

WORKING PAPER · NO. 2019-121

A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap

Carolyn Sloane, Erik Hurst, and Dan Black

OCTOBER 2019



**Becker
Friedman
Institute**

Initiative for the
Study of Gender
in the Economy

Macroeconomic
Research
Initiative

Ronzetti Initiative
for the Study of
Labor Markets

A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap *

Carolyn Sloane

Erik Hurst

Dan Black

September 29, 2019

Abstract

This paper explores the importance of pre-market human capital specialization in explaining gender differences in labor market outcomes among the highly skilled. Using new data with detailed undergraduate major information for several cohorts of American college graduates, we establish many novel facts. First, we show evidence of a gender convergence in college major choice over the last 40 years. Second, we highlight that women today still choose college majors associated with lower potential wages than men. Third, we report gender differences in the mapping from major to occupation. Even conditional on major, women systematically choose lower potential wage and lower potential hours-worked occupations than men. Fourth, we document a modest gender convergence between the 1950 and 1990 birth cohorts in the mapping of major to occupation. Finally, we show that college major choice has strong predictive power in explaining gender wage gaps independent of occupation choice. Collectively, our results suggest the importance of further understanding gender differences in pre-labor market specialization including college major choice.

*We thank seminar participants at the University of Chicago for helpful comments. Author affiliation and contact information: Carolyn Sloane, University of California, Riverside, csloane@ucr.edu; Erik Hurst, University of Chicago Booth School of Business, erik.hurst@chicagobooth.edu; and Dan Black, University of Chicago, Harris School of Public Policy, danblack@uchicago.edu.

1 Introduction

There is an extensive literature studying the fundamental changes to women’s labor market outcomes in the U.S. during the post-War period.¹ Some of this literature has documented the phenomenal growth in the relative supply of college-educated women.² This shift has changed the gender composition of the college-educated labor force. At the same time, the specialized skills of college-educated men and women have also been evolving. While post-schooling labor market specialization (e.g. occupational choice) has been identified as an important determinant of gender gaps in wages and employment, the impact of pre-market specialization (e.g. major choice) has not received as much attention.³ The menu of major choices offered by US post-secondary institutions is large. Given changes over time in occupational specialization, we suspect that cross-cohort dynamics in pre-market specialization may be important. Until recently, large, multi-cohort data linking detailed college major choice to subsequent labor market outcomes was not previously available. Due to this empirical constraint, there is relatively little work on the impact of *pre-market human capital specialization* on gender wage gaps among the highly-educated across cohorts.

In this paper, we use new data to explore the importance of gender differences in pre-labor market specialization on gender differences in labor market outcomes. Specifically, the paper addresses the following four questions: (1) To what extent have the pre-labor market human capital specialization decisions of college-educated men and women converged over time? (2) Does gender influence the mapping of college major to subsequent occupational specialization and has this mapping changed over time? (3) How much of college gender wage and employment gaps can be attributed to major choice independent of occupational choice? (4) How has the relationship between major choice and gender wage and employment gaps evolved over time?

To address the above questions, we exploit newly released data from the American Community Survey (ACS) which added questions starting in 2009 on pre-labor market specialization, namely major choice, for millions of college-educated individuals.⁴ These questions

¹See, for example, Altonji and Blank (1999), Bertrand and Hallock (2001), Black and Juhn (2000), Blau and Kahn (1997), Blau and Kahn (2000), Goldin (1992), and Jacobsen et al. (1999). For a recent, detailed review of this literature, see Blau and Kahn (2017).

²See Becker et al. (2010), Charles and Luoh (2003), DiPrete and Buchmann (2006), Goldin et al. (2006), and Jacob (2002).

³The literature documenting both changes in occupational sorting by gender over time and the contribution of occupational choice to gender labor market disparities includes Bayard et al. (2003), Blau et al. (1998), Blau et al. (2014), Cortes and Pan (2018), Goldin (1992), Groshen et al. (1987), Hsieh et al. (2019), Macpherson and Hirsch (1995), and Pan (2015).

⁴The ACS has been used by Altonji et al. (2014) to measure the evolution of wage inequality within major and Altonji et al. (2016) to estimate major-specific returns.

are asked regardless of age. This allows us to explore how pre-labor market human capital specialization decisions have changed across cohorts.

