that changes in graduate degree attainment contribute about 0.05 to the narrowing of the gender gap.

The composition of graduate fields for men and women also changed substantially. Between the 1950s and 70s birth cohorts, the shares of women choosing Business and Professional Degrees, including MD and Law, increased to catch up with men. More women choose MBA or Engineering, but as men's shares also increased, the gender gap in the proportions persists.

To get an initial sense of the role of college major and graduate degree attainment in changes in the gender gap over time, we use a regression model with a dummy for male, interacted indicators for 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 1/3 for the early cohorts and 1/4 for the later cohorts. Additionally controlling for graduate field matters less. Adding additional controls for occupation further reduces the gap, consistent with the findings of Sloane et al. (2021) for undergraduate degrees.[3]

Section 4 presents the methodology that we use to decompose cohort trends in the gender gap for full time workers. For earnings, we 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 includes gender interacted with college major specific age polynomials, parental education, and gender specific race and Hispanic origin dummies. We

[3]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.

also control for flexible interactions between age and birth cohort. To account for the fact that we see cohorts at different ages, we use a regression adjustment to normalize our results to refer to the simple average of annual earnings between the ages of 28 and 52. 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. The constant relative returns case simplifies the interpretation of the trends in the gender gap. We then turn to a specification that allows relative returns to vary across cohorts.

We have five main results in section 5. The first result is 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 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. From the mid fifties forward, there is little change in the residual gap. 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.

How much of the decline in the residual gap occurred because women, conditional on education, moved to better paying occupations relative to men? Combining decompositions of the gender gap in the component of earnings related to occupation with the earnings gap decompositions, we find that in 1931 86% of the cohort residual gap is within occupation. Both the within and across occupations residual gaps decline to zero by 1953 (relative to 1961). Thus much of early steep decline in the birth cohort specific residual gap is within occupation.

The second result is that gender differences in the relative return to undergraduate and graduate degree combinations contribute very little to the decline in the gender gap.[4] This is true even though we see large 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. About 75% of the relative return gap is within occupation, with little variation across cohorts.

The picture is similar in the variable returns specification. With that spec-

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

ification, 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.

Our third set of results concerns the "education gap" contributed by gender differences in college major, graduate degree attainment and graduate field. They 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 then begins to fall after the 1936 cohort. The decline is particularly rapid between the 1940 and 1952 birth cohorts, averaging at 0.005 per year, arriving at 0.124 for the 1952 cohort. The education gap continues to decline between 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 varies in a narrow range around 0.20 across cohorts and accounts for very little of the decline in the gap between the early 1930s and the early 1980s. It equals 36% of the cross cohort average of the total gap and 56% of the total gap in 1984. The constancy of the gap sharply 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 constancy of 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 2019, 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).

The fourth set of results concerns the role of specific college majors in the narrowing of the education gap. Using the constant returns specification, we measure the contributions of each of 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 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 a number of fields, with Business, Biology, Nursing, and Health contributing to the decline and Engineering and Computer Science/Math working in the opposite direction.

The final set of results measures the contribution 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 contributes only 6%.[5]

This paper relates to a large literature studying the gender gap and the role of college majors in contributing to this gap. This literature is reviewed in Altonji et al. (2012), Altonji et al. (2016), and Patnaik et al. (2020). Altonji et al. (2012) documents trends in college majors by men and women for those who graduated in the mid 1970s to 2008, showing that the women were increasingly pursuing science and business related degrees, and decreasingly pursuing liberal arts, English, and education degrees.[6] They conduct a Blinder-Oxaca decomposition by graduation year, showing that differences in coefficients explain approximately half of the male female gap in hourly wage rates, while differences in weights across college degrees and final education 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. Altonji et al. (2016) further build on Altonji et al. (2012) by showing how men and women differ in their acquisition of graduate degrees. In a more recent review, Patnaik et al. (2020) provide further discussion on the change in rates

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

[6] In this paper, we do not discuss causes behind the changes in major choice, graduate school attendance, and graduate field that we observe. See Bronson (2019), Zafar (2009), Gemici and Wiswall (2014) and Abramitzky et al. (2022) for discussion of mechanisms.

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, and a small increase in STEM majors for women from 1950 to 1960, as well as an increase from the 1985 to 1993 birth cohorts. These papers and the papers they cite build on a large prior existing 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). Blau and Kahn (2017) similarly documents trends in gender gap. We provide gender specific estimates of trends in college major choice, graduate degree attainment, and graduate field for the 1931-1984 birth cohorts, and thus can consider both undergraduate field and graduate education, which has grown dramatically.

The paper most closely related to our own is Sloane et al. (2021). This paper uses 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 that major. They find that these differences narrow across birth cohorts, but that women still choose majors and occupations with lower expected earnings (for men). As this paper uses ACS data, they are not able to study trends as far back in time or consider graduate fields of study or degrees, but are able to more exhaustively consider the interactions between major and occupation given the large sample sizes in the annual ACS surveys.

2 Data

The paper uses three datasets. The National Center for Science and Engineering Statistics' National Survey of College Graduates (NSCG) is our primary data source. Building upon the data construction described in Altonji and Zhong (2021), we add NSCG 2017 and 2019 and make a few adjustments to make the data a representative sample of people with college degrees.[7] We

[7] We exclude people who fall into one of NSF's Science and Engineering Statistics-eligible education categories but entered the NSCG sampling frame through a Science and Engineering Statistics-related occupation.

use surveys that are nationally representative, namely the constructed NSCG data from 1988, the matched Census data from 1990, and NSCG data from 1993, 2003, and 2010 to 2019. In most of the empirical analyses we restrict the sample to people born between 1931 and 1984.

The second data set 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 people with 4 or more years of college in the 1960 to 1980 sample, and people with a bachelor's degree or higher in the sample after 1980. In the empirical analysis we deflate income to 2013 dollars, and 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. 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 we use the NSCG as 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.[8] 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 of the paper.

[8]The summary statistic tables from the 1949 to 1966 Degrees and Other Formal Awards Conferred biannual survey report the number of diplomas conferred for 3 broadly defined college majors and 5 broadly defined graduate fields of study.

[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.

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 A.1 and A.2, we show the college major trend for 15 majors and 19 graduate fields.[10] The two vertical dash lines separate the birth cohorts from 1931 to 1984 into three periods in which the patterns of the trends vary. We discuss the trends based on Figures 1 and A.1.

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 throughout 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 have declined to 22.1% and 10.8% in 1949, respectively. Between the 1935 and 1950 birth cohorts, men were diversifying 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% over the same period. It overtook Education as the most popular major for women in 1958. Meanwhile, men saw renewed growth in Engineering and Business and sustained growth in Computer Sci-

[10]There are 4 majors we cannot show, because the cell count for either men or women are too low in the early birth cohorts. We merge together Nursing and Health Administration graduate fields for male graduates for similar reasons.

ence & Math, which emerged as the third most popular major for men in the 1960s. The once-steady male participation in the Education and Humanities BAs declined to 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 quickly went beyond making up for this decline, while men's rates stagnated until 1970 before seeing growth again in the late 1970s and throughout the 1980s.

3.2 Trends in gender differences in earnings and occupational sorting

We use a simple regression of log earnings on a male dummy and controls to begin our analysis on the trends in earnings of men and women with higher education. 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 1994. Column 1 includes the baseline controls, which consist of parents' education dummies, a cubic in age interacted with a male dummy, a cubic

[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)

in birth year, and race indicators interacted with gender. The age cubics are constructed such that the coefficients in the table capture the estimated residual earnings gap at age 35. We also use a regression index of interactions between and year of birth as a control. The index is estimated using the Census/ACS data, as discussed in 4.3.2.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), one group at a time. The nested regression specifications allow us to identify the contribution of each set of control to explaining the gender gap. All regressions compare individuals with college degrees working full time.

The baseline regression coefficients shows that, without education or occupation controls, the estimated log earnings gap is 0.54 (0.023) for the first birth cohort group. The gap declines to 0.46, 0.37, and 0.35 for the second, third, and fourth birth cohort groups, respectively. The column 2 shows that controlling for college majors shrinks the gender gaps for the four cohort groups to 0.35, 0.31, 0.26, and 0.26 respectively. This is a 35% reduction in the earliest cohort and a 26% reduction in the fourth (1964-94) cohort group. The difference between column 2 and column 3 shows that graduate degree attainment and field explain another 0.015 to .022 of the gender gap for the first three cohorts, but only 0.005 for the (1964-94) group. This result shows 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 0.05 to 0.29, 0.24, 0.19 and 0.20, respectively. This suggests that occupational sorting also plays an important role in the gender pay gap of highly educated workers, consistent with prior work.

4 Birth cohort specific decompositions of the gender gap in earnings

This section develops our decomposition of the gender gap in earnings. In 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 returns to degrees are constant across cohorts. Sections 4.3 and 4.4 discusses estimation of the key inputs into the decomposition: the earning 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 $\bar{y}_{o(it)}^{ba(it)}$ and a within occupation component $\tilde{y}_{it}$:

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

We define the occupation specific components to be the same for men and women. Variation across genders within occupation is part of $\tilde{y}_{it}$. We return to this point below. We consider a specification in which they are constant across birth cohorts, and a specification in which they vary with birth cohort and age.[12]

The regression equation for log earnings is

$$Y_{it} = \alpha_{cg}^{sb} + X_{1it}^{s}\beta_1^s + X_{2it}^{s}\beta_2^s + Z_i^s\Gamma^s + u_{it}. \tag{1}$$

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

In our "constant returns" 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

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

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

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 and 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 age_{it} up to the second order plus $b_i^3 \times a_{it}$ and $b_i \times a_{it}^3$, all interacted with $S_{s(i)}$. We include it because age profiles are likely to vary with birth cohort. Such variation means that 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 evaluate the gender specific age polynomial terms in X_{1it}^s and X_{2it}^s at their average levels between age 28 and 52.[13] Note that because calendar time $t = b_i + age_i$, our measures account for secular shifts 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 affecting all cohorts 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 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 $\alpha_{cg}^{s0} + \alpha^{sb} + Z_i^s\Gamma^s$.

We assume that the expectation of $\bar{y}_{o(it)}^{ba(it)}$ conditional on a_{it}, t, b, cg, b, race/ethnicity and s is

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

where $\bar{u}_{it}$ is the error term. 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

[13] 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 using 26-59 (not reported).

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. They combine the regression model parameters with estimates of the degree probabilities. The goal is to isolate the contribution to the gender gap in earnings coming from: (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. In this section, we provide a decomposition 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 that gender gap to be

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

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 $Pr\left(c(i) = c, g(i) = g \mid s, b\right)$. Then given equation (2) one may write $GAP(b)$ as

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

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

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

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. 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 return 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, 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, difference in the probability of graduate school attendance conditional on college major, and the distribution of graduate field condition on graduate school attendance and undergraduate major. Because G_i is the deterministic function $1(g_i > 0)$ of g_i, we have

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

Simplifying the notation, the formula is

Education Gap (b) =

$$\begin{aligned}
&\sum_{cg} \alpha_{cg}^{m0} \times \Big(\Pr{}_b^m(g,G|c) - \Pr{}_b^f(g,G|c)\Big) \times \Pr{}_b^f(G|c) \times \Pr{}_b^f(c) && \text{grad field gap} \\
&+ \sum_{cg} \alpha_{cg}^{m0} \times \Pr{}_b^f(g|G,c) \times \Big(\Pr{}_b^m(G|c) - \Pr{}_b^f(G|c)\Big) \times \Pr{}_b^f(c) && \text{grad enroll gap} \quad (5) \\
&+ \sum_{cg} \alpha_{cg}^{m0} \times \Pr{}_b^f(g,G|c) \times \Big(\Pr{}_b^m(c) - \Pr{}_b^f(c)\Big) && \text{BA field gap} \\
&+ \Delta ED_b^{23} && \text{approx. error}
\end{aligned}$$

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_{se}^{23} is an approximation error that turns out to be negligible.[14]

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 turning to earnings. We describe how we estimate the age-birth cohort interaction term, discuss our use of OLS. Finally, we present evidence on rends 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_c^{fb} 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 $P_{g|c}^{fb}$ by fitting a b-spline to $G_{g(i)}$ using females who majored in c. $P_{G|c}^{fb}$ is $(1 - P_{g=0|c}^{fb})$, 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 $P_{g|G,c}^{fb}$ and $P_{g,G|c}^{fb}$. We use the same approach for males. We employ sample weights in this analysis, which is important because the NSCG overrepresents STEM fields. Appendix F.2 shows that results are similar when we estimate P_c^{fb} and P_c^{mb} as a three year moving average of b,s specific major probabilities.

[14]It is the sum of terms involving second order and third order interactions among the gender differences in $Pr(c)$, $Pr(G|c)$ and $Pr(g|G,c)$.

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

The age-birth cohort interaction term $X^s_{2it}\beta^s_2$ is identified in the NSF data for the early and late cohorts only by function form. To see this, note first that our regression sample is restricted to graduates between ages 23 and 59. The year of our earnings observations ranges from 1989 to 2019. Consequently, individuals born in the 1931 are only observed at age 58, while those born in 1984 are only observed between ages 23 to 35. While functional form restrictions mean that the profiles are technically identified, we would be relying heavily on functional form to extrapolate to combinations of b_i and age_{it} that are far outside of our sample. 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

$$Y_{it} = \sum_{s\in\{f,m\}} \left[a^s_0 S_{s(i)} + a^s_1 S_{s(i)} G_{it} + X^s_{1it}\beta^{s*}_1 + X^s_{2it}\beta^{s*}_2 + Z^{s*}_i \Gamma^{s*}\right] + u_{it}, \qquad (6)$$

where G_i is an indicator for whether i has graduate education.[15] 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.

[15]For the 1960-1990 Census, $G_i = 1$ if the individual has five or more years of post secondary 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.[16] 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.[17] In the case of the return to graduate degrees conditional on undergraduate field, Altonji and Zhong (2021) and Altonji et al. (2023) assume the $\alpha^{s0}_{cg}{=}\alpha^{s0}_{c}+\alpha^{s0}_{g}$ and use a 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 individual who obtain a graduate degree to identify α^{s0}_{cg} , 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^{s0}_{g} (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 there is differential selection into our analysis sample over time. The NSF data does not contain

[16] 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^{f}_{cg} 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 replace the term $\sum_{g=0}^{\mathcal{G}} \alpha^{s}_{cg} S_{s(i)} C_{c(i)} G_{g(i)t}$ with the additively separable specification $\sum_{c=1}^{\mathcal{C}} \alpha^{s}_{c} S_{s(i)} C_{c(i)} + \sum_{g=1}^{\mathcal{G}} \alpha^{s}_{g} S_{s(i)} G_{g(i)t}$. We use $\hat{\alpha}^{s}_{c} + \hat{\alpha}^{s}_{g}$ as the estimate of α^{s}_{cg} for the c, g combinations with fewer than 30 observations in the decompositions below.

[17] See Altonji et al. (2012) and Altonji et al. (2016) for a discussion of the methodological challenges and surveys of empirical studies of the return to undergraduate field.

Altonji and Zhong (2021) provide a formal discussion of bias in the use of OLS to estimate the return to graduate degrees. Kirkebøen et al. (2014) provide a strategy for estimating return 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 up a minimum grade requirement the University of California Santa Cruz to estimate the return to majoring in economics and obtain results similar to OLS.

measures of ability and only includes college graduates. We therefore rely on Census and ACS data to study selection into college, and a combination of four data sets to study selection on ability into college, graduate school, and working full time. Here we briefly discuss the role of selection into our sample, which we develop more fully in Appendix J. We study selection into college and graduate school based on test scores using data from Project Talent (PT), the NLS72, NLSY79, and NLSY97. Each of these data sets 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 J.2 shows that the gap between male and female test scores grew between the PT cohort born in the early 1940s to the NSLY79 cohort born in the early 1960s, before shrinking moderately for the NLSY97 cohort born in the early 1980s. This increase corresponds to period when the share of men earning BAs is relatively stable, while the share of women earning BAs is increasing rapidly (See Appendix Figure J.3). We find similar sorting 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 1980 and 2015 using the PSID and the full population of workers.

Lastly, in Appendix J 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 we measure over time.

4.4 Allowing relative returns to degrees to vary across cohorts

The assumption that the gender-specific returns to college majors and graduate fields remain constant for birth cohorts ranging from the 1930s to the 1980s 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 c, g 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,

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

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} from the δ_{cg}^{sb}. 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

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

where $\bar{P}_{cg}^{f}$ and $\bar{P}_{cg}^{m}$ are the unweighted averages over b of the c, g 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

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

Decomposing the double sum, we have

$$
\begin{aligned}
GAP(b) = & \sum_{cg} \left(\alpha_{cg}^{m0} - \alpha_{cg}^{f0}\right) P_{cg}^{fb} && \text{rel. return gap, base year returns} \\
& + \sum_{cg} \alpha_{cg}^{m0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{education gap, base year returns} \\
& + \alpha^{mb} - \alpha^{fb} && \text{cohort } b \text{ residual gap} \\
& + \Delta Z_b && \text{demographic control gap} \\
& + \sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb} && \text{rel. return gap, varying returns} \\
& + \sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{education gap, varying returns}
\end{aligned} \tag{8}
$$

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. They are 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 captures the degree to which movements in $\delta_{cg}^{mb} - \delta_{cg}^{fb}$ tend to be more positive in fields that are common among for 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 in which relative returns change in the same way for men and women, with $\delta_{cg}^{mb} = \delta_{cg}^{fb}$ for all b and c, g, the sixth term captures the degree to which cross cohort changes in returns are higher in fields that are more popular among males than females. [18]

One can decompose both of the terms that comprise the education gap into the contributions of changes in the gender gap in P_{cg}^{sb} into the contribution of changes $P_{c}^{mb} - P_{c}^{fb}$, $P_{G|c}^{mb} - P_{G|c}^{fb}$ and $P_{g|c,G=1}^{mb} - P_{g|c,G=1}^{fb}$, using a formula analogous to equation (5) for the education 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 the δ parameters δ_{cg}^{mb}. [19]We discuss the decomposition of the Education gap, below, focusing

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

[19]The formula is provided in Appendix C, equation (10).

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.

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 $\bar{y}^{ba}_{o(it)}$. One way to do this would be to estimate a regression with fixed effects for the interaction between birth cohort, age and occupation, and then use these fixed effects as dependent variables based on the individual's occupation, age, and birth cohort. We use this option to supplement our main analysis rather than as our preferred approach.[20] 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. This will lead the birth cohort specific component of the occupation premiums to pick up the effects of c and g on earnings that work within 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 imposes the assumption that $\bar{y}^{ba}_{o(it)}$ does not depend on b and so can be written as $\bar{y}^{a}_{o(it)}$. [21] Using these estimates, we then construct $\bar{y}^{a}_{o(it)}$ as our dependent variable based on the individual's occupation and age.

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. First, we construct $X^s_{2it}\bar{\beta}^s_2$, the index of interaction terms between age a_{it} and b_i in equation (3), using our new dependent variable.[22] Second,

[20]Due to cell size limitations, we use occupation-birth cohort fixed effects and occupation specific third order age polynomials rather than fully interacting birth cohort, age and occupation.

[21]Specifically, 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, 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 refers to the simple average between ages 28 to 52.

[22]We substitute $\bar{y}^a_o$ 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). For the constant-returns case, we assume $\bar{\alpha}^{sb}_{cg} = \bar{\alpha}^{s0}_{cg} + \bar{\alpha}^{sb}$, while for the dynamic case we assume $\bar{\alpha}^{sb}_{cg} = \bar{\alpha}^{s0}_{cg} + (\bar{\alpha}^{sb} + \bar{\delta}^{sb}_{cg})$. 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 $\bar{\delta}^{sb}_{cg}$. For the constant returns case we use the modified versions of equations (4) and (5), and for the dynamic returns case we use the modifed versions of equations (8) and (10).[23]

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 ACS/census data because sampling error from these data sets 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 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. For simplicity, we start with decompositions that assume the relative returns to degrees are constant across birth cohorts, which we then relax. The second two subsections consider the gap in the occupation premium, and how much of the overall earnings gap this can explain. We again begin with the assumption that relative returns to degrees are constant across birth cohort, 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. 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

[23] In all specifications with age varying occupation premiums, we rescale the dependent variables 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.

rate of decline slowly falls until about 1953, when the gap is 0.333. After that the decline is much more gradual, averaging only 0.004 per year until 1978 birth cohort, when the gap is 0.273. Between 1978 and 1984 the gap declines by 0.002 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. It is constructed from our regression model and estimates of the postsecondary education outcome probabilities. It is not directly observable in the data.[24]

Next we turn to the relative return gap (Figure 2 panel A, orange line). The formula is $\sum_{cg}\left(\alpha_{cg}^{m}-\alpha_{cg}^{f}\right)\times P_{cg}^{fb}$. It includes the 1961 value of the gender gap component that affects all cg categories equally. It 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. The relative returns are constant, so the small amount of variation in this term is driven entirely by changes in the degree mix 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 return gap. 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 is a change in the gender gap across cohorts that is shared across all undergraduate and graduate degrees. It 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. The work of Goldin (2014, 2006, 2021), Lemieux (2006) Ruggles (2015), Blau and Kahn (2017), and many other contributors to the literature on long terms trend in the gender gap suggests changes in total labor experience at a given age, lower fertility, shifting gender norms and preferences affecting occupation choice, and reduced discrimination against women have

[24] In Appendix E, 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 Census/ACS based estimates of the path of $GAP(b)$ are similar to the NSCG based estimates.

all played a role.[25]

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, 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 overall wage gap from 1931 through the mid 1950s, while most 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 contribution of changes in college major, changes in graduate school attendance given college major, and changes in graduate field given graduate school attendance and college major to the education gap. The BA field gap (pink 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 Altonji et al. (2012)and Sloane et al. (2021)' analyses, which abstract from graduate education and do not consider cohorts before 1950.

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 BA field gap and 6% (0.007) for the Grad field

[25] As this paper focuses on college graduates that work full time, the composition of men and women in this sample over time may also play an important role in the evolution of the gender gap. Appendix J documents the rapid increase in college and graduate degree attainment and full time work for women during the time period we study and considers how this may affect selection into our sample. Using multiple data sets which contain test scores, Appendix J shows that the gap in test score between men and women increased between the 1940s and the 1960s birth cohorts, which broadly corresponds to when there was a rapid increase in the share of women graduating from college, earning graduate degrees, and working full time.

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 to the green line for the education gap shown in Figure 2.

The field specific results show that the decline in the education gap between 1930s and 1960s cohorts is 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, English, and Other Humanities that is offset by small increases in the contribution of Computer Science/Math, Engineering and Psychology. The decline after the late 1970s is due to a number of 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

In appendix F 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 B. The analysis is restricted to cohorts after 1944 because HEGIS starts in 1966. The results are broadly similar to the NSCG based estimates.

Second, we use a simple 3 year moving averages to estimate P_c^{fb} and P_c^{mb} rather than b-splines. This does not make much difference.

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 (8). 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. Using the gender-cohort-cg specific returns increases the estimate of the gender gap among highly educated workers in the early cohorts and reduces it 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. This component is common to all education choices. The residual gap is normalized to 0 in 1961. It starts at 0.336 in 1931 and decreases sharply from the 1930s to the late 1940s birth cohortss and then more slowly from 1950 to 0 in 1961. It remains near 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 accounts for about 36% of the gender gap on average and 56% of the total gap for the 1984 birth cohort. But in contrast to the constant returns case, it contributes very little to the decline in the gap.

5.2.1 Decomposing the Education Gap

We now show that the surprising relative constancy of the education gap across cohorts is the net result of two offsetting cohort trends. 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 c, g across cohorts (yellow line).[26] The education gap is reproduced as the green line. One can see that 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.

[26]Shifts in α^{mb}that are the same for all cg and therefore do not affect on the education gap.

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 out of Education, English and the Humanities, and into Business 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 dominated by women declined across cohorts. The two effects cancel out and generate the relatively constant cohort-specific education gap.

Using the estimates of the cohort specific component of the occupation premiums from 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) (not shown). 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

$$\begin{aligned}
&\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{Edu gap, varying return} \\
& && (9) \\
=&\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{m,avg} - P_{cg}^{f,avg}\right) && \text{Edu Gap, const. probability gap} \\
+&\sum_{cg} \delta_{cg}^{mb} \left(\left(P_{cg}^{mb} - P_{cg}^{fb}\right) - \left(P_{cg}^{m,avg} - P_{cg}^{f,avg}\right)\right), && \text{interaction term,}
\end{aligned}$$

where $P_{cg}^{s,avg} = \frac{1}{54}\sum_b P_{cg}^{sb}$ is a cohort invariant gender specific average probability of obtaining cg. Figure 5 panel E graphs the three terms. 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 $\sum_{cg} \delta_{cg}^{mb} \left(P_{cg}^{m,avg} - P_{cg}^{f,avg}\right)$ (dark blue dashed line) shows that about 0.05 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 $\left(P_{cg}^{mb} - P_{cg}^{fb}\right) - \left(P_{cg}^{m,avg} - P_{cg}^{f,avg}\right)$ (turquoise dashed line). Relative to men, women moved away from majors with rising returns.

In Figure 4 panel D, we focus on the education gap evaluated at the base year returns for men (α_{cg}^{m0}) (light green long dashed line) and decompose it by education choices using the analog of equation (5). Similar to the constant return specification, the gap is mainly due to gender differences in undergraduate field, with differences in graduate field contributing a smaller share. The contribution of graduate attendance falls from 0.004 in the 30s to -0.005 in 1984.

5.2.2 Decomposing the Return Gap

In Figure 4 panel B, we decompose the return gap (orange line). From equation (8), 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) is the component of the cohort-specific returns that is constant over birth cohort. Its stable magnitude is expected given our findings based on the constant returns specification. It shows that the value of the average return average across cg is about 0.2 and that the change in cg choices of women over birth cohorts does not change their average return relative to men by very much. 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 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.[27]

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

So far we have focused on decomposing the total gap. As described in Section 4.1, we can express the observed log earnings as an occupational earnings component and a within-occupation component, which we also refer to as the occupation premium. Here we study the role of occupation by decomposing just the occupation component rather than observed log earnings. Recall that differences in earnings within an occupation are part of the within-occupation component, so all differences here capture differences in occupations.

Figure 6 panel A shows the decomposition of gender differences in the occupation premium, $\bar{y}_{o(it)}^{a}$. Across birth cohorts, the total gap (black line) drops from 0.202 to 0.098. Its share of the overall gap in earnings increases from one-third to one-half. The decline is almost linear from 1931 to 1963 with a slope of -0.003 per year. From 1963 to 1984, the gap remains constant at around 0.108.

[27]The value of $\sum_{cg}\left(\delta_{cg}^{mb}-\delta_{cg}^{fb}\right)P_{cg}^{fb}$ 0.007 in 1931, reaches its minimum of -0.015 in the 1940s, is close to zero between the mid 50s and mid 60s, and ends at -0.008 in 1984.

The cohort residual gap (gray long dashed line) starts at 0.035 in 1931 and 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 is the average of the cross-occupation return for men and women, weighted by the P_{cg}^{fb}, the birth cohort specific cg probabilities for women (orange dashed line). Its magnitude is around 0.060, implying that men work in occupations that pay 5.83% more than those of women with the same education choices. Changes across cohorts in the female cg probabilities do not lead to much variation in the relative return gap. Because the occupation premiums we use are the same for men and women, the 0.060 gap arises because for a given c, g pair, women end up in lower paying occupations than men.

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 with a small increase up to 0.107 in 1936, followed by a steady decline to 0.053 in 1963. Then the education gap widens very slowly to 0.056 in 1978 and falls back to 0.043 by 1984. Figure 6 panel B decomposes the education gap into the contribution of changes in college major, changes in graduate school attendance given college major, and changes in graduate field given graduate school attendance and college major.[28] The magenta short dashed line graphs the contribution of undergraduate field choice. On average across cohorts, undergraduate field accounts for 81% of the level of the education gap in the occupation premium and for most of its decline. Business and Engineering contribute to the decline in the college major gap, as do Education, English/Languages/Literature, and Performing arts. Business contributes 0.016 in 1931, 0.011 in 1949, and 0.001 in 1984. Education contributes 0.031 in 1931 and 0.011 in 1984. Psychology and Social Work and Computer Science/Math contribute a partially offsetting increase in the education gap after 1960.

Gender differences in graduate attendance conditional on major (purple dashed line 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) contributes 0.015 in 1931. 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 although college major choices give men better access to high-paying occupations, the gender differences in going to graduate school and in graduate field conditional on going play a secondary role in long term trends.

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

5.3.1 Within and Across Occupation Sources of the Earnings gap

In Figure 7, panel A we stack the earnings components and the occupation components to show the portion of the gaps that arises within and across occupation. The estimates are based on the constant returns specification. The thickness of the color bands shows the importance within and across occupations of the education gap, the relative return gap, the cohort residual gap, and the demographic gap. The trend in the overall gap in the occupation premium (magenta dashed line) is much flatter than the trend for earnings (black solid line). The education gap is about 59% across occupation and about 41% within occupation. In contrast, 71% of the relative return gap comes from within occupation. Eighty-six percent of the cohort residual gap is within occupation in 1931, and both components drop to zero by 1955 (relative to 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.

Panel B performs a similar exercise for the Education gap, decomposing the three parts of the Education gap "across" and "within" occupation components. The contribution of undergraduate field to the gender gap comes largely through occupation. The same is true, to a great extent for graduate field. In contrast almost all of the effect of graduate attendance is within occupation. 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 shows the result of decomposing the occupation premium $\bar{y}^a_{o(it)}$ based on the cohort-specific relative return estimates. 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 1980 cohorts. These estimates are similar to the estimates based on the constant returns specification.

The cohort residual gap in the occupational premium drops from 0.036 to zero, indicating that shifts in the male-female gap in the occupational premium that are common to all cg choices account for 0.036 of the 0.104 decline of the gap. These values are similar to the the constant returns case.

The path of the education gap (green line) is consistent with the constant return specification reported in Panel A of Figure 6. It contributes 0.087

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

The cohort-specific return gap (orange line) starts at 0.035 in 1931 and rises to 0.059 in 1984. Panel B shows that the gender gap in the occupation premium 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 driven almost entirely by the 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. The fact that it begins with negative values and increases to zero catches our attention. Because $\bar{y}_{o(it)}^{a}$ does not depend on gender or birth cohort, the term $\sum_{cg}\left(\bar{\delta}_{cg}^{mb} - \bar{\delta}_{cg}^{fb}\right)P_{cg}^{fb}$ reflects cross cohort trends in gender differences in the relationship between occupation and cg. The upward slope means that, while the cohort-specific return gap initially favors women, this advantage of women choosing higher-paying occupations than men from the same education fields (relative to the 1961 birth cohort) gets smaller and eventually vanishes. It increases from -0.026 to -0.002.

As a robustness check and to allow us to explore the possibility that changes across cohorts in occupation premiums also contributed to the trend in the gender gap in the occupational component of earnings, we also perform decompositions using cohort specific occupation premiums, $\bar{y}_{o(it)}^{ba}$. 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. 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.

When we use $\bar{y}_{o}^{ba}$ as the dependent variable for our decomposition, we obtain the results in Figure G.2. The behavior of the total gap, education gap, relative return gap, and cohort residual gaps are very similar to the estimates in 8 based on $\bar{y}_{o}^{a}$. This is reassuring.

We now return to the question of the extent to which the upward slope in the relative return 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 across b in the occupation premiums.

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 G.1 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 $\bar{y}_o^{b}$.[29] 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 line), 0.023 is caused by the change in the mapping from cg to occupation that favors men, while only 0.006 is caused by changes in the occupation premiums that favor men (yellow lines in panels B and C, respectively). This further decomposition tells us that, although most components in the gender pay gap decrease over birth cohorts and favor women, the change over cohorts in the mapping from cg to occupation choices favors 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 in 1931, and the overall earnings gap in 1931 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 to this 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 30s, we contribute new facts about the contribution of type of higher education to the large reduction in the gender gap that occurred prior the 1950s cohorts. By incorporating graduate education, we can assess the importance of the 21% point change in the gender gap graduate degree attainment across birth cohorts. We can also assess the contribution of changes in what men and women study in graduate school. The extension back in time is not straightforward because the earliest wave of the National Survey

[29]To create panel B and panel C of this figure, we replace the dependent variable in the regression model (3) with$\bar{y}_o^{0at}$ and λ_o^b, respectively, as defined in section 4.5. The lines are defined in the same way as panel A.

of College Graduates only supplies earnings data only 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.

The introduction provides a detailed summary of the results, which we will not repeat in detail here. 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 education values. Most of the decline is within occupation, especially for the early cohorts. The residual gap varies little from 1951 to the late 70s, after which it resumes its decline. The factors behind residual component are not the subject of our paper, but the extensive literature on long terms trend in the gender gap suggests changes in total labor experience at a given age, 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 contribute very little to the decline in the gender gap over the full time period.

Third, we study and further decompose the "education gap", the contribution of college major choice, graduate degree attainment and graduate field to the gap. When evaluated at fixed relative returns to each degree type, we find that the education gap declines substantially and is an important part of the narrowing of the gender gap. But to our surprise, this decline is mostly offset by cohort trends in the relative returns to specific fields that worked in favor of men against women. Overall, the education gap varies in a narrow range around 0.2 and accounts for very little of the decline.

We close with some 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 (1) changes across cohorts into college and (2) changes in selection into full time work conditional on having a college degree contribute to changes in the earnings gap among college educated workers. In Appendix J we provide evidence based on test scores that differential trends in selection may have worked against the narrowing the gender gap. One could supplement our research by drawing on results 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 undergradate and graduate degrees. Our results assuming constant returns are robust to using the method developed in Altonji and Zhong (2021), that allow us to treat graduate degree choices as endogenous, but we do not have an alternative to OLS for college major. It is possible that selection into the degrees has changed across cohorts in ways that alter the estimates of relative returns. Finally, we leave to future research the task of distinguishing among the many factors, including discrimination, that have contributed to the large changes in education choices of both men and women that we document.

References

Abramitzky, R., V. Lavy, and M. Segev (2022). The effect of changes in the skill premium on college degree attainment and the choice of major. *Journal of Labor Economics just accepted*.

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. _eprint: https://doi.org/10.1146/annurev-economics-080511-110908.

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.

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.

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

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. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/iere.12040.

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.

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.

Kirkebøen, L. J., E. Leuven, and M. Mogstad (2014, December). Field of Study, Earnings, and Self-Selection. SSRN Scholarly Paper ID 2545214, Social Science Research Network, Rochester, NY.

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.

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

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.

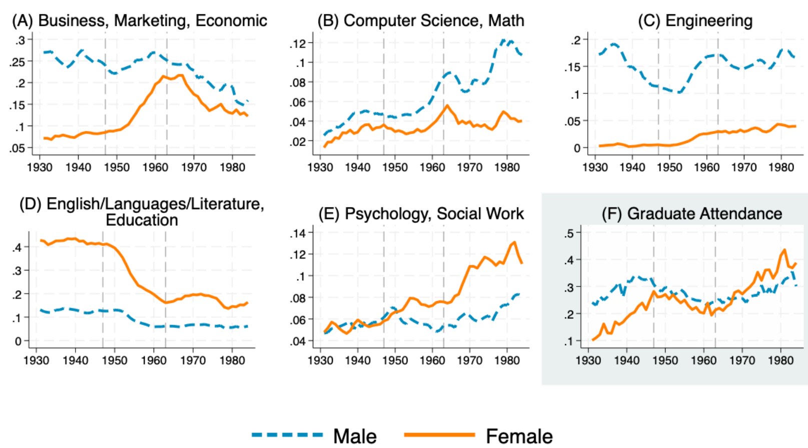
Zafar, B. (2009, February). College Major Choice and the Gender Gap. SSRN Scholarly Paper ID 1348219, Social Science Research Network, Rochester, NY.

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

	(1)	(2)	(3)	(4)
Male (1931-39 birth cohorts)	0.543***	0.349***	0.337***	0.287***
	(0.023)	(0.027)	(0.026)	(0.024)
Male (1940-47 birth cohorts)	0.458***	0.313***	0.291***	0.235***
	(0.026)	(0.025)	(0.024)	(0.023)
Male (1948-63 birth cohorts)	0.373***	0.261***	0.246***	0.194***
	(0.026)	(0.026)	(0.025)	(0.024)
Male (1964-94 birth cohorts)	0.350***	0.259***	0.254***	0.198***
	(0.028)	(0.028)	(0.027)	(0.025)
Constant	10.87***	10.96***	10.82***	10.98***
	(0.027)	(0.028)	(0.027)	(0.026)
Baseline controls	Y	Y	Y	Y
College major		Y	Y	Y
Grad field of study			Y	Y
Occupation				Y
N	409,358	409,358	409,358	378,296

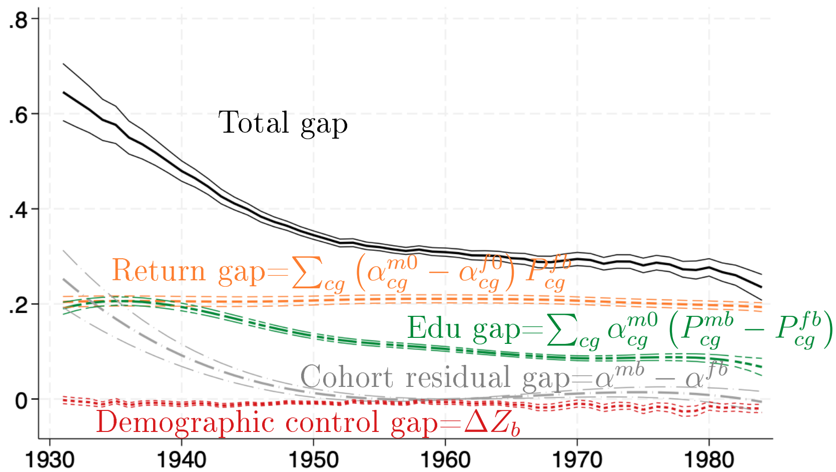
Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The reported table estimates are for our race reference group, White and non-Hispanic individuals, at the average age between 28 and 52. Baseline controls include a cubic in age interacted with a male dummy, a cubic in birthyear, race indicators interacted with gender, parent education dummies, and an adjustment factor constructed from the ACS and decennial census.

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

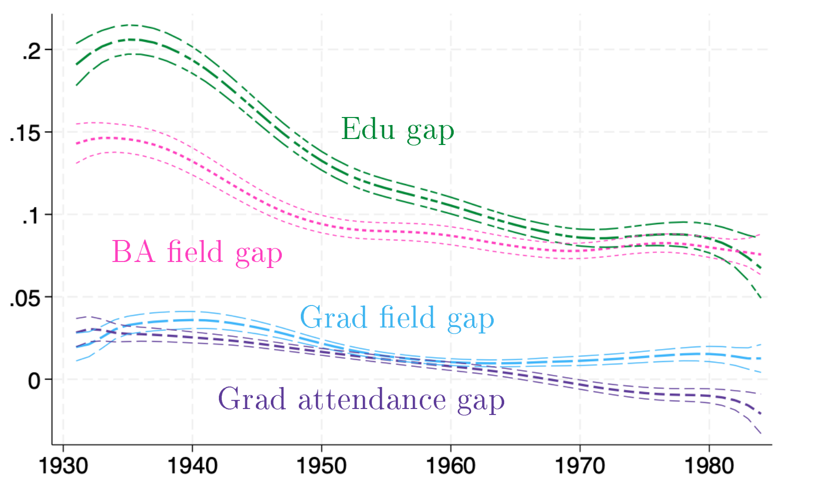


Notes: Panels A-E show the proportion of men and women in specific college majors by birth cohorts between 1931 and 1984. Panel A is Business, Marketing, and Economics majors. Panel B is Computer Science and Mathematics. Panel C is Engineering. Panel D is English/Languages/Literature and Education. Panel E is Psychology and Social Work. Panel F shows the proportion of men and women with graduate degree 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 calculated using the NSCG.

Figure 2: OLS Decomposition of Log Earnings, Constant Returns



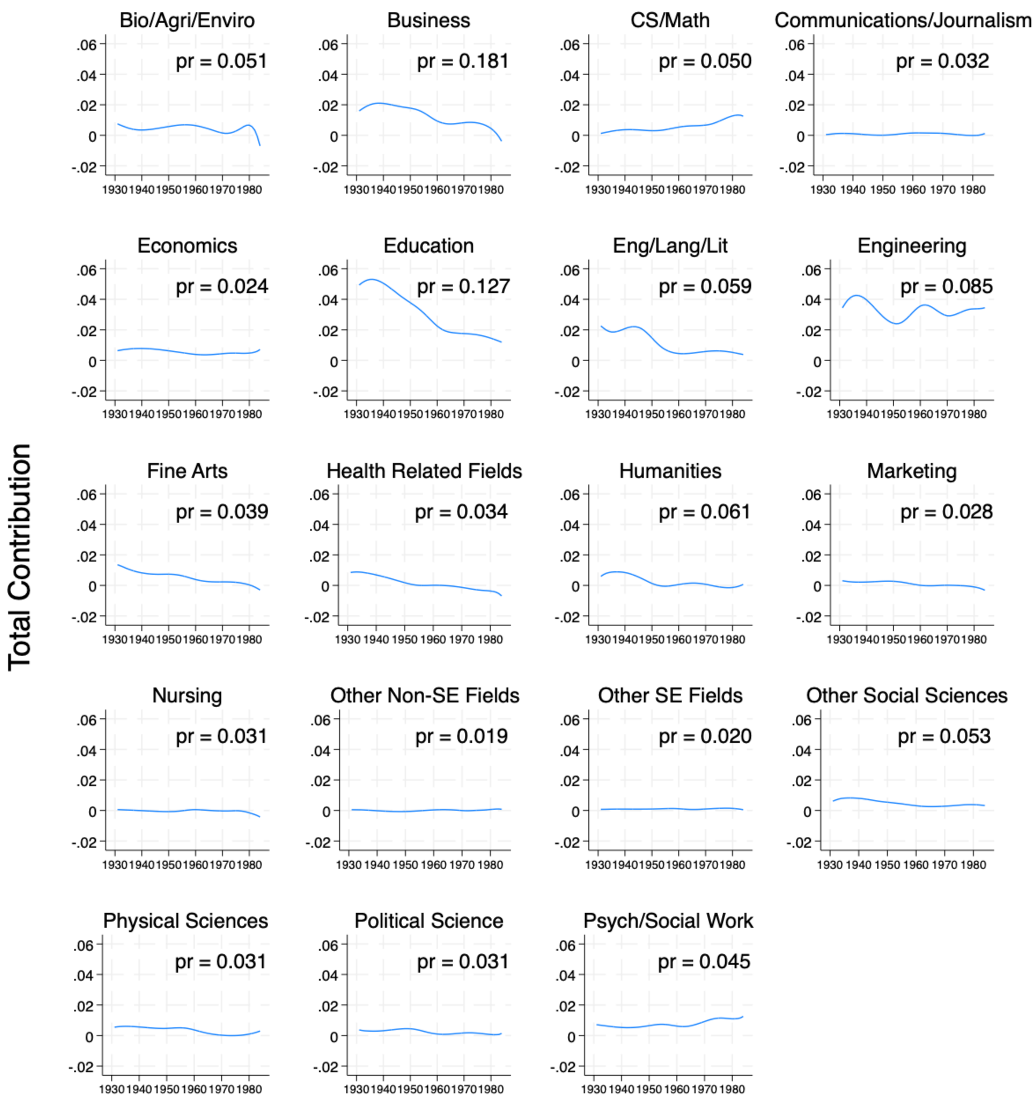
(A) Total gap



(B) Education gap

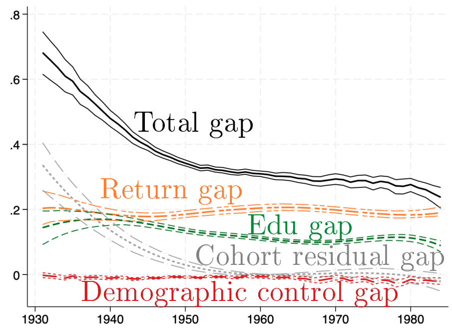
Notes: Panel A shows the decomposition of the predicted gender gap in log earnings for each birth cohort averaged from age 28 to 52. The black line shows the total gender log earnings gap, the orange line shows the portion of the gap at birth year b explained by the gender differences in returns to degrees, the green line shows the education contribution, the gray line shows the cohort contribution that is not related to education fields, and the red line shows the contribution of the demographic controls. The estimates are constructed using OLS estimates of equation (4). The NSCG base year samples are used with cross sectional weights. Ages 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 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

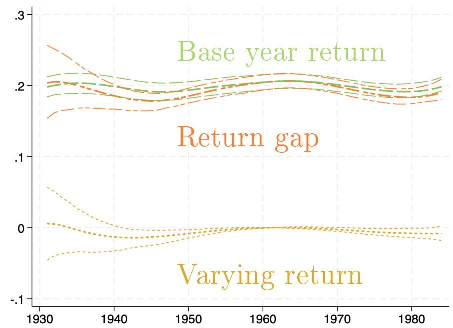


Notes: The figure shows the education gap disaggregated into the contribution of each college major. The subfigures plot $\sum_g \alpha_{cg}^{m0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right)$. The curves in the major specific panels sum to the Education gap graphed as the green line in Figure 2.