We begin by assessing the extent to which men and women choose different college majors and how those differences have evolved over time. We document that women made substantial progress in traditionally male-dominated majors such as business and the physical and life sciences. Younger birth cohorts of women are even more likely to be biology majors than men. In contrast, although *relative* growth of women in engineering is large, engineering remains a male-dominated field of study. The fraction of women majoring in education, a historically female-dominated field of study, declined substantially. However, the decline among males was even larger making the education field even more heavily female-concentrated among recent birth cohorts.

In addition to using traditional indices to summarize trends in gender segregation, we develop a new index that measures the potential wage gap between women and men based on major choice. We define a major's potential wage based on the median log hourly wage of native, white men aged 43-57 who matriculated with that major. With this measure, we compare the major choice of women relative to the major choice of men for each birth cohort where the units of the index are in potential log wage differentials. We find that across all birth cohorts women systematically choose majors with lower potential wages relative to men. College-educated women born in the 1950s matriculated with majors that had potential wages that were 12% lower than men from their cohort. That gap fell to about 9% for the 1990 birth cohort. While there has been convergence in major choice between men and women during the last 40 years, the youngest birth cohorts of women still choose majors with lower potential wages than men. We also highlight that the trend in the gender similarity of major choice is non-monotonic with much of the convergence occurring between the 1950 and 1975 birth cohorts with a modest divergence for recent cohorts.

In the next part of the paper, we measure gender differences in the mapping of college major to subsequent occupational specialization and explore how this mapping has changed over time. We show that convergence in major choice is of a similar order of magnitude as the convergence in occupational choice. While women systematically are in lower wage occupations conditional on major choice, this gap has narrowed over time. For example, for the 1950 birth cohort, women who majored in engineering chose subsequent occupations with potential wages that were 14 percent lower than men from their cohort who majored in engineering. For the 1990 birth cohort, however, women who majored in engineering ended up working in occupations with roughly the same potential wages as men who majored in engineering. These patterns have nothing to do with women earning less than men within an occupation as we only measure an occupation's potential wages based on what men earn

in that occupation. These patterns stem from the fact that the subsequent occupations chosen by women who majored in engineering used to differ markedly from men but now the occupational choice conditional on major has converged.

The patterns for engineering are broadly similar to the patterns for all majors. For the 1950 birth cohort, women chose occupations *conditional on major choice* with potential wages that were about 11% lower than comparable men with the gap being larger for higher potential wage majors (like biochemical engineering and economics). The gender gap in occupational choice conditional on major narrowed meaningfully for recent cohorts. Interestingly, we find this convergence is driven by movements of women at the top of the major-pay distribution into higher-pay occupations.

These results show that not only has there been a gender convergence in pre-labor market specialization (college major) and post-schooling specialization (occupational choice), there has also been a gender convergence in the mapping from major choice to occupation. Compared to their predecessors, younger women are in occupations that are closer in potential wages to those chosen by male peers with the same major. We also show that the potential desire for lower hours worked on the part of women explains part, but not all, of these patterns.

In the final part of the paper, we ask how much of the gender gap in wages can be explained by controlling for both undergraduate major and for current occupational choice and how these relationships have evolved over time. We do this for all cohorts pooled together and separately by 10-year birth cohorts. The latter analysis lends itself to a decomposition exercise to assess how much of the change in gender wage gaps can be explained by changes in undergraduate major choice and current occupational choice. We find that the gender gap in wages for all college-educated cohorts in the 2014-2017 ACS was 23 log points after controlling for simple demographics such as highest degree completed, age, race, and state of residence. Further controlling for both major choice and current occupation reduced the wage gap to only 11 log points - a 50% reduction. Most importantly, we find that controlling for major choice has strong predictive power above and beyond controlling for just occupational choice. The gender gap in wages controlling for demographics and occupational choice is 14.3 log points. Adding undergraduate major as a control in addition to occupation and demographics further reduces the gender wage gap by 3 log points. We then compare recent (1978 to 1987) birth cohorts to older cohorts (1958 to 1967). We document that the role of college major in these cohorts remains remarkably stable, but there is a sharp