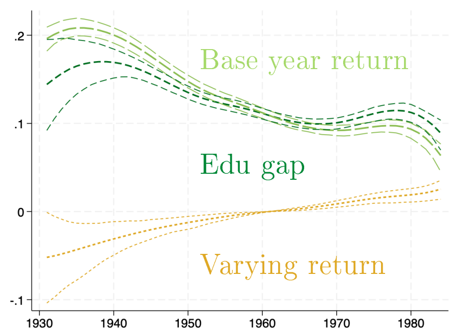
Figure 4: OLS Decomposition of Log Earnings, Cohort Specific Relative Returns



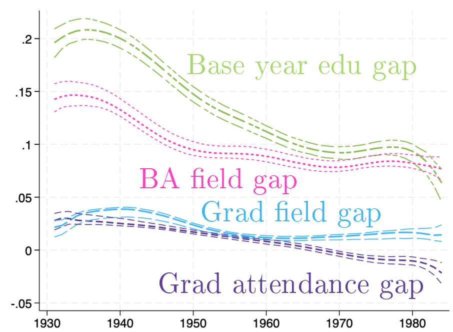
(A) Total gap



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



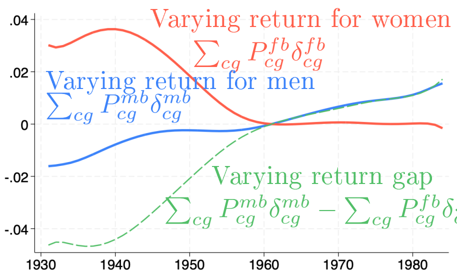
(C) The role in the education gap of base year and cohort varying relative returns



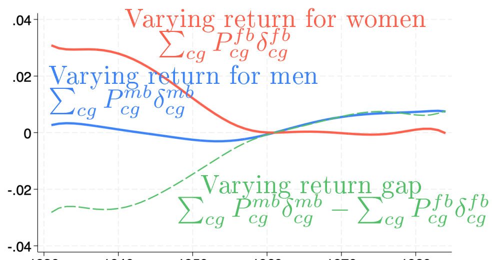
(D) The roles of undergrad field, grad attendance, and grad field in the education gap

Notes: Panel A shows the predicted gender gap in log earnings for each birth cohort at the average age distribution. 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 contribution, the gray line shows the cohort contribution that is not related to education fields, and the red line shows the contribution of the demographic controls. The coefficient estimates are from regression model (8). Panel B decomposes the return gap, $\sum_c \sum_g \left(\alpha_{cg}^{m0} + \delta_{cg}^{mb} - \alpha_{cg}^{f0} - \delta_{cg}^{fb}\right) \times P_{cg}^{mb}$, 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, $\sum_c \sum_g \alpha_{cg}^{mb}\left(P_{cg}^{mb} - P_{cg}^{fb}\right)$, 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.

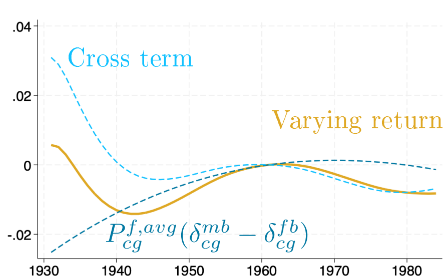
Figure 5: Varying Returns Trend of Log Earnings and Occupation Premium



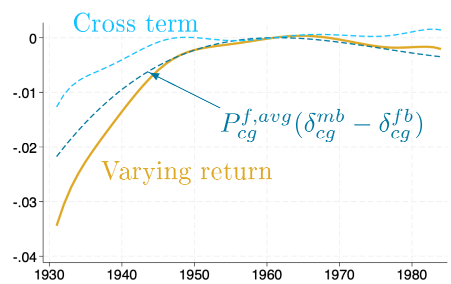
(A) Varying returns for men and women, log earnings



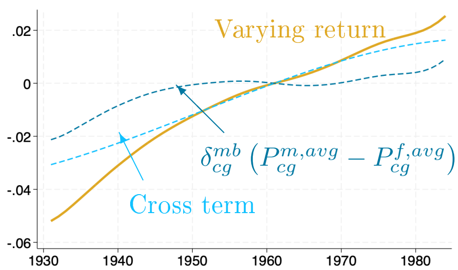
(B) Varying returns for men and women, occupation



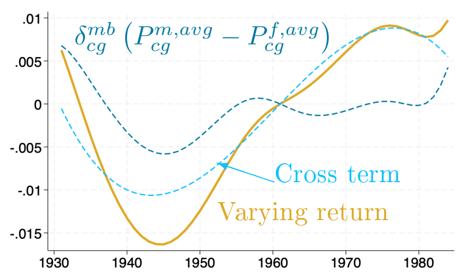
(C) Return gap, log earnings



(D) Return gap, occupation



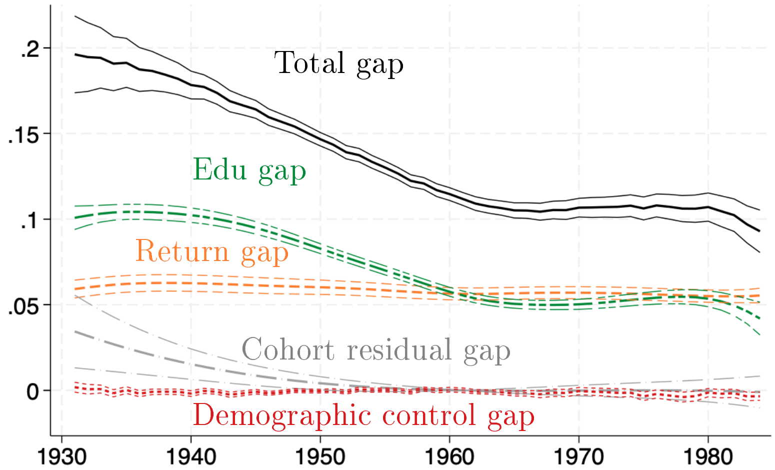
(E) Education gap, log earnings



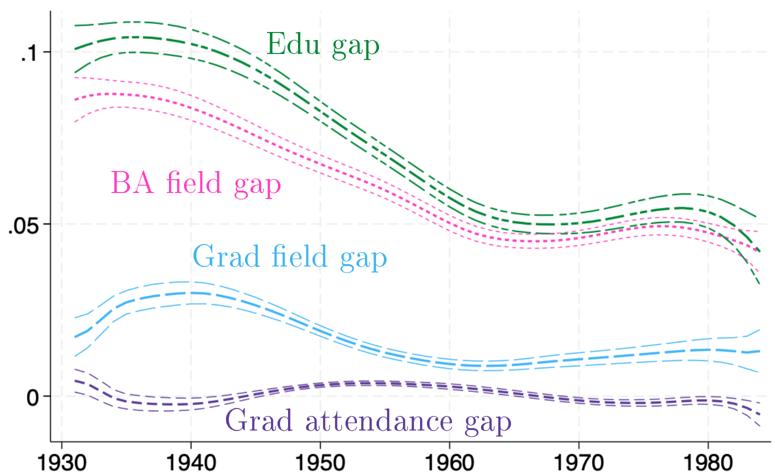
(F) Education gap, occupation

Notes: Panel A shows the overall gap using the gender-specific, cohort varying relative returns in the decomposition of log earnings (green dash line) and the raw sums by gender (blue for men and red for women). Panel B shows the same statistics in the decomposition of occupation premium. Panel C shows the return gap using the gender-specific, cohort varying relative returns in the decomposition of log earnings (yellow line), $\sum_c \sum_g \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb}$, and this line is the same as the yellow line in Figure 4 panel B. This line is decomposed into two components, following equation (9). The dark blue line uses the average probability for women (summand labeled on the figure), and the turquoise line is the cross term. Panel D shows the same decomposition for occupation premium, in which the yellow line is the same as the yellow line in Figure 4 panel C. Panel E shows the education gap using the gender specific, cohort varying returns in the decomposition of log earnings. The yellow line is the same as Figure 8 panel B. Panel F shows the same decomposition for occupation premium.

Figure 6: OLS Decomposition of Occupation Premium, Constant Returns



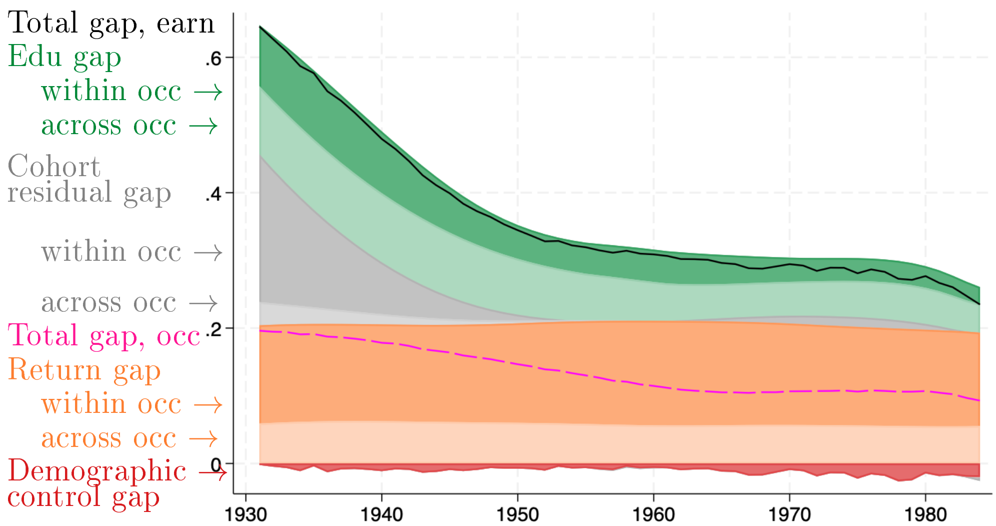
(A) Total gap



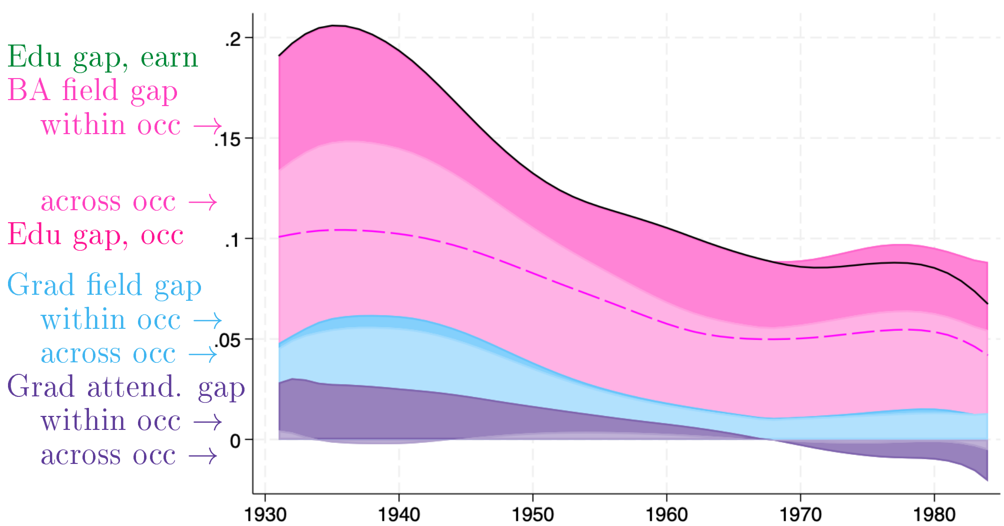
(B) Education gap

Notes: The figure shows predicted gender gap in occupation premium for each birth cohort at the average age distribution. 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, Constant Decomposition



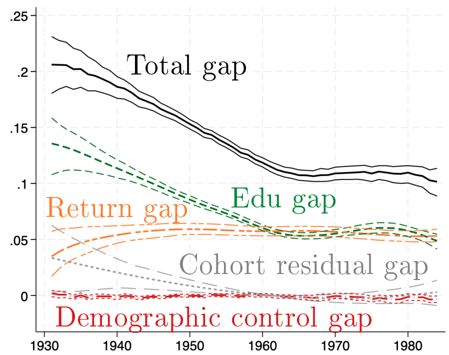
(A) Total gap by occupational effects



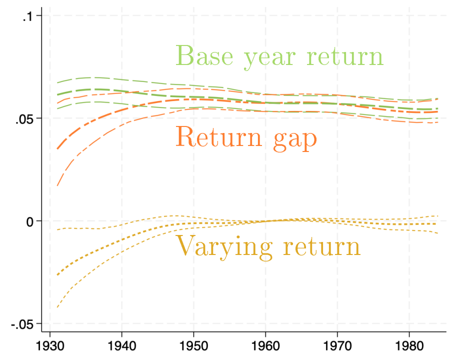
(B) Education gap by occupational effects

Notes: The figure 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 gaps earnings and occupation gaps are smaller than the sum of the positive gaps.

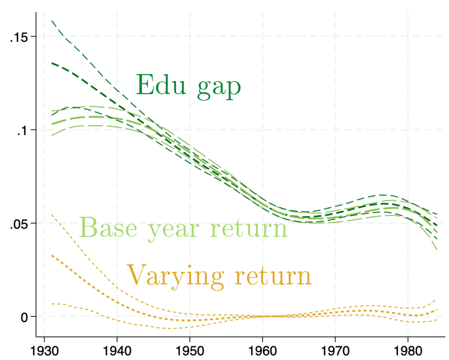
Figure 8: OLS Decomposition of Occupation Premium, Cohort Specific Relative Returns



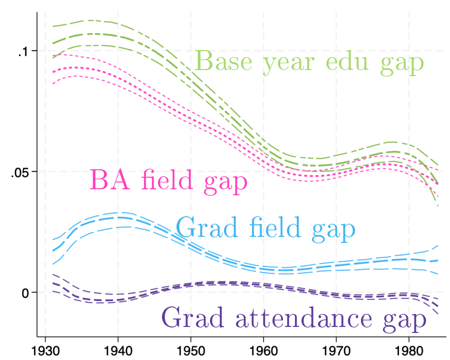
(A) Total gap



(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

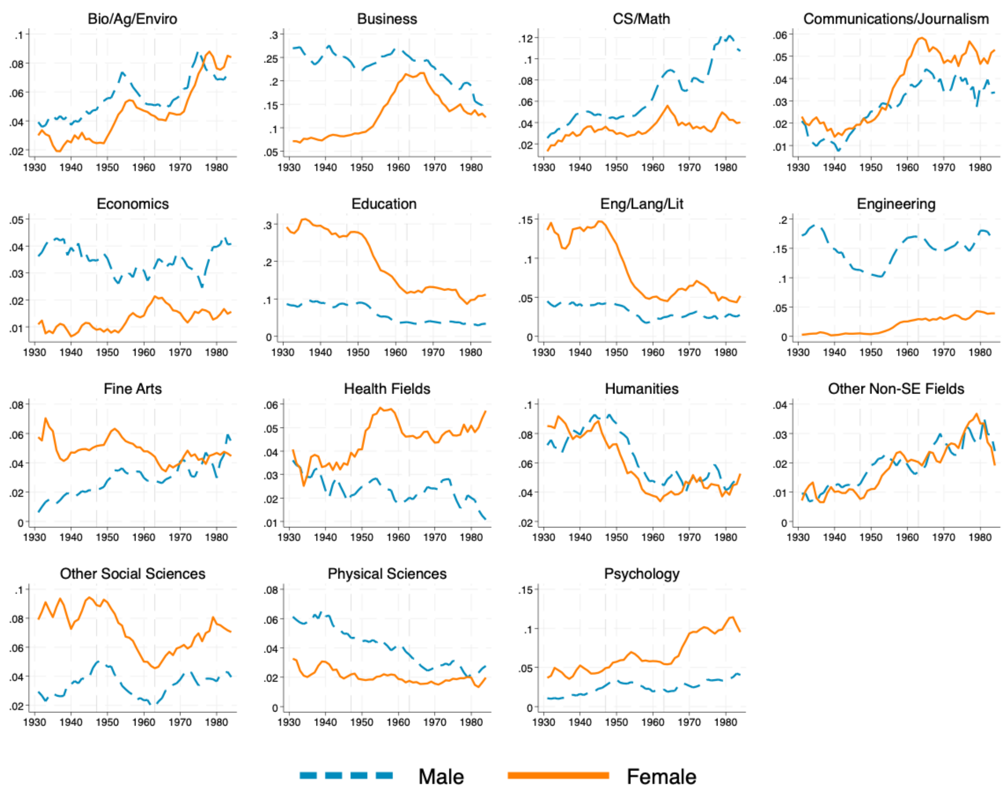
Notes: This figure shows the predicted gender gap in occupation premium for each birth cohort at the average age distribution. The occupation premiums are estimated as described in section 4.5, and are used as the dependent variable in equation (8) to estimate the gender gap. The definitions of the lines are the same as Figure 4.

Online Appendix

A Aggregate Trends of College and Graduate Majors by Gender

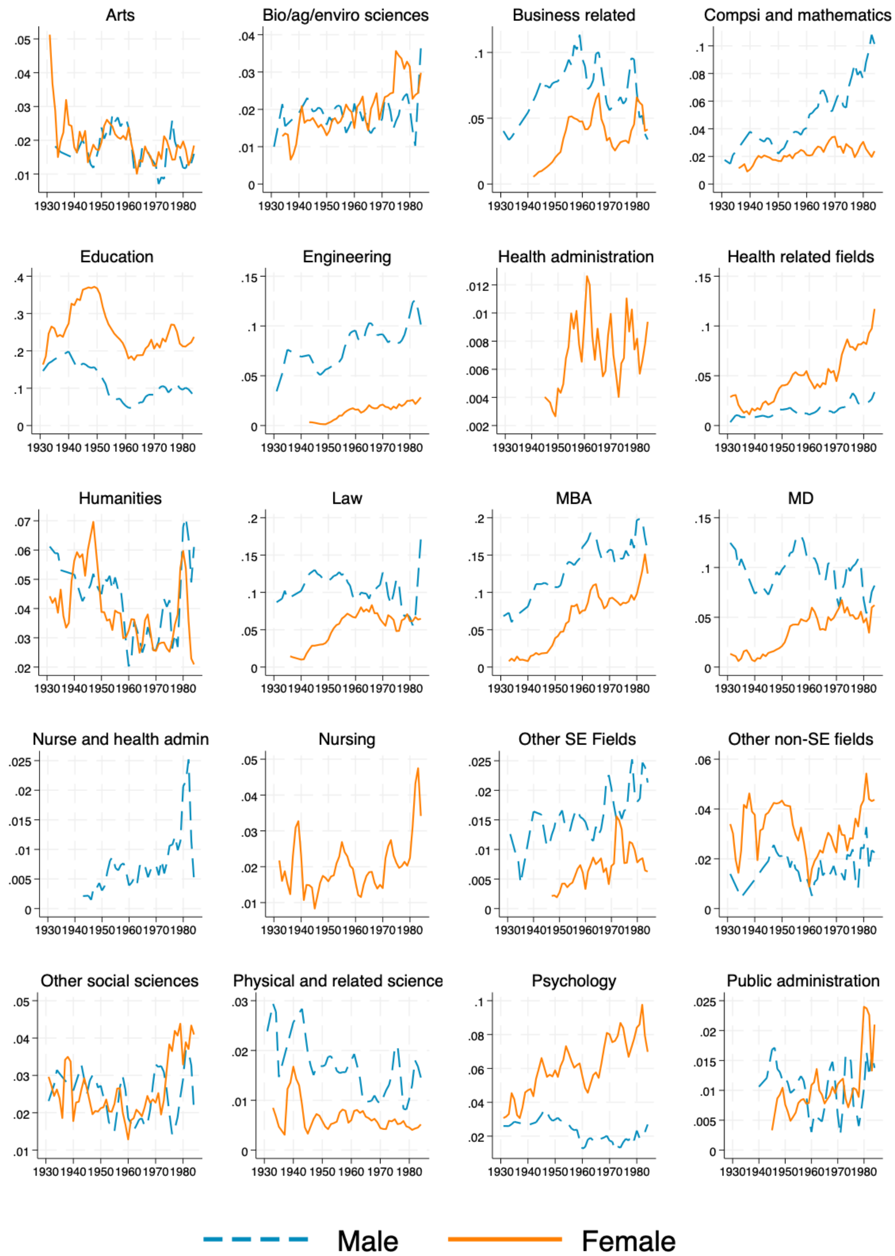
We estimate the probabilities of men and women choosing each college major by birth cohort. Figure A.1 shows the aggregate trends. We also estimate the probabilities fo men and women choosing each graduate major conditioning on having a graduate degree by birth cohort and show the results in Figure A.2.

Figure A.1: Aggregate Trends of College Majors by Gender



Notes: The figure shows the proportion of men and women in specific college majors by birth cohorts between 1931 and 1984. The blue dash line shows the male proportion and the orange solid line shows the female proportion.

Figure A.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 between 1931 and 1984. 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. We merge together Nursing and Health Administration fields for men for similar reasons.

B IPEDs, HEGIS, and Degrees and Other Formal Awards Conferred Biannual Surveys

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 could be many year ago. We 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. Furthermore, it has the disadvanatage that we must make an 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 from those who reside in the US.

The construction of the HEGIS/IPEDS data set relies on multiple sets of 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, they used three different taxonomies. The first two taxonomies, 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 cross walk between the HEGIS data set and the NSF degree classification used in the NSCG surveys. We then used the existing crosswalk to aggregate into our 19 undergraduate and graduate degrees. There is no observation for 1970.

The IPEDS data spans from 1985-2019. All years of the data use the CIP codes to classify degree subjects, though the taxonomy is 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 specification. 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 B.1 and B.2 show the spline created from the marginal distributions of undergraduate degrees in the NSCG (blue), the marginal distributions from HEGIS/IPEDS (red), and their difference (green). 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, allowing us to compare the marginals of birth cohorts going back to 1944. As expected, we do the NSCG shows more graduates with engineering and computer science/math degrees than HEGIS/IPEDS for both men and women. In business, the spline is very similar to the HEGIS/IPEDS for men. For women, the spline over-estimates the number of women getting business degrees in early birth cohorts, then switches to under-estimating in the 1970(CHECK) birth cohort. For education degrees, we again see the spline and HEGIS/IPEDS show very similar results for men. However for women, HEGIS/IPEDS shows women receiving education degrees 10 percentage points more than from the spline between 1944 and 1950. This difference than quickly moves towards 0 for the rest of the birth cohorts.

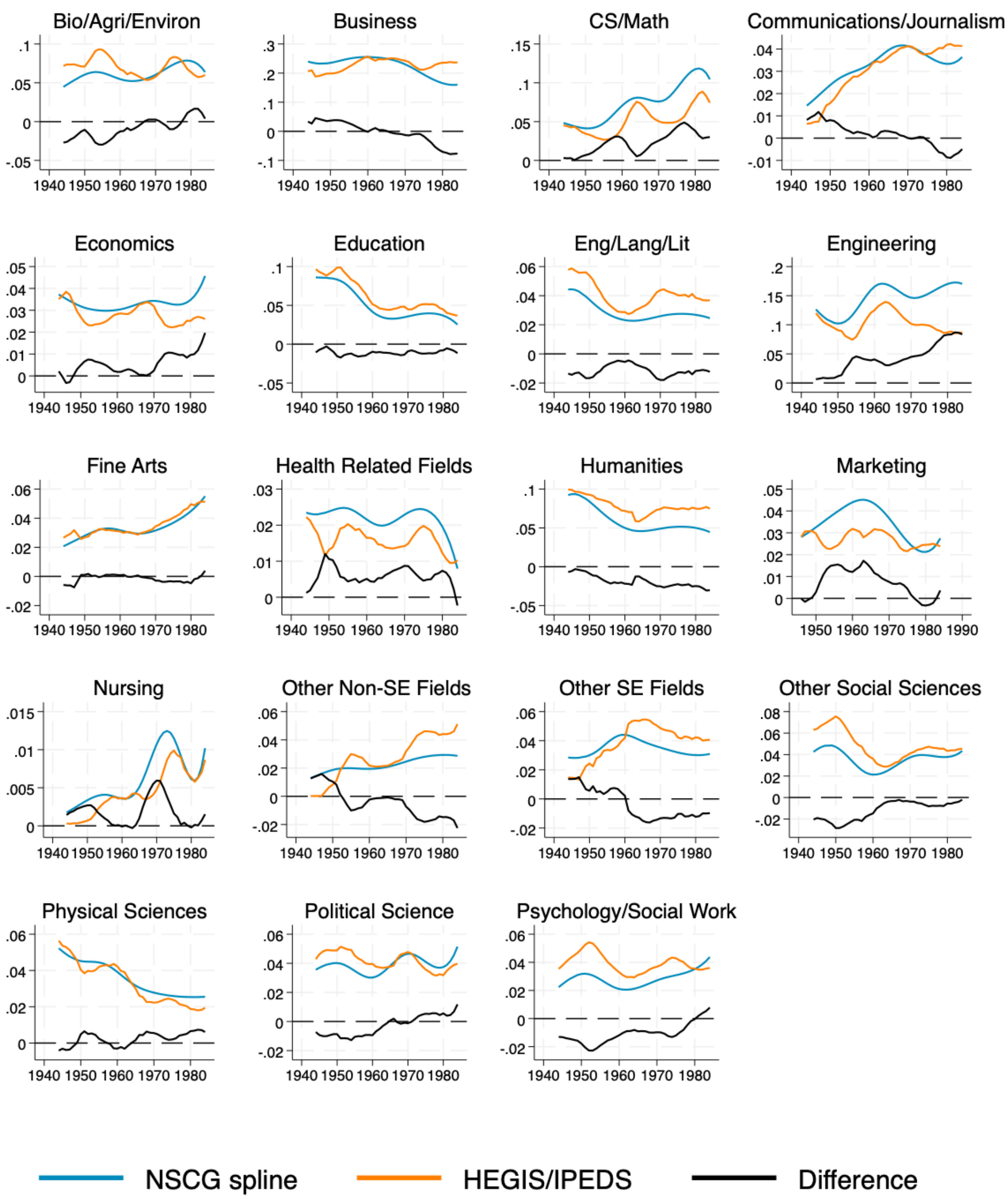
Figures B.3 and B.4shows the same marginal distributions for graduate majors for men and women respectively. Comparisons with the NSCG are harder for graduate degrees because there is a large variation in the age of attainment for graduate degrees, even within subject and gender. We impute the birth 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 there were more men getting graduate degrees in engineering than in our NSCG sample, hovering around 0.1 of all graduate degree earners compared to around 0.07 respectively. The NSCG does a good job at capturing the large growth in women receiving Law, MBA, and MD degrees in earlier birth cohorts (between 1940 and 1960). For less common graduate degrees, the trends in graduate degrees can be seen more clearly. We see large drops in humanities degrees in earlier birth cohorts, especially for women. For non science and engineering (SE) fields, there is a large decrease in the number of women receiving these degrees, while we a see slight increase in the number of men receiving them. For SE-related graduate degrees, we see growth in the number of both men and women attaining graduate degrees, with the gap widening over time in favor of men. Graduate degrees in health administration and nursing have become more popular over time, though still relatively small.

Using the by subject conferral counts from the Degrees and Other Formal Awards Conferred biannual survey report and yearly conferral counts from "120 years of American Education", [30]

[30] Using the by subject conferral counts from the Degrees and Other Formal Awards Conferred biannual survey report and yearly conferral counts from "120 years of American Education", we are able to track the trends in three undergraduate and five graduate degrees going back to the gradaution year1950. This is the first, to our knowledge, time series data of degree majors disaggregated by gender going back this far. However, we do not use this in our analysis due the small coverage of majors. We show these in figures B.5 and

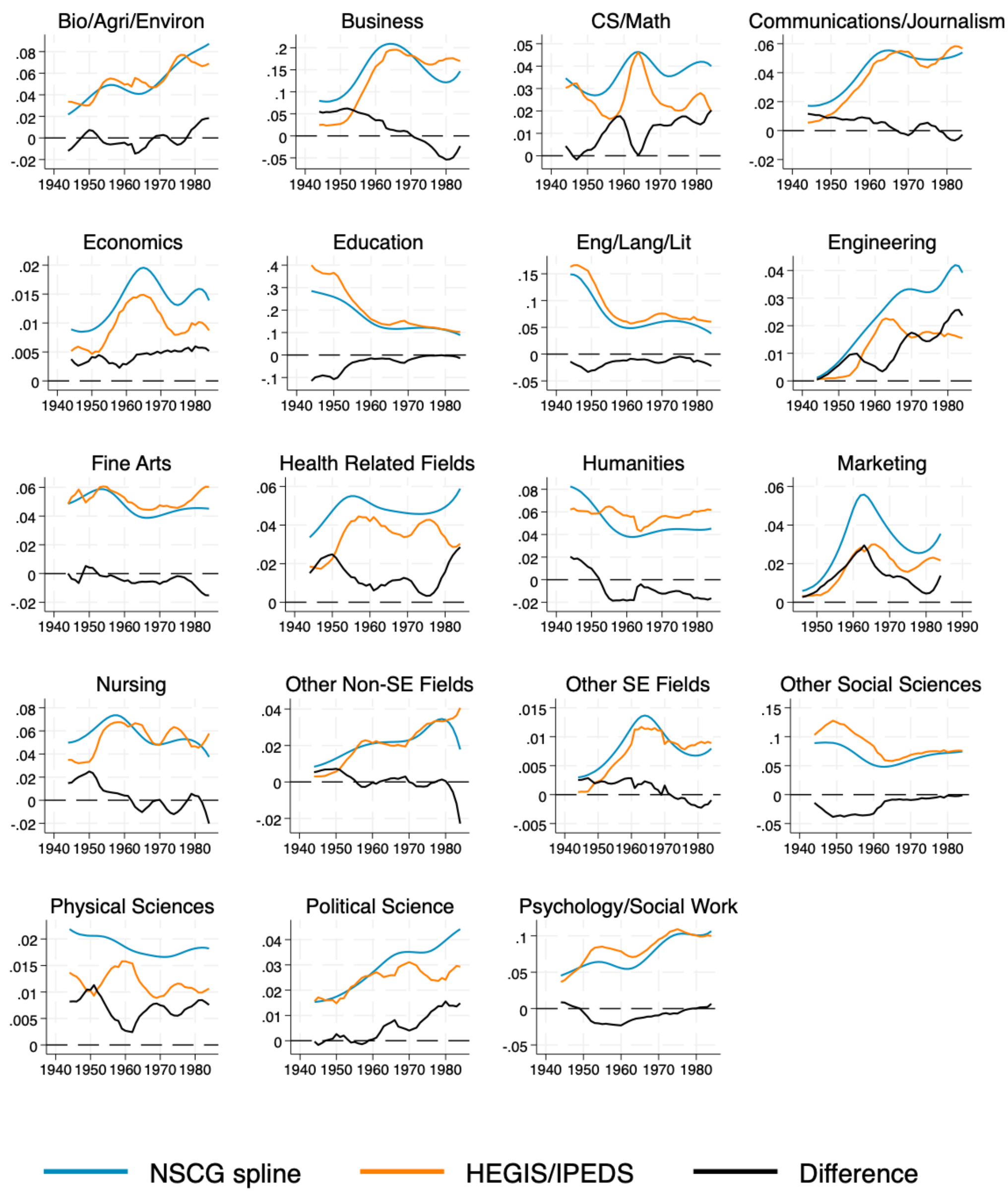
B.6. Note that the horizontal axis in these graphs refers to graduation year, not birth year. The Degrees and Other Formal Awards Conferred biannual survey report can be found on the NCES website. The book "120 Years of American Education", a report written by the NCES, is also available online.

Figure B.1: Aggregate Trends of Undergraduate Fields by Gender in HEGIS/IPEDS: Males



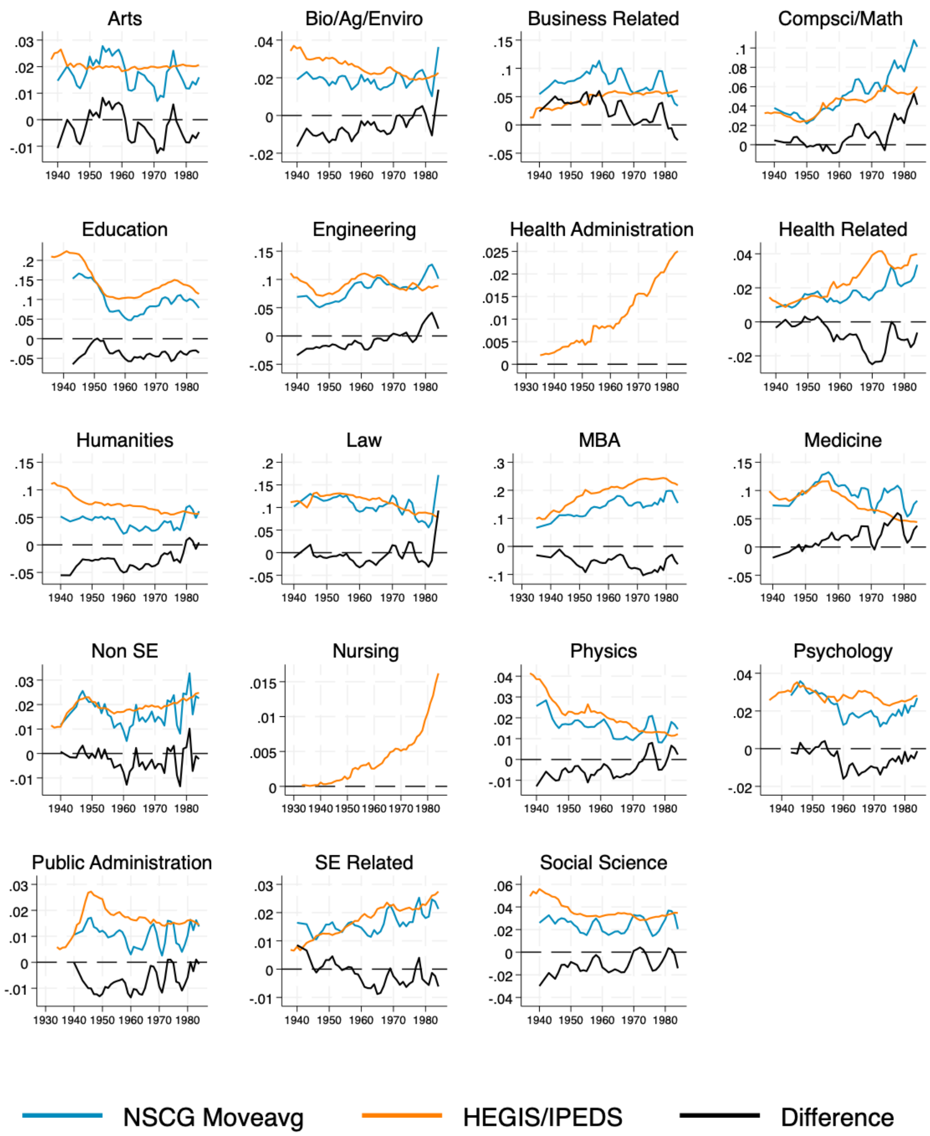
Notes: This figure compares the flows in men receiving a given college major by birth cohort between the spline calculated using the NSCG (blue) and HEGIS/IPEDS (orange) data. The black line shows the difference between the two. We assume the age of obtaining a college degree is 22.

Figure B.2: Aggregate Trends of Undergraduate Fields by Gender in HEGIS/IPEDS: Females



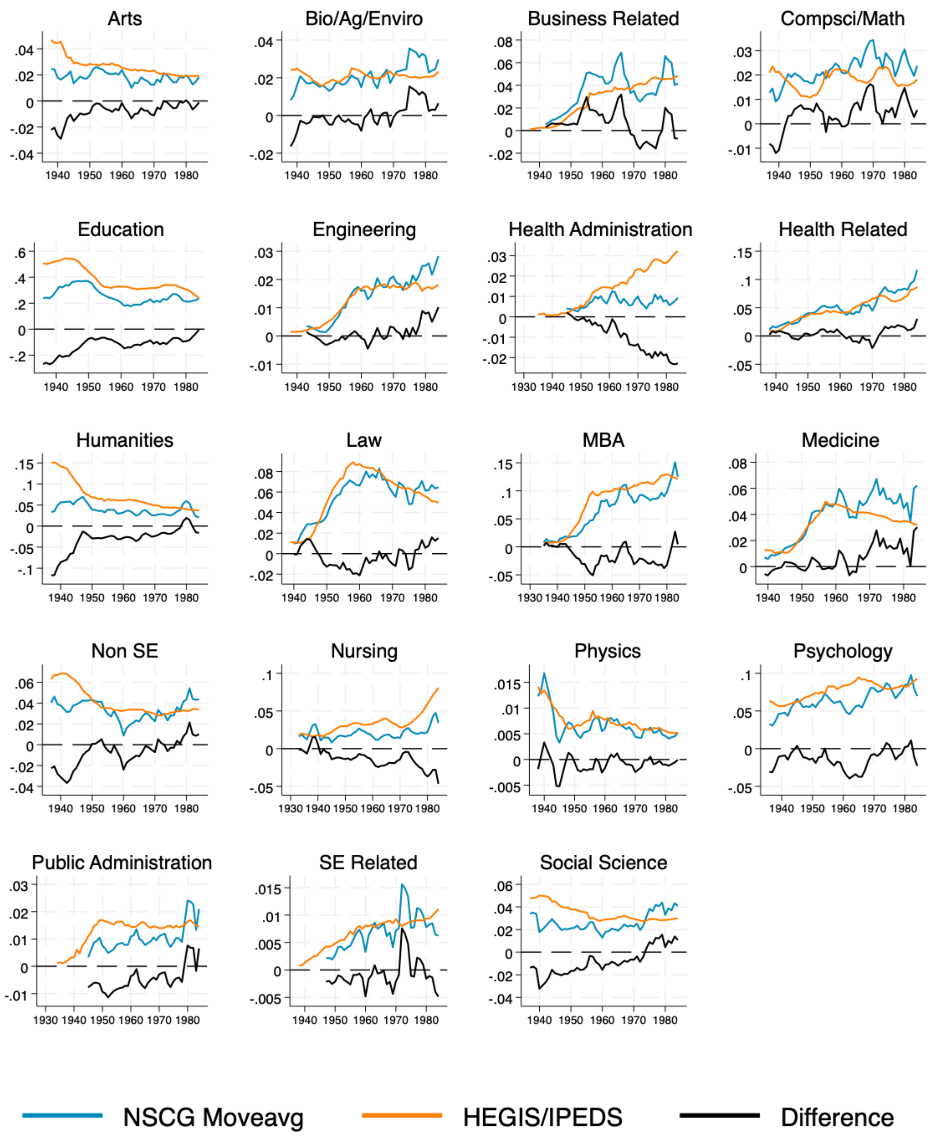
Notes: This figure compares the flows in men receiving a given college major by birth cohort between the spline calculated using the NSCG (blue) and HEGIS/IPEDS (orange) data. The black line shows the difference between the two. We assume the age of obtaining a college degree is 22.

Figure B.3: Aggregate Trends of Graduates Fields by Gender in HEGIS/IPEDS: Males



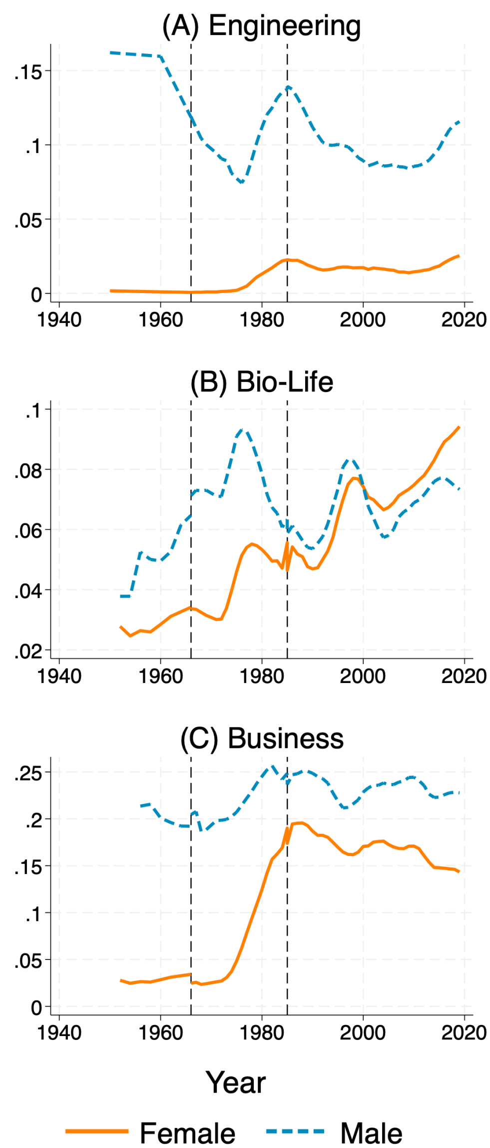
Notes: The figure compares the marginal distributions of men receiving advanced degrees between the HEGIS/IPEDS (orange) data and the splines (blue) created from the NSCG data. The difference is shown in black. To estimate the birth cohort, we calculated the average age of attainment for each graduate degree in the NSCG and subtracted it from the year of conferral in the HEGIS/IPEDS data.

Figure B.4: Aggregate Trends of Graduates Fields by Gender in HEGIS/IPEDS: Females



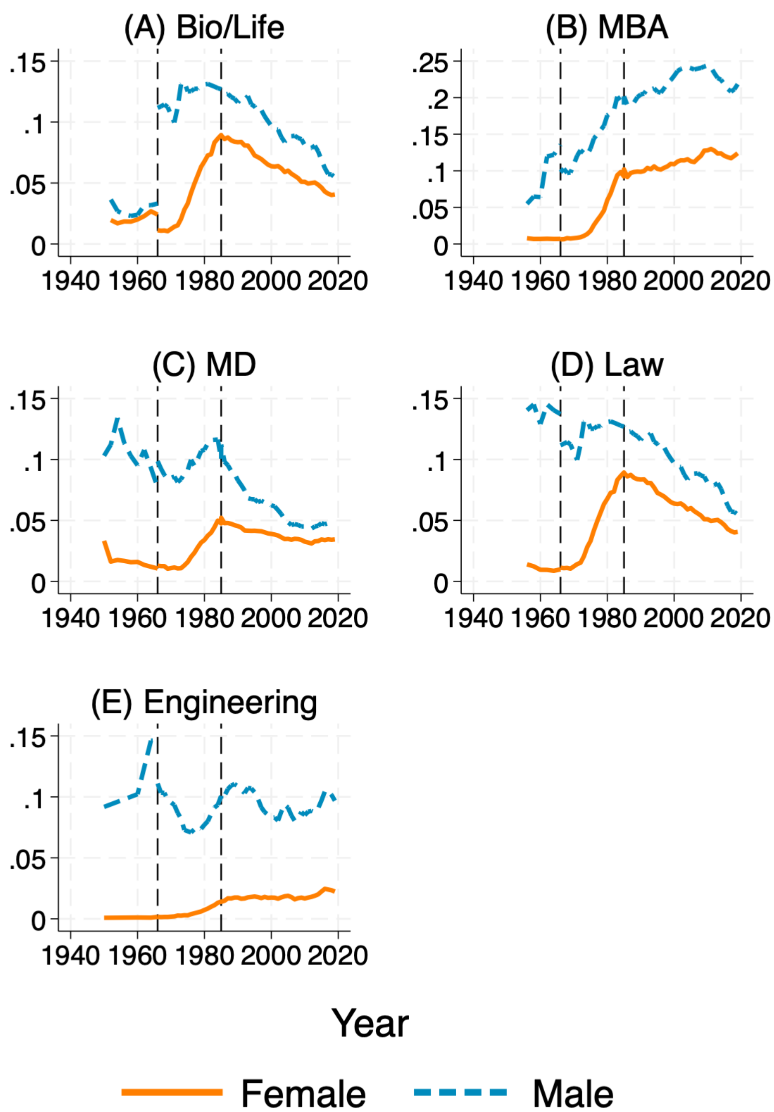
Notes: The figure compares the marginal distributions of women receiving advanced degrees between the HEGIS/IPEDS (orange) data and the splines (blue) created from the NSCG data. The difference is shown in black. To estimate the birth cohort, we calculated the average age of attainment for each graduate degree in the NSCG and subtracted it from the year of conferral in the HEGIS/IPEDS data.

Figure B.5: Trends in Undergraduate Field, Institution Data



Notes: This figure shows the marginal distribution of men and women for undergraduate Engineering, Biology/Life science, and Business majors going back to 1950. The x-axis shows the year of conferral. The first vertical dotted line in 1966 shows the switch from the Degrees and Other Formal Awards Conferred biannual survey report to HEGIS, and the second vertical dotted line shows the switch from HEGIS to IPEDS.

Figure B.6: Estimates of Graduate Degree Attainment for Selected Fields, Institution Data



Notes: This figure shows the marginal distribution of men and women for advanced degrees in Biology/Life science, Business, MD, Law, and Engineering going back to 1950. The x-axis shows the year of conferral. The first vertical dotted line in 1966 shows the switch from the Degrees and Other Formal Awards Conferred biannual survey report to HEGIS, and the second vertical dotted line shows the switch from HEGIS to IPEDS.

C Additional Details and Formulas for the Decomposition

We provide additional details on the decomposition formulas.

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

$$
\begin{aligned}
&\text{Education Gap } (b) = \\
&\sum_{cg} \alpha_{cg}^{m0} \left(\Pr{}_b^m(g,G|c) - \Pr{}_b^f(g,G|c)\right) \Pr{}_b^f(G|c) \Pr{}_b^f(c) && \text{grad field gap } (\alpha_{cg}^{m0}) \\
&+ \sum_{cg} \alpha_{cg}^{m0} \Pr{}_b^f(g|G,c) \left(\Pr{}_b^m(G|c) - \Pr{}_b^f(G|c)\right) \Pr{}_b^f(c) && \text{grad enroll gap } (\alpha_{cg}^{m0}) \\
&+ \sum_{cg} \alpha_{cg}^{m0} \Pr{}_b^f(g,G|c) \left(\Pr{}_b^m(c) - \Pr{}_b^f(c)\right) && \text{BA field gap } (\alpha_{cg}^{m0}) \\
&+ \sum_{cg} \delta_{cg}^{mb} \left(\Pr{}_b^m(g,G|c) - \Pr{}_b^f(g,G|c)\right) \Pr{}_b^f(G|c) \Pr{}_b^f(c) && \text{grad field gap } (\delta_{cg}^{mb}) \\
&+ \sum_{cg} \delta_{cg}^{mb} \Pr{}_b^f(g|G,c) \left(\Pr{}_b^m(G|c) - \Pr{}_b^f(G|c)\right) \Pr{}_b^f(c) && \text{grad enroll gap } (\delta_{cg}^{mb}) \\
&+ \sum_{cg} \delta_{cg}^{mb} \Pr{}_b^f(g,G|c) \left(\Pr{}_b^m(c) - \Pr{}_b^f(c)\right) && \text{BA field gap } (\delta_{cg}^{mb}) \\
&+ \Delta ED_b^{23} && \text{approx. error}
\end{aligned}
\tag{10}
$$

Table I.6 reports the estimates of this decomposition.

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

In the Blinder-Oaxaca decomposition, it is important that we alter the base of the decomposition and redo the calculation. If the results are different in the alternative specification, we will discuss the rationale. In this appendix, we briefly present the equations of the decomposition if we use female coefficient as the base.

D.1 Birth cohort specific decompositions, constant returns

The Blinder-Oaxaca style decomposition using the female education coefficients as the weights and the male cg probabilities is as follows:

$$
\begin{aligned}
GAP(b) = & \sum_{cg} \left(\alpha_{cg}^{m0} - \alpha_{cg}^{f0}\right) P_{cg}^{mb} && \text{relative return gap} \\
& + \sum_{cg} \alpha_{cg}^{f0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{Education gap} \\
& + \alpha^{mb} - \alpha^{fb} && \text{cohort } b \text{ residual gap} \\
& + \Delta Z_b && \text{demographic control gap}
\end{aligned} \tag{11}
$$

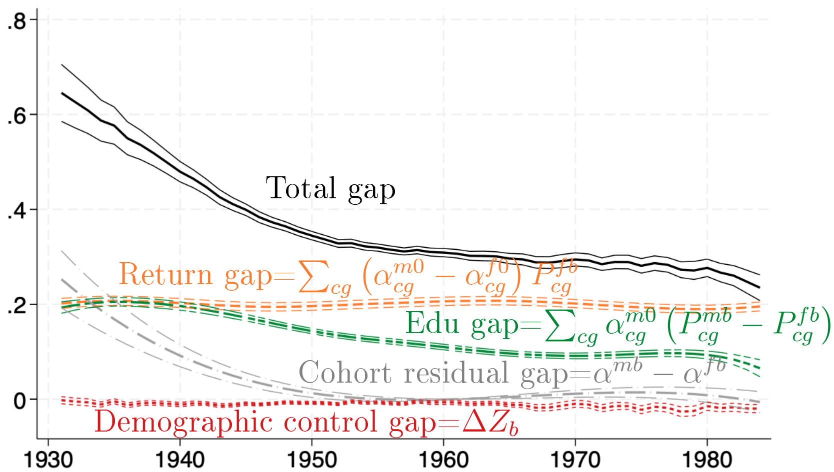
The equation is comparable with equation (4) in the main text.

The decomposition of the education gap using the female coefficient is:

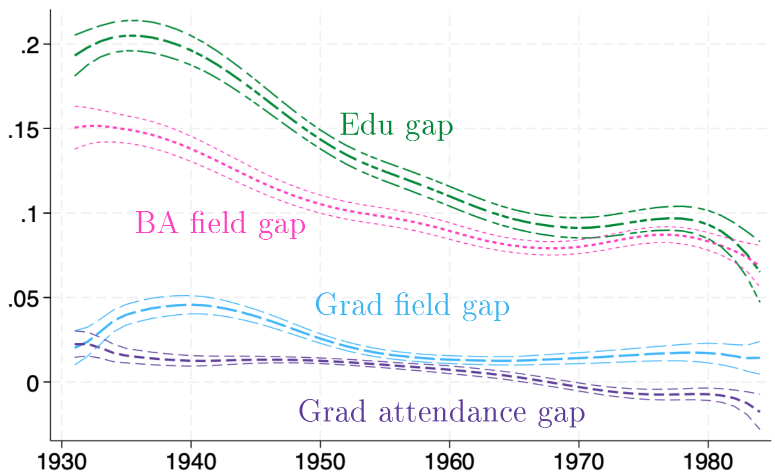
$$
\begin{aligned}
& \text{Education Gap } (b) = \\
& \sum_{cg} \alpha_{cg}^{f0} \times \left(\Pr{}_b^m(g,G|c) - \Pr{}_b^f(g,G|c)\right) \times \Pr{}_b^m(G|c) \times \Pr{}_b^m(c) && \text{grad field gap} \\
& + \sum_{cg} \alpha_{cg}^{f0} \times \Pr{}_b^m(g|G,c) \times \left(\Pr{}_b^m(G|c) - \Pr{}_b^f(G|c)\right) \times \Pr{}_b^m(c) && \text{grad enroll gap} \\
& + \sum_{cg} \alpha_{cg}^{f0} \times \Pr{}_b^m(g,G|c) \times \left(\Pr{}_b^m(c) - \Pr{}_b^f(c)\right) && \text{BA field gap} \\
& + \Delta ED_b^{23} && \text{approx. error}
\end{aligned} \tag{12}
$$

Figure D.1 shows the decompositions of the total gap and the education gap of log earnings using the female coefficients and male probability following equations (11) and (12). The estimates are very consistent with the male coefficient version shown in the main text. Figure D.2 shows the decompositions of occupation premium using female coefficients. The estimates of the return gap is, on average, 0.01 higher than the male coefficient version.

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



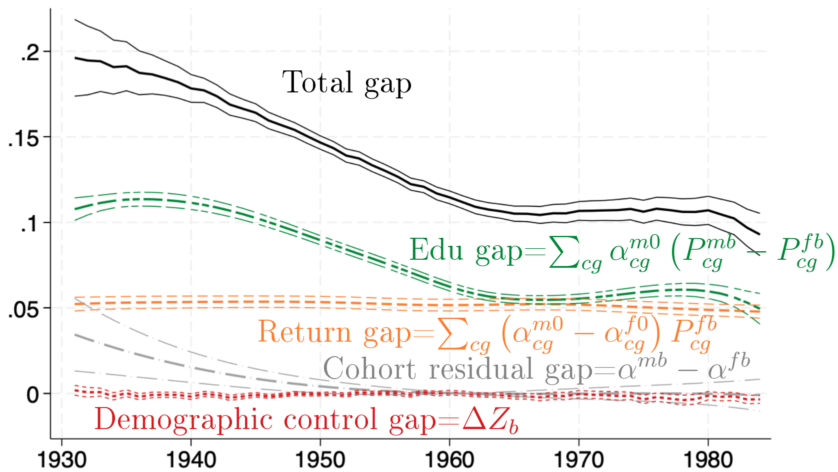
(A) Total gap



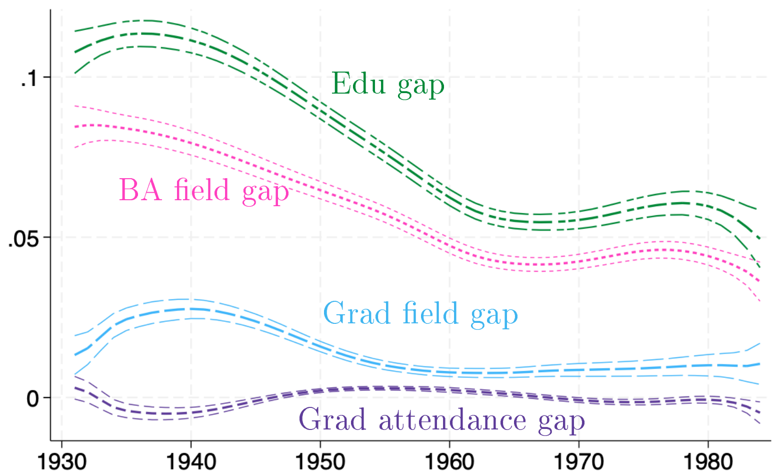
(B) Education gap

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

Figure D.2: OLS 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.

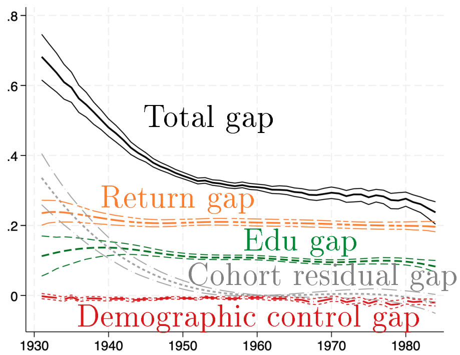
D.2 Cohort varying relative returns

The decomposition formula with cohort varying relative returns using female coefficient is:

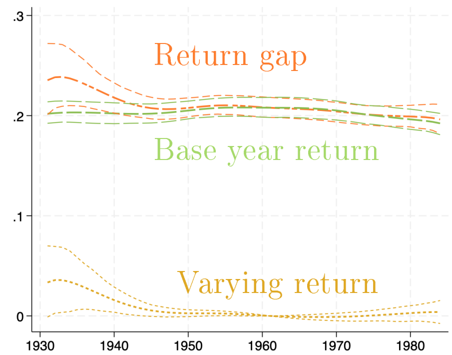
$$
\begin{aligned}
GAP(b) = & \sum_{cg} \left(\alpha_{cg}^{m0} - \alpha_{cg}^{f0}\right) P_{cg}^{mb} && \text{rel. return gap, base year returns} \\
& + \sum_{cg} \alpha_{cg}^{f0} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{education gap,base year returns} \\
& + \alpha^{mb} - \alpha^{fb} && \text{cohort } b \text{ residual gap} \\
& + \Delta Z_b && \text{demographic control gap} \\
& + \sum_{cg} \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{mb} && \text{rel. return gap, varying returns} \\
& + \sum_{cg} \delta_{cg}^{fb} \left(P_{cg}^{mb} - P_{cg}^{fb}\right) && \text{education gap, varying returns}
\end{aligned} \tag{13}
$$

Figure D.3 and Figure D.4 are the cohort varying relative return decomposition of log earnings and occupation premium following equation (13). When using the female coefficient and male probability, the education gap 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 gaps. We discussed their rationale in section 5.4.

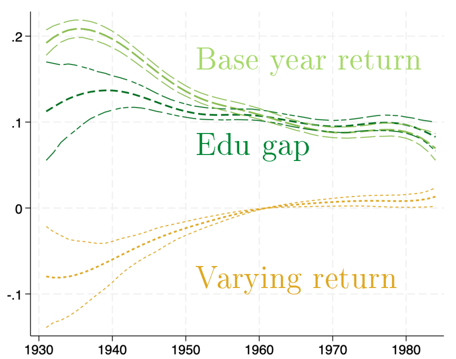
Figure D.3: OLS Decomposition of Log Earnings, Cohort Specific Relative Returns, Female Coefficients



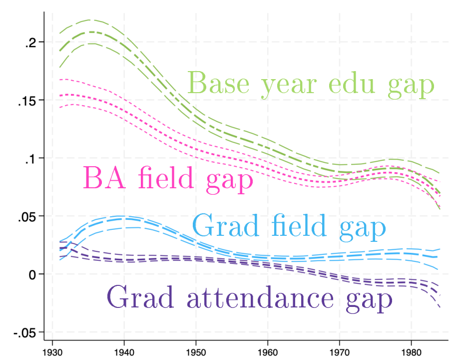
(A) Total gap



(B) The Role in the Return gap of base year and cohort 1.6cmvarying relative returns



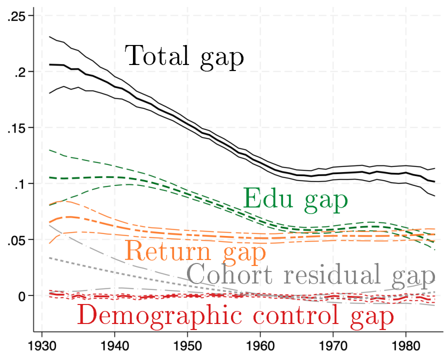
(C) The Role in the Education gap of base year and cohort varying relative returns



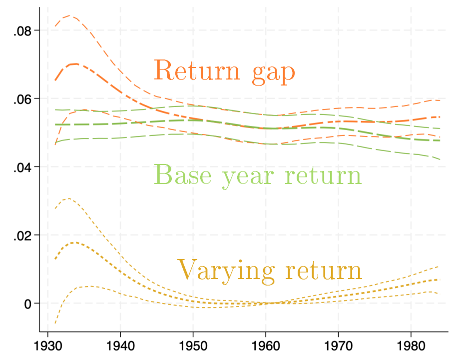
(D) The roles of undergrad field, grad attendance, and grad field in the education gap

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

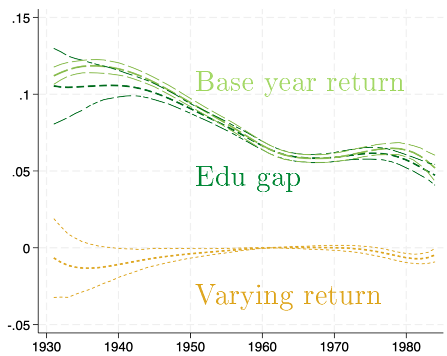
Figure D.4: OLS Decomposition of Occupation Premium, Cohort Specific Relative Returns, Female Coefficients



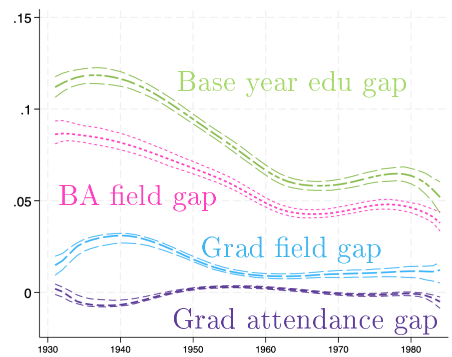
(A) Total gap



(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 log earnings 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 4.

E 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

$$Y_{it} = \alpha^s_{G(i)} + X^s_{1it}\beta^s_1 + X^s_{2it}\beta^s_2 + Z^s_i\Gamma^s + u_{it}. \tag{14}$$

In this regression, α^s is the interaction between gender and a binary education variable indicating whether the $G(i)$ is 0 for individual with only a college degree and 1 for individuals with a graduate degree. It is based on years of education.[31] The vector X_{1it} contains the triple interaction among gender, the graduate education dummy and a cubic age polynomial, as well as a gender specific cubic birth cohort polynomial. The demographic control Z^s_i contains interactions between gender and race and Hispanic dummies. The excluded category for men and for women is a white non-Hispanic. Recall that equation (1) contains an adjustment factor X^s_{2it} estimated from the Census/ACS data to account for the potentially poor estimation quality of the cohort differences in earnings gaps in some age ranges. We do not need this factor when we estimate regression model (14) because the data coverage in early birth cohorts is much better in here.

The compositions, $pr\,(c \mid s, b)$, is estimated on the sample of people above 35 years old, and smoothed using a spline basis, then winsorized and normed 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

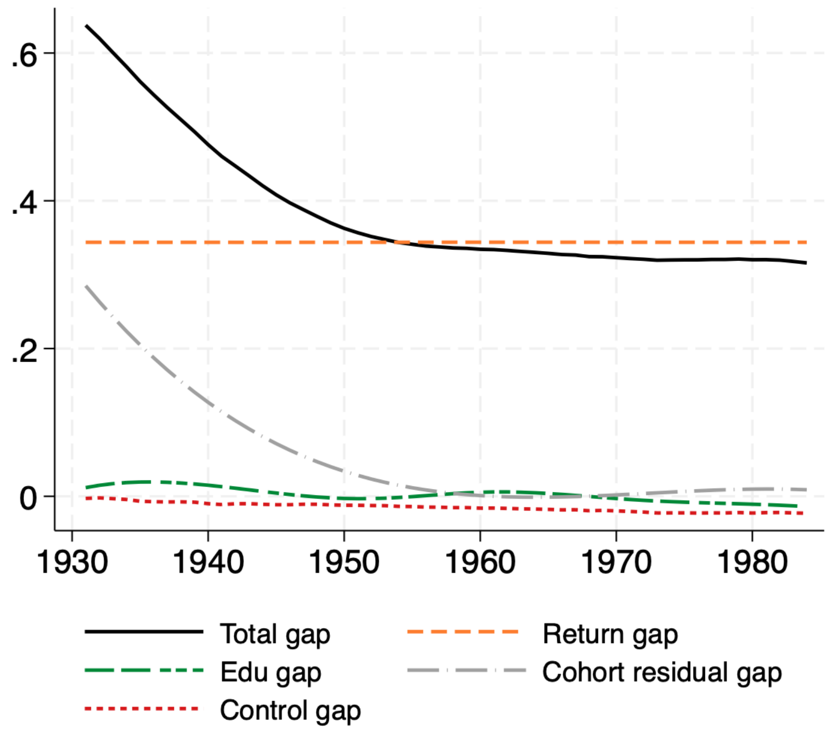
[31] The 1960-1990 Census uses years of schooling attended and years of schooling completed as the education measures rather than degree attainment. This poses challenges in defining the population of individuals with a college degree and with a graduate degree. Based on Park (1996); Frazis et al. (1995); 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, we set $G(i)$=1 for individuals with six or more years of completed schooling and G(i)=0 for individuals with four or five years of completed schooling. For the 2000 Census and the ACS we use responses to direct questions about degree attainment.

$$GAP(b) = \sum_{G=0,1} \left(P_G^{mb} \alpha_G^m - P_G^{fb} \alpha_G^f \right) + (\alpha^{mb} - \alpha^{fb}) + \Delta Z_b$$

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

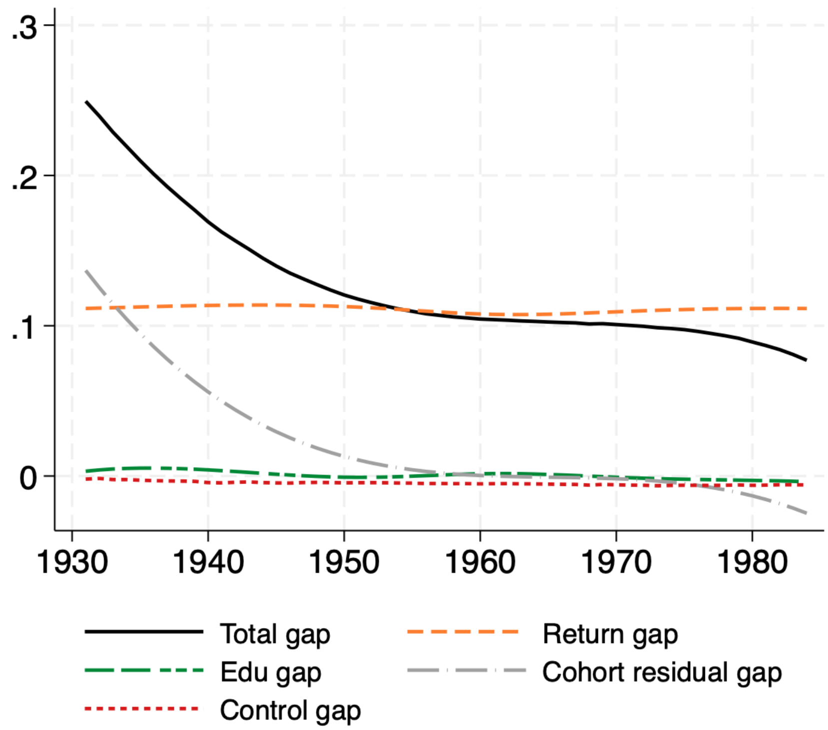
Figure E.1 presents the decomposition results following equations above. The total gaps of log earnings and occupation premiums are consistent with our main specification. The Census/ACS data has limited information on higher education history. As a result, compared with the results for the constant returns specification based on the NSF data (Figure 2 and 6), the results here show larger return gaps, much smaller education gaps, and slightly larger cohort residual gaps.

Figure E.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 b explained by the gender differences in returns to education level (Graduate versus college only), the green line shows the education contribution, the gray line shows the cohort contribution that is not related to education level, and the red line shows the contribution of the demographic controls. The coefficient estimates are from regression model (14). OLS coefficients were used. The data comes from Census/ACS. Ages 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. The residual gap is normalized to be zero in 1961.

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



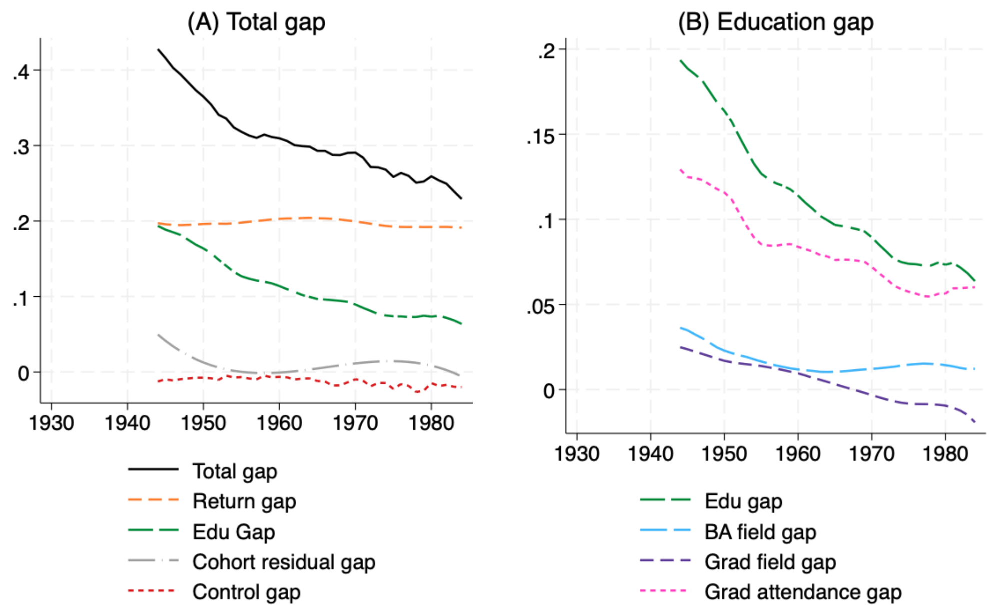
Notes: The figure shows predicted gender gap in occupation premium for each birth cohort at the average age distribution. The data is from Census/ACS.

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

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

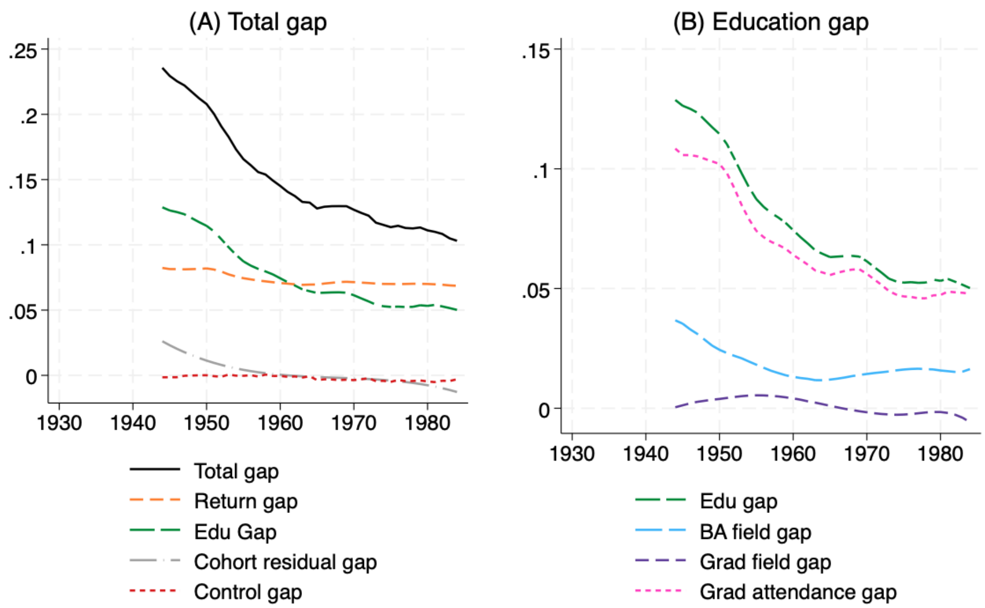
One concern is whether the the NSCG data is sufficiently representativeness to construct the undergraduate degree probabilities P_c^{fb} and P_c^{mb}. As check, we perform decompositions using probability estimates based on the HEGIS/IPEDS data. We assume that the age of conferral is 22 to impute birth 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. This amounts to an average of 0.015 difference between the two decompositions between 1971 and 1984.

Figure F.1: OLS 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. Panel A and B can be compared to panel A and B from 2. Since HEGIS only goes back to 1966, we can only observe birth cohorts after 1944.

Figure F.2: OLS Decomposition for Occupation Premium using $pr(c)$ from HEGIS, Constant Returns

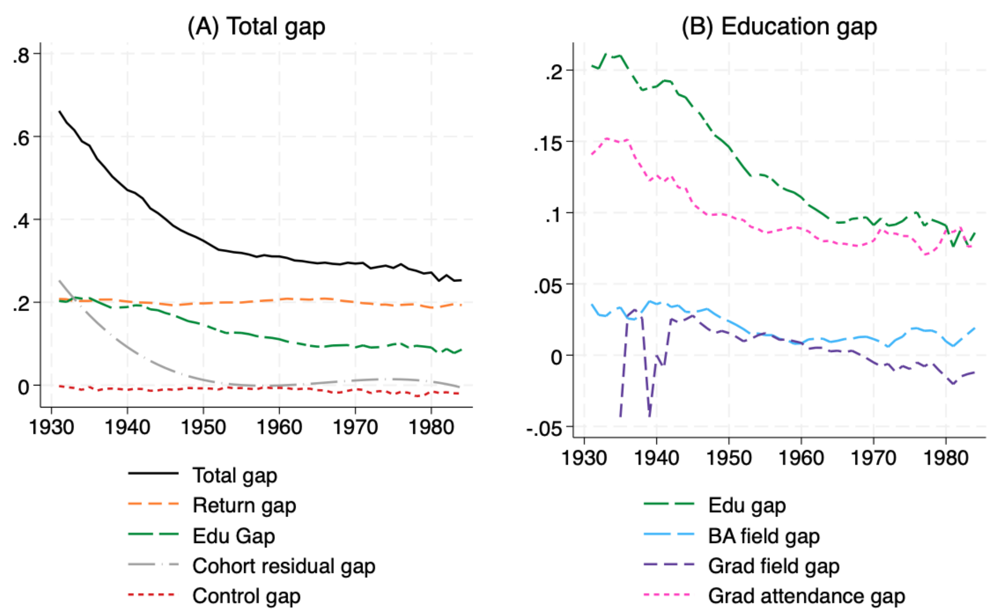


Notes: This figure shows the decomposition results for the education gap when we replace the NSCG based estimates of P_c^{fb} and P_c^{mb} with the estimates from the HEGIS/IPEDS data. We assume that the age of attainment of the undergraduate degree is 22. Panel A and B can be compared to panel A and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

F.2 Using Moving Averages to Estimate Undergraduate Degree Probabilities

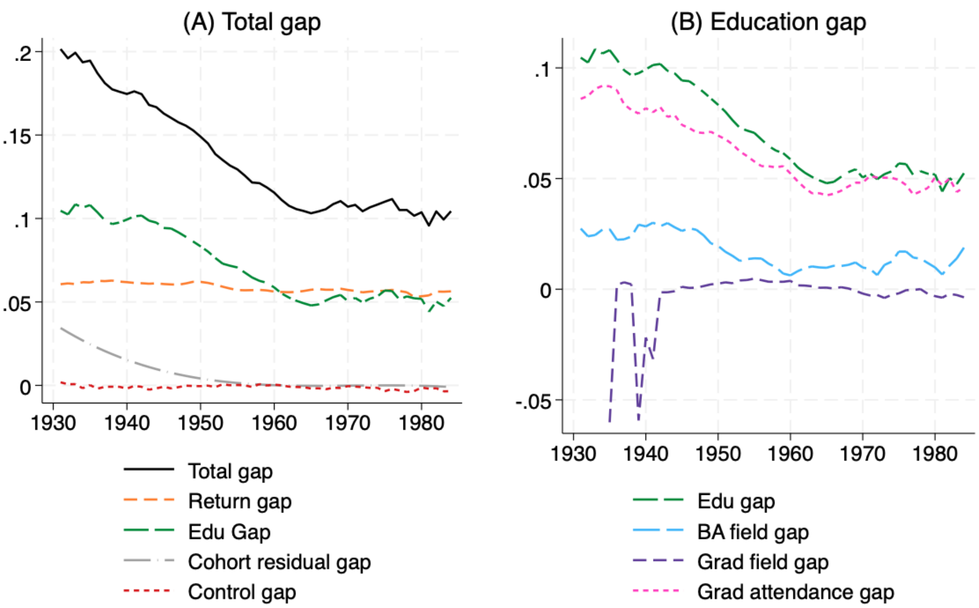
In the earnings and occupation premium decomposition we use b-splines to estimate P_c^{fb} and P_c^{mb}. It is possible that the use of splines leads to overfitting and biases the decompositions. To check on this, we replicate the decomposition using 3-year moving averages of the birth year specific and gender specific distributions of undergraduate degrees, by gender. Panel A shows the decomposition into the total gap, return gap, education gap, cohort residual gap, and control gap. Figure F.3 and F.4 show the decomposition with constant returns with log earnings and occupation premium as the dependent variable respectively. The decomposition results are similar to our main results.

Figure F.3: OLS 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. Panel A and B can be compared to panel A and B from Figure 2.

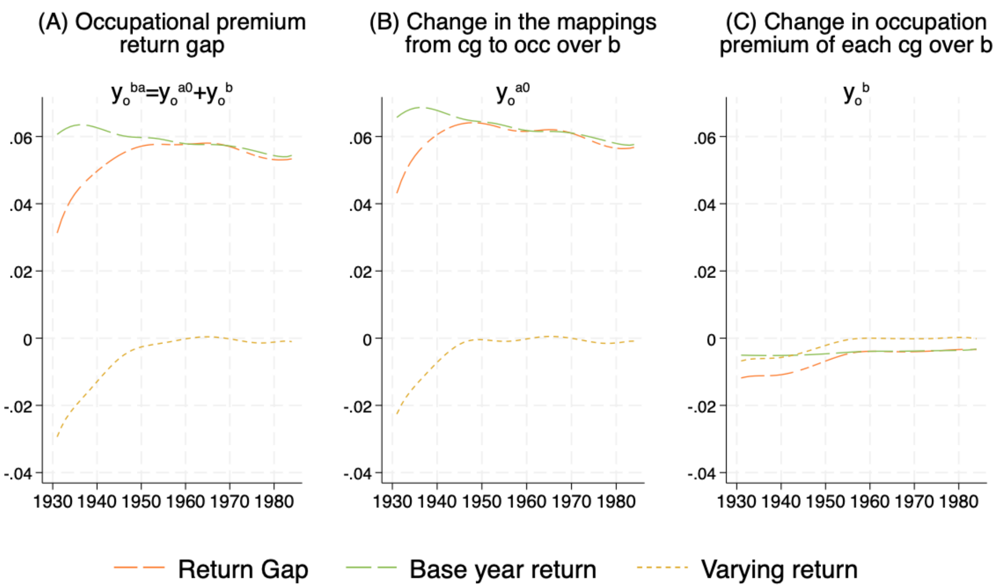
Figure F.4: OLS Decomposition for Occupation Premium using Moving Average Estimates of P_c^{fb} and P_c^{mb}, Constant Returns



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. Panel A and B can be compared to panel A and B from Figure 6.

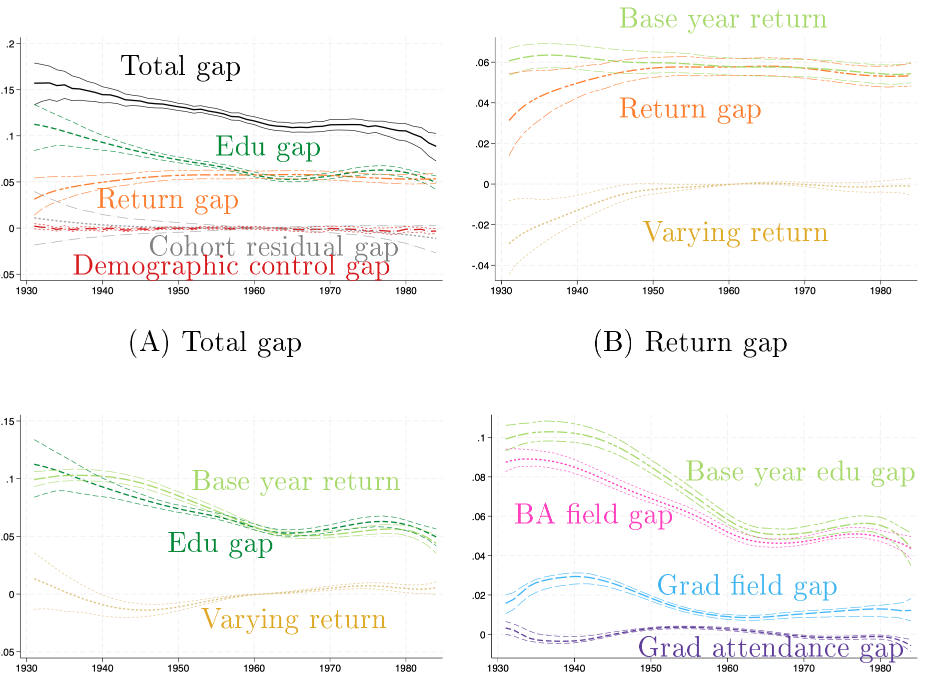
G Decomposition of the Gender Gap Using Cohort and Age Specific Occupation Premiums

Figure G.1: Decomposing the Varying Return Gap for Occupation Premium



Notes: This figure decomposes the cohort-specific relative return gap in pink in panel A. Panel A is the same as Figure 8 Panel B. In panel B and panel C of this figure, we replace the dependent variable in the regression model (3) with $\bar{y}_o^{0at}$ and δ_o^b, respectively, as defined in section 4.5. The lines are define in the same way as panel A. By construction of the dependent variable, the sum of panels B and C equals to panel A. In all three panels, the varying relative return line is the graph of $\sum_c \sum_g \left(\delta_{cg}^{mb} - \delta_{cg}^{fb}\right) P_{cg}^{fb}$, but the values of $\delta_{cg}^{mb} - \delta_{cg}^{fb}$ depend on the dependent variable for the regression model that is used in the panel. The y-axis is consistent across panels, which facilitates comparison of the magnitudes.

Figure G.2: OLS Decomposition of 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 at the average age distribution. 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 section 4.4.1and 4.5. The definitions of the lines are the same as Figure 4, but refer to the occupation premium rather than earnings.

H Decomposition of Earnings and Occupation Premium by Fields of Study

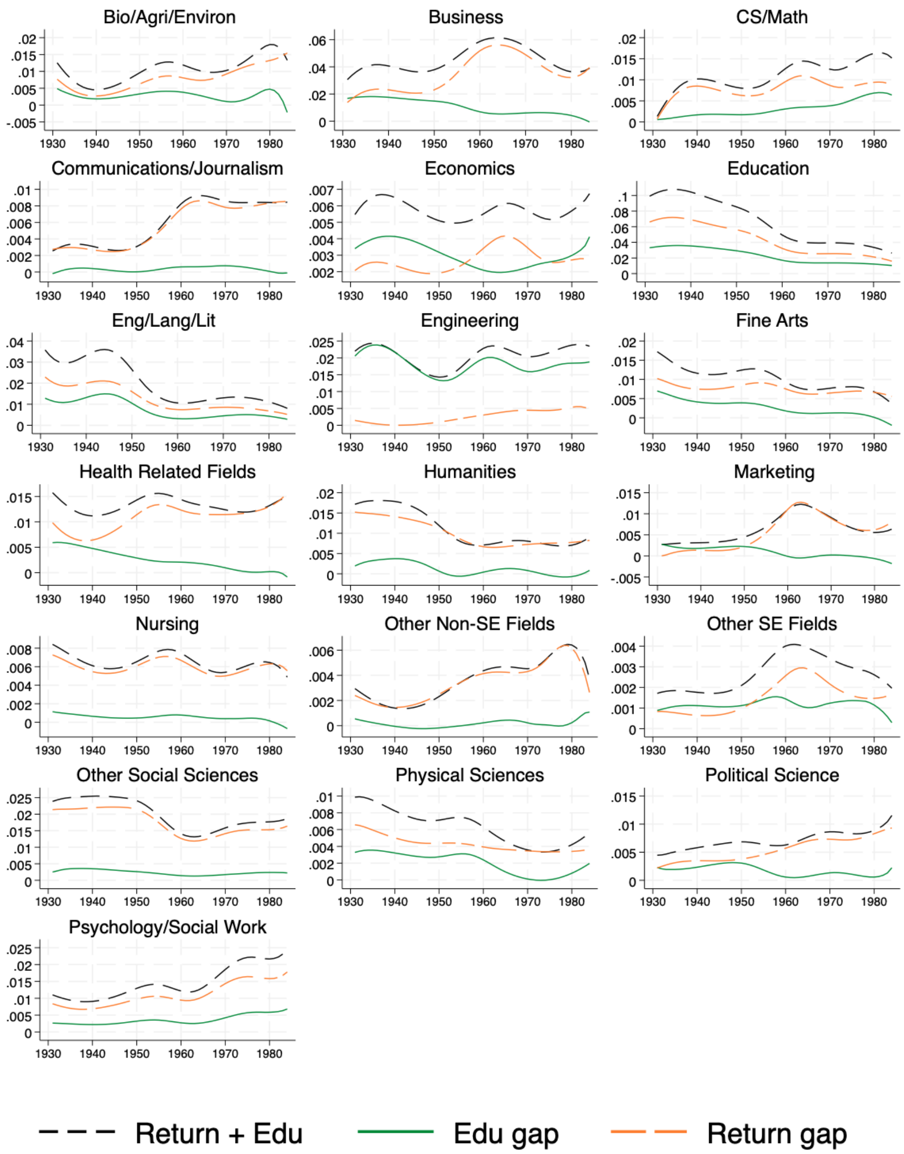
In this appendix, we disaggregate the decompositions by college majors. For all figures in this appendix, the color scheme and the definition of the lines are consistent with the main text. When the magnitudes of the gaps are too distinct, we use multiple y-axes. The majors are ordered alphabetically. We will list the primary contributing majors for each decomposed gap.

Figure H.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. For the total gap, Engineering and Business contribute the most to the decreasing trend. Education has an increasing trend, going against the overall decreasing trend. For the return gap, Education sees the largest drop from 0.07 down to 0.02, while Business increased from 0.015 to 0.055 in the 1960s and back down to 0.04 in 1984. The education gap is much more volatile than the return gap and dominates the trends of the major-specific total gap. Figure H.2 shows the disaggregation using the female coefficients and male composition. It shows the same patterns as the male coefficient version.

Figure H.3 disaggregates the return gap and the education gap in Figure 6. Business and Engineering contribute the most to the decrease in the total and education gaps. The gaps for Education major increased from -2 to -0.5. The return gaps have much smaller magnitudes. Education major decreases from 0.025 to 0.007. Business increases from 0.004 to 0.014 in the 1960s. Figure H.4 shows the disaggregation using the female coefficients and male composition.

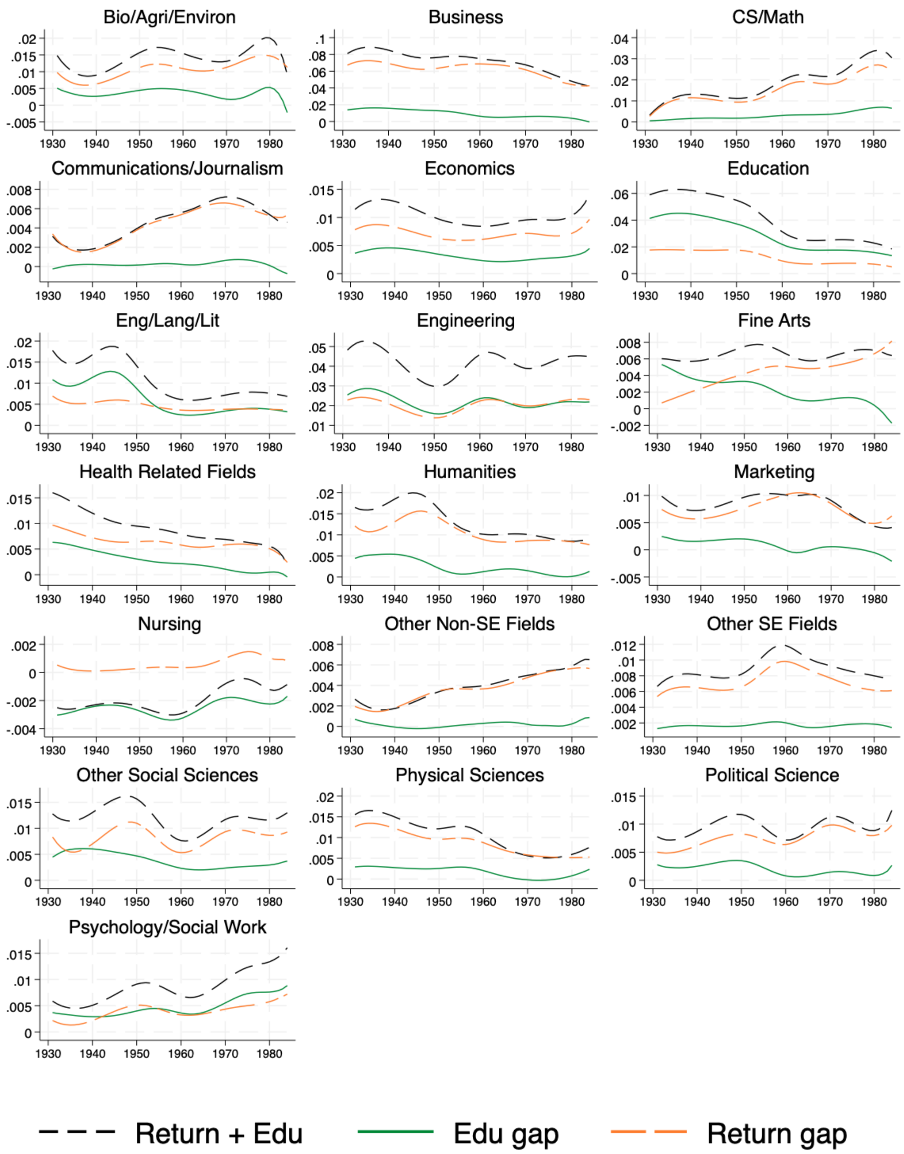
Figure H.5 disaggregates the lines in Figure 4 and Figure H.6 uses female coefficients. Figure H.7 disaggregates Figure 8, and Figure H.8 uses female coefficients. The contributors are the same as the constant return specification.

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



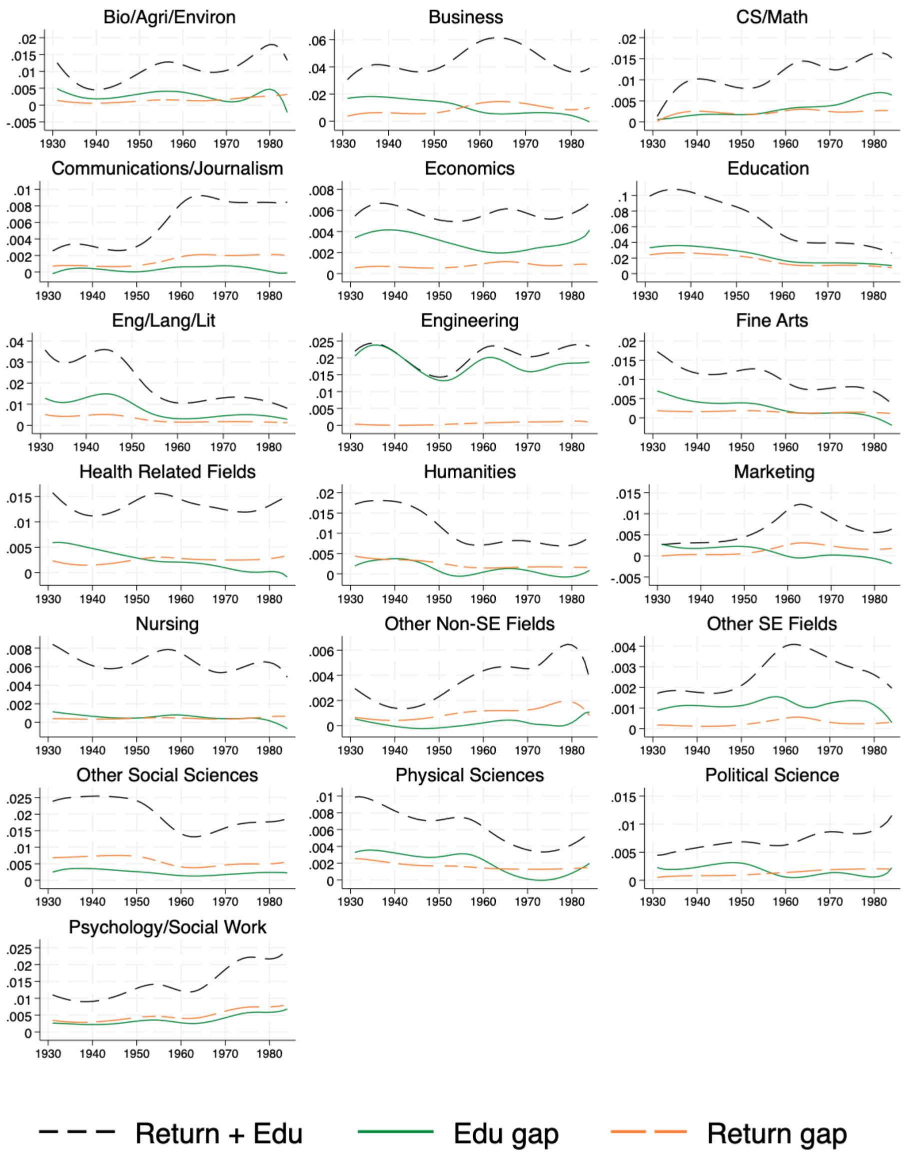
Notes: This Figure shows the by degree contribution towards the gaps shown in figure 2. The green line is the sum of the education gap for a given college major. The orange line is the sum of the return gap for a given college major. So they sum over all graduate fields plus the category of never attended graduate school (BAonly). The black line is the sum of the education and return gap, and does not include the birth cohort fixed effect or the demographic controls.

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



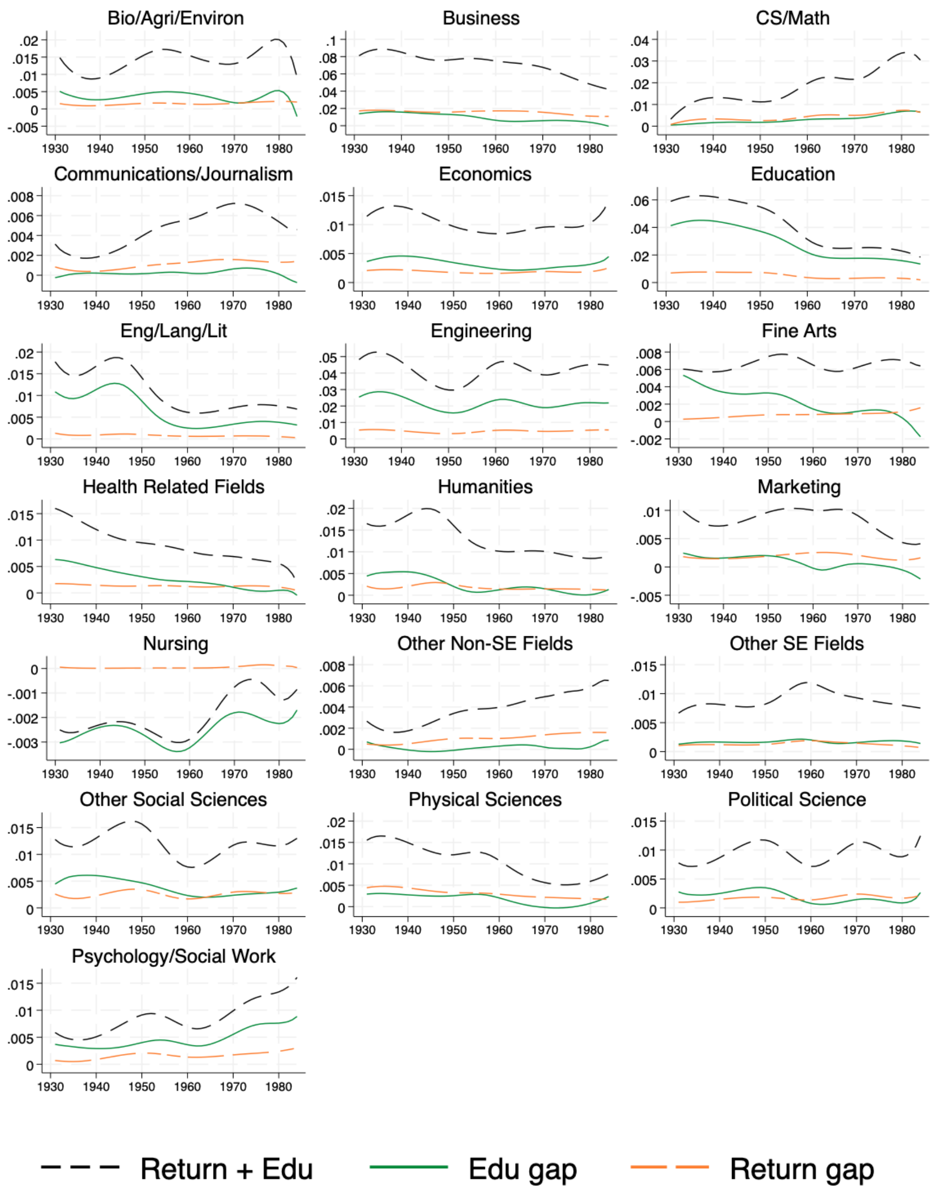
Notes: Figure shows the by degree contribution towards the gaps shown in Figure D.2. For the definition of the lines, see the notes of Figure H.1.

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



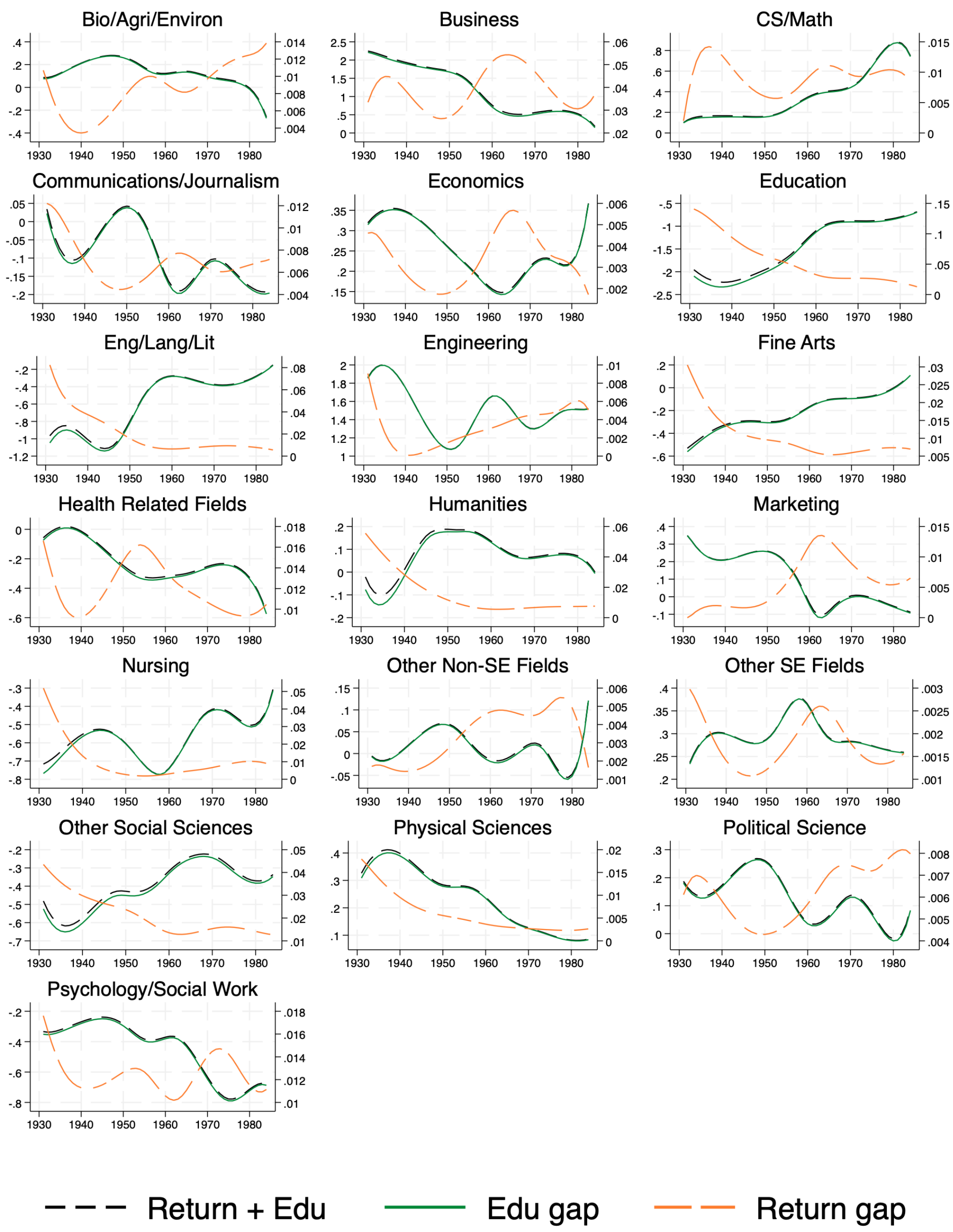
Notes: Figure shows the by degree contribution towards the gaps shown in Figure 6. For the definition of the lines, see the notes of Figure H.1.

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



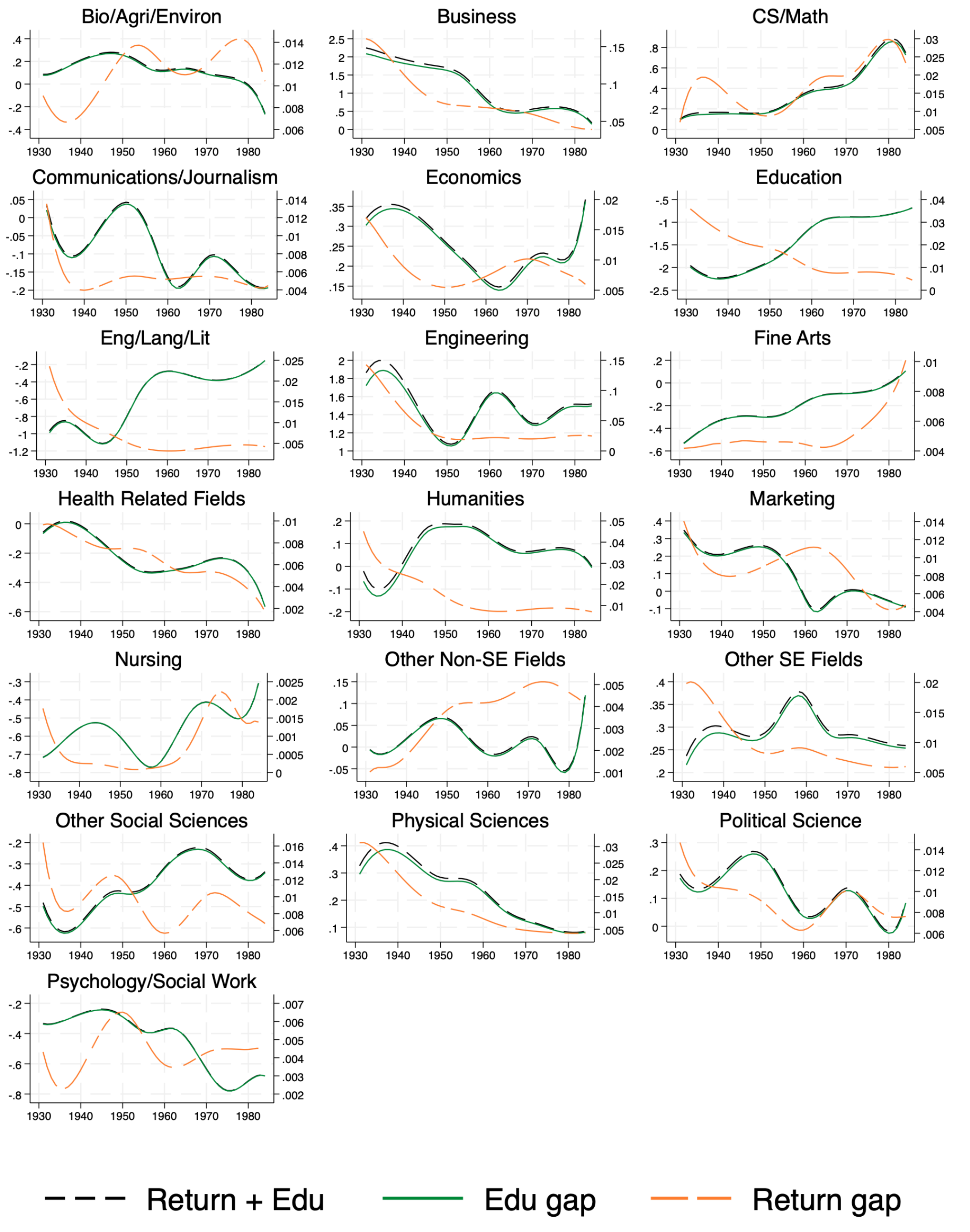
Notes: Figure shows the by degree contribution towards the gaps shown in Figure D.2. For the definition of the lines, see the notes of Figure H.1.

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



Notes: Figure shows the by degree contribution towards the gaps shown in Figure 4. For the definition of the lines, see the notes of Figure H.1.

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

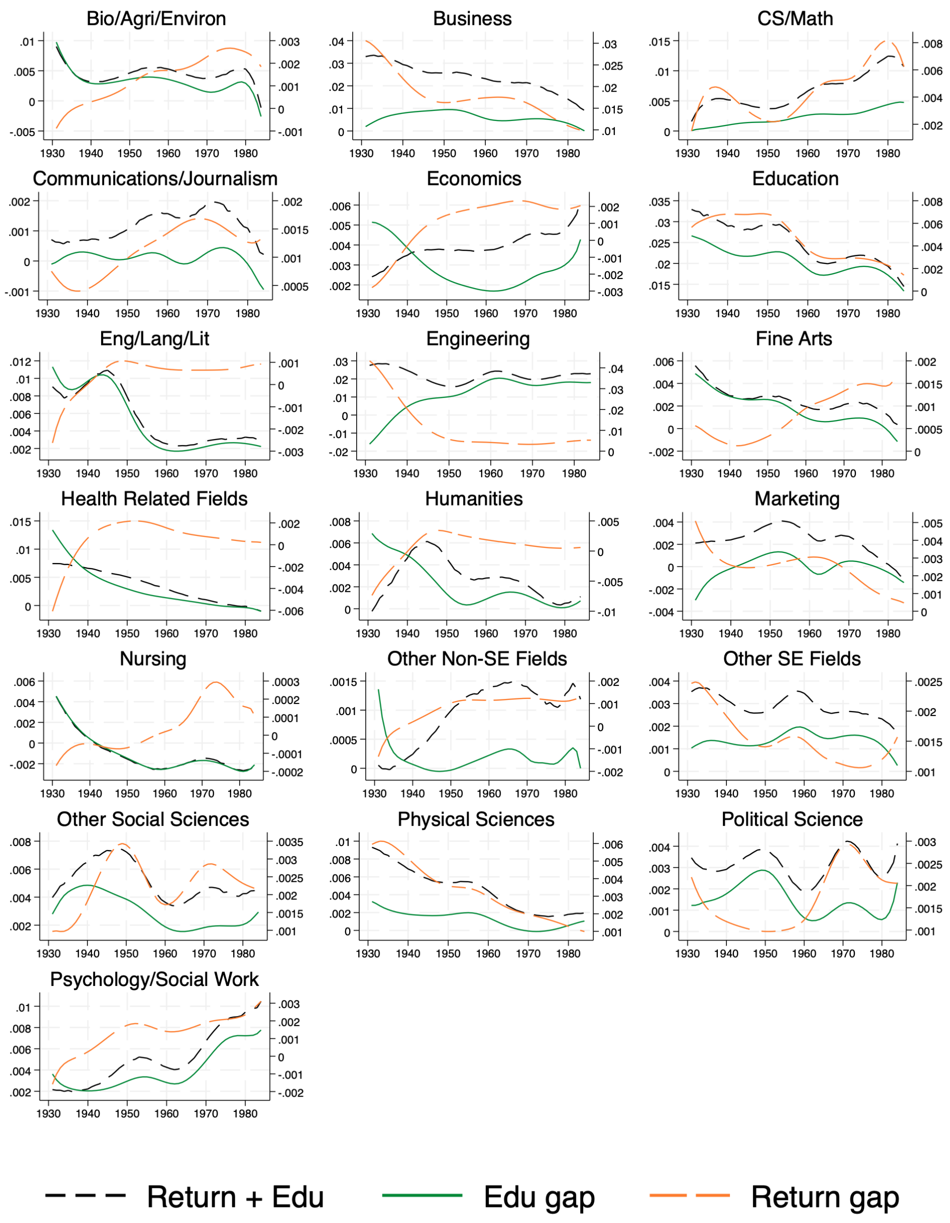


Notes: Figure shows the by degree contribution towards the gaps shown in Figure D.3. For the definition of the lines, see the notes of Figure H.1.

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

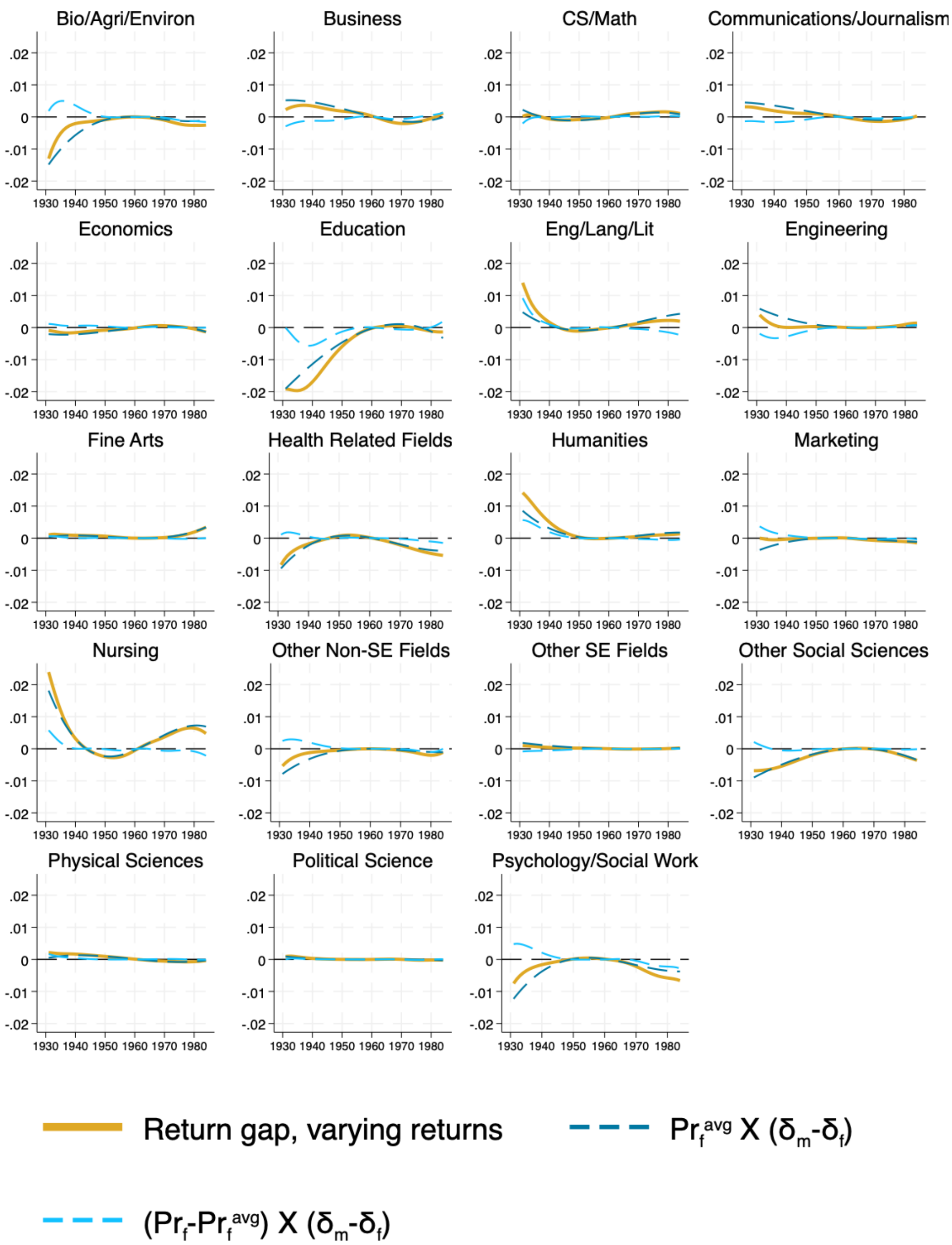
Notes: Figure shows the by degree contribution towards the gaps shown in Figure 8. For the definition of the lines, see the notes of Figure H.1.

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



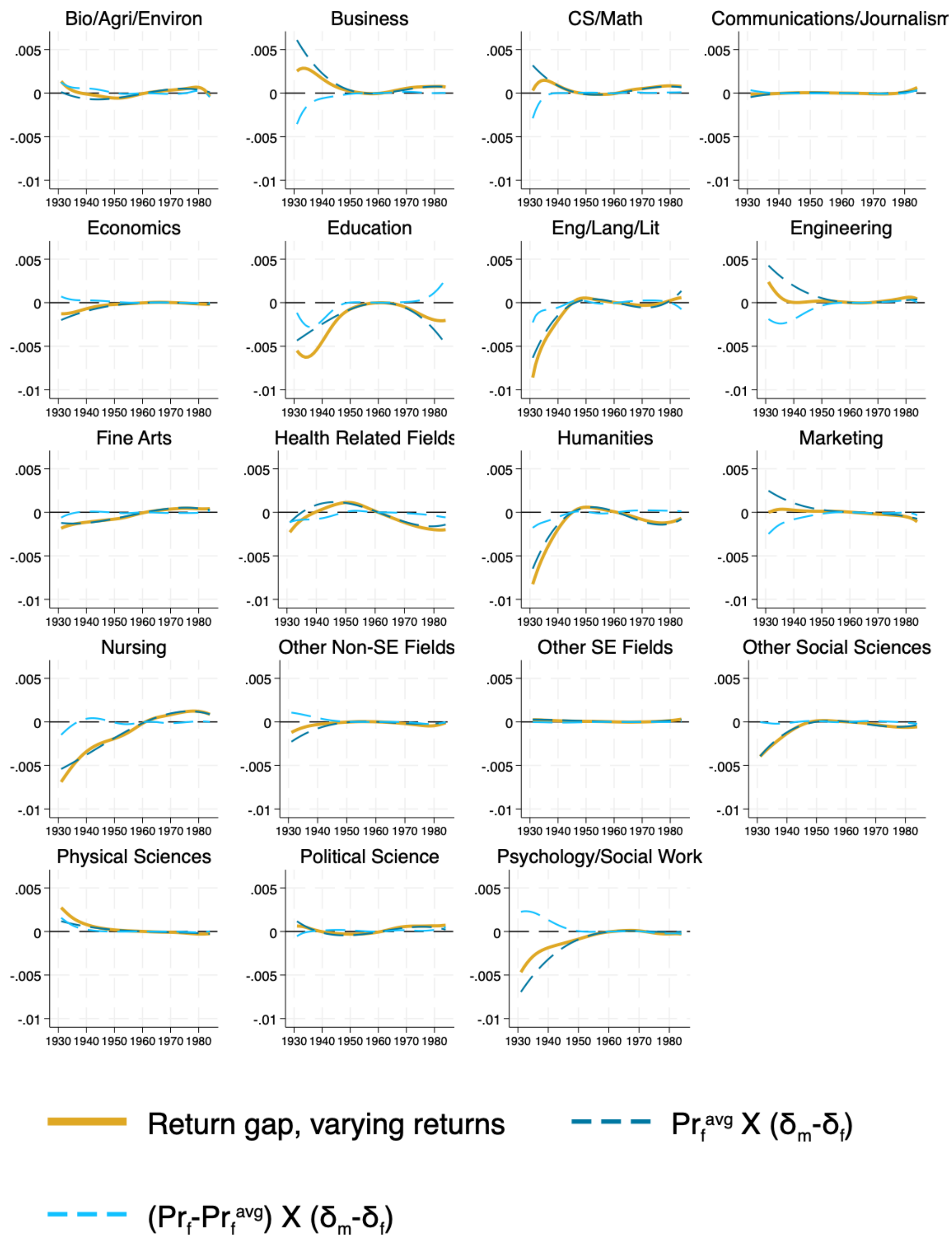
Notes: Figure shows the by degree contribution towards the gaps shown in Figure D.4. For the definition of the lines, see the notes of Figure H.1.

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



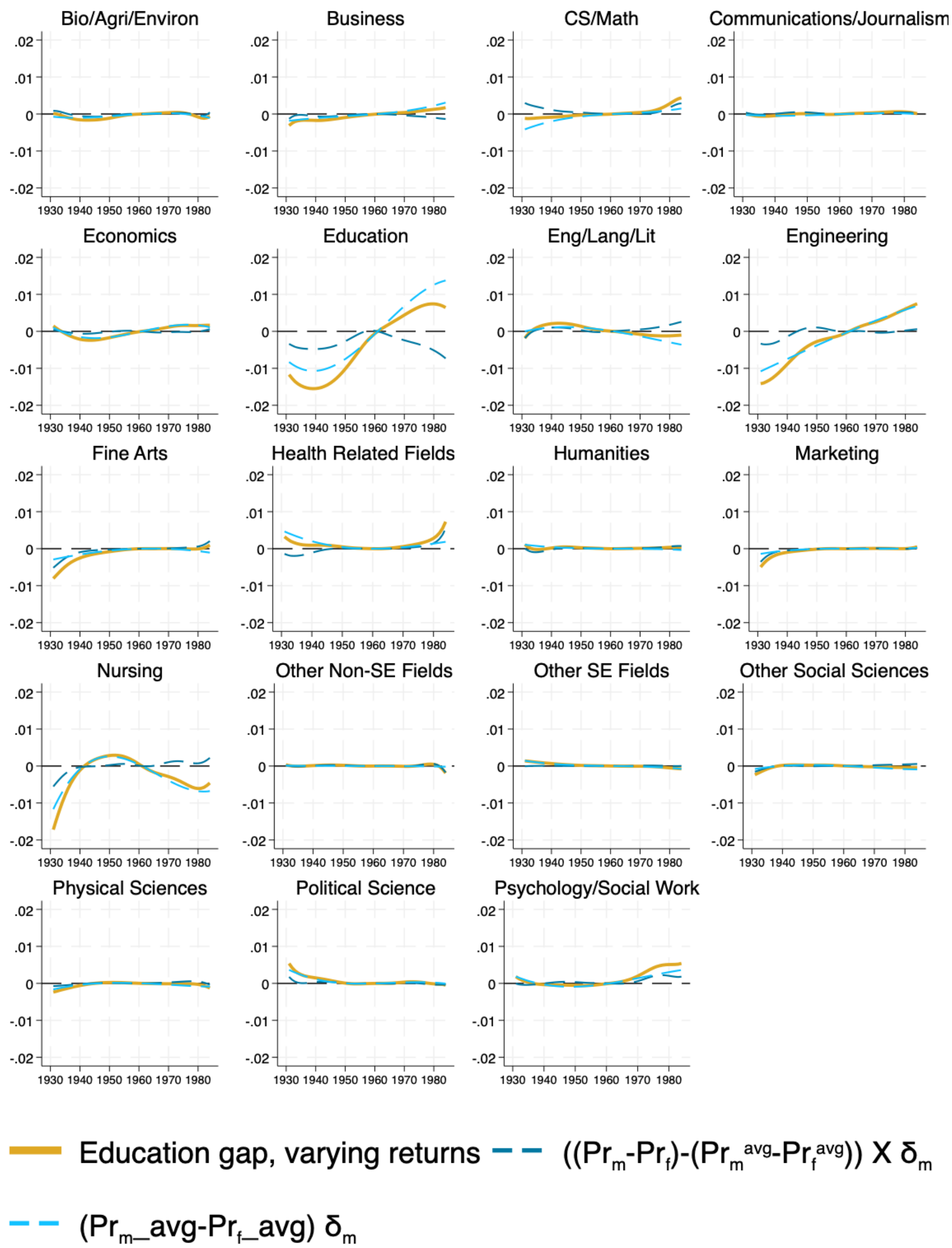
Notes: Figure shows the contribution of each college major towards the gaps shown in Figure 5 panel C.

Figure H.10: Decomposing the Varying Return Gap for Occupation Premium by College Major, Male Coefficient



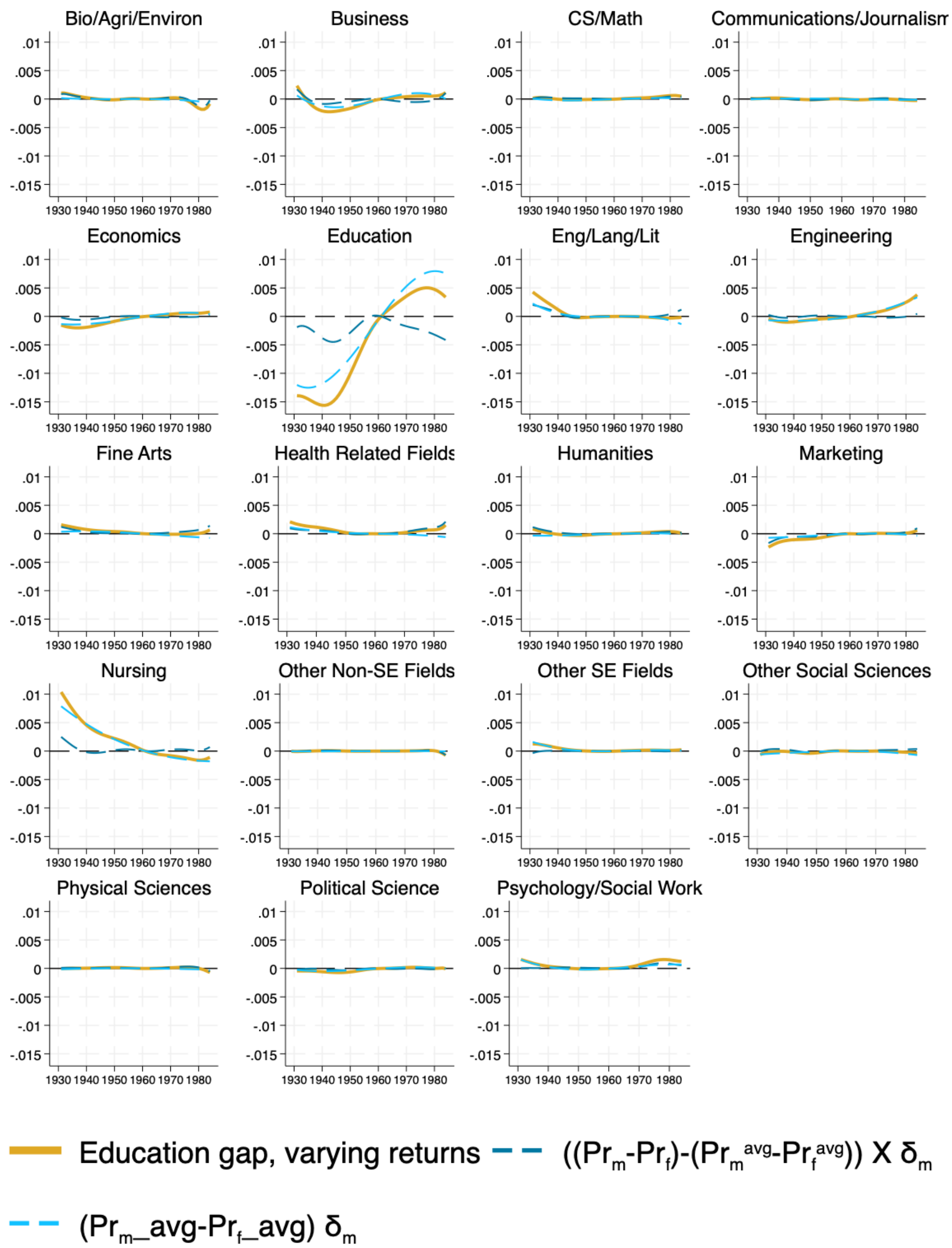
Notes: Figure shows the contribution of each college major towards the gaps shown in Figure 5 panel D.

Figure H.11: Decomposing the Varying Return Education Gap for Log Earnings by College Major, Male Coefficient



Notes: Figure shows the contribution of each college major towards the gaps shown in figure 5 panel E.

Figure H.12: Decomposing the Varying Return Education Gap for Occupation Premium by College Major, Male Coefficient



Notes: Figure shows the contribution of each college major towards the gaps shown in figure 5 panel F.

I Decomposition Tables

Table I.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: Table shows the predicted gender gap in log earnings for each birth cohort at the average age distribution shown in panels 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 restricted to be between 23 and 59. By construction, Total gap = Return gap + Education gap + Birth Cohort residual gap + Demographic control gap.

Table I.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: Table shows the predicted gender gap in occupational premiums for each birth cohort at the average age distribution shown in panel A of Figure 4. 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. The methodology is 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 restricted to be between 23 and 59. By construction, Total gap = Return gap + Education gap + Birth Cohort residual gap + Demographic gap.

Table I.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: 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 8. Bootstrapped standard errors are reported in parentheses. The standard errors are estimated from 200 bootstrap iterations. The methodology is explained in section 4.5. By constant, Total Gap = Birth Cohort Residual Gap + Return gap (Base Year Return) + Return gap (Varying Return) + Edu gap (Base Year Return) + Edu gap (Varying Return) + Demo Control Gap.

Table I.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: 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 component and cohort varying component of the returns and education gaps respectively. The coefficient estimates are from regression model 8. Bootstrapped standard errors are reported in parentheses. The standard errors are estimated from 200 bootstrap iterations. The methodology is explained in section 4.5. By construction, Total Gap = Birth Cohort Residual Gap + Return gap (Base Year Return) + Return gap (Varying Return) + Edu gap (Base Year Return) + Edu gap (Varying Return) + Demo Control Gap.

Table I.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: 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 (8). 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 (8) 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 (8) when we decompose the gender gap in occupation premium. It is the black line in Figure 8 panel A. The overall gap within occ 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 I.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: 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 10. 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 occ education gap (second to last line) 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 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.

J 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 our sample 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 don't 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 data sets are longitudinal studies that include standardized achievement or cognitive ability tests and follow individuals into adulthood. Our earliest longitudinal study is Project Talent, 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, 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, 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, a survey of individuals who were 12-16 at the start of 1997.

Starting with SAT scores, Figure J.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 average quantitative scores. The x-axis on both plots is birth cohort, assuming exams are taken at age 17. The y-axes are 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 for 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 over time. Figure J.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 J.1 reports all graphed averages and differences plotted in Figure J.2.

Beginning with college graduates, Figure J.2 panel A shows that college graduates had an average test score percentile between 65 and 75 in all four data sets.[32] 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 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 J.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 J.2 show that the gender gap in test scores among college graduates working full time increased substantially from the PT to the NLSY79 and

[32] 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.

then shrunk some from the NSLY79 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.[33] Figure J.4 panel 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 NSLY97 (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 J.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 J.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 NSLY79 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 J.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

[33]Note that we do not observe labor supply at the same age in every data set. In particular, we observe individuals in the PT data around age 29 and individuals in the NLS around age 32. These small differences may affect differences between data sets, in particular for women who may be more likely to work part time when there is a young child in the household.

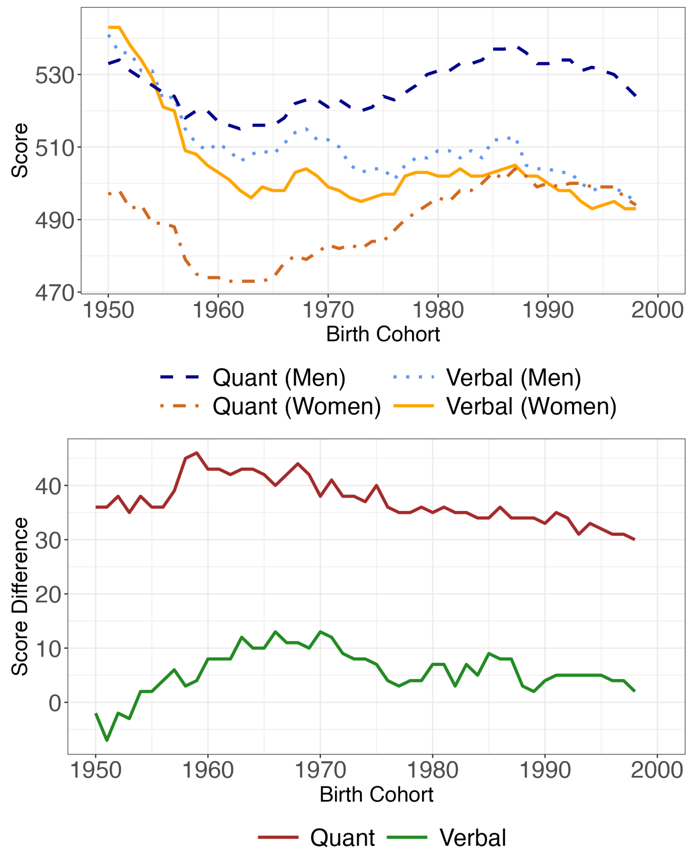
full time fell between the late 1940s and early 1960s birth cohorts as the share of women graduating from college, earning graduate degrees, and working full time all increased. Since we cannot control for test scores in our primary analysis, we miss this trend, which may explain part of the residual gender gap in earnings.

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

Birth	Prop w/ BA			Prop w/ MA			Prop 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

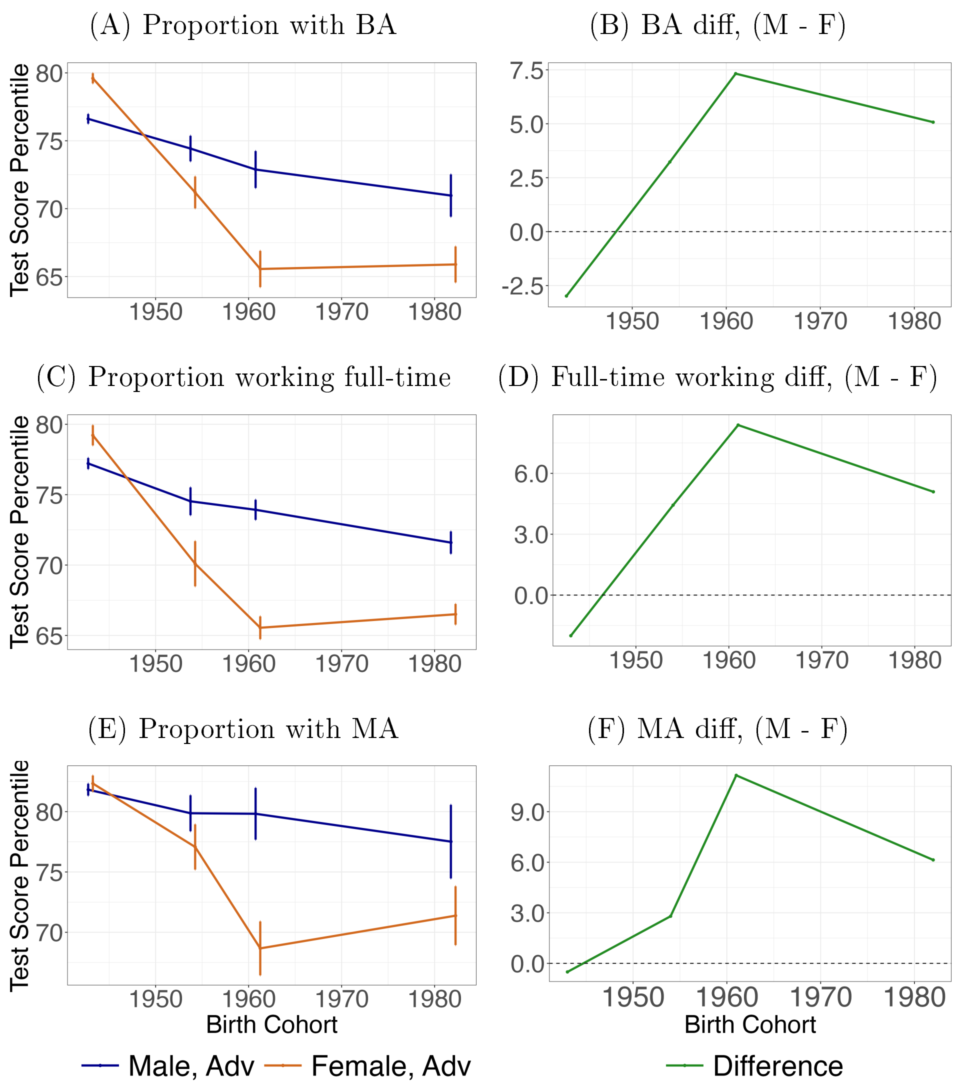
Notes: Table shows the average test score percentile for males, females, and their difference for those who selected into: college graduation (BA), full-time work (working FT), and graduate degrees (MA). Those who selected into full-time work and graduate degrees are restricted to college graduates. 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 J.1: SAT Scores for Men and Women Over Time



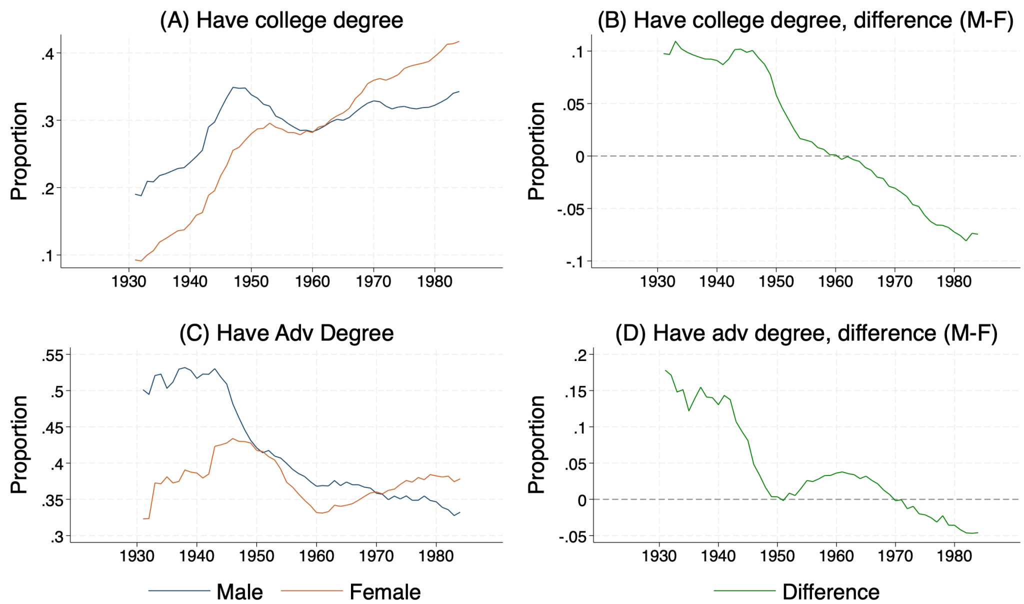
Notes: Figure plots the average SAT scores on the quantitative and verbal sections for college-bound high school seniors who took the SAT. Data is from the College Entrance Examination Board annual reports, which was then collected and harmonized and reported by the NCES in the Digest of Education Statistics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

Figure J.2: Selection into college graduation, graduate degree attainment, and full-time work on test scores by gender



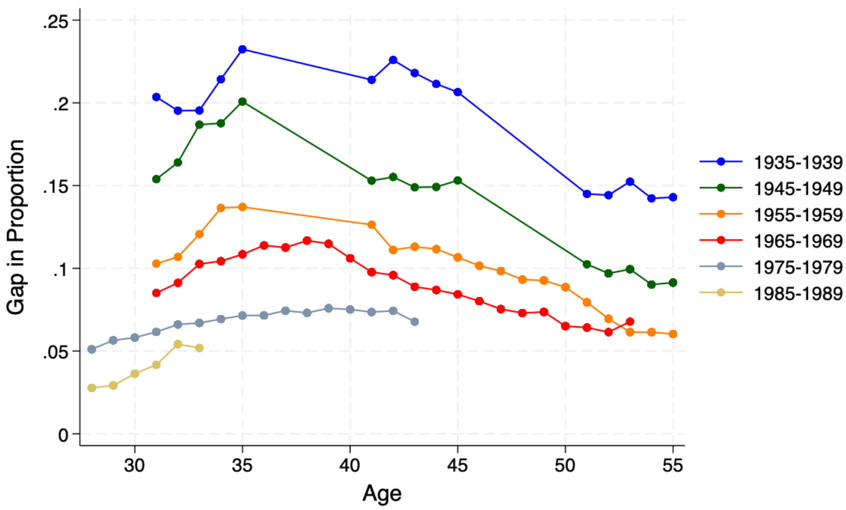
Notes: Panels A, C, and E show the selection into college graduation, graduate degree attainment, and full-time work on test scores by gender, respectively. Panels B, D, and E show the difference between the two genders for each respective left-hand panel. The graduate degree attainment and full-time work estimates are conditional on receiving a BA. These figures use use data from Project Talent, The National Longitudinal Study 1972, the National Longitudinal Survey of Youth 1979, and the National Logitudinal 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 NSLY97. Test score percentiles are the weighted percentile of achievement or cognitive tests given in each survey.

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

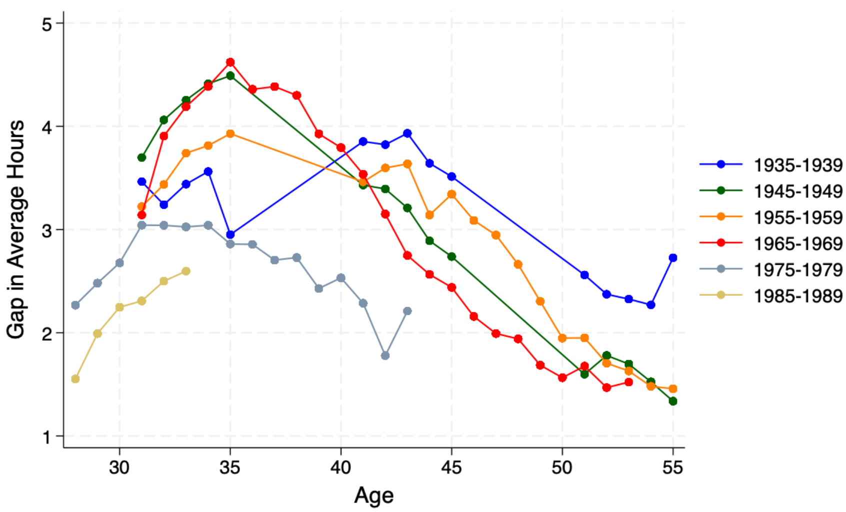


Notes: This figure uses data from the ACS (2001-2018) and the decennial Census (1960, 1970, 1980, 1990, 2000). Panel A shows the proportion of the birth cohort that have an undergraduate degree disaggregated by gender. Panel B shows the difference (male - female) of panel A. Panel C shows the proportion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

Figure J.4: Gender Gap of 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 This figure shows the difference in the proportion of the population, conditioned 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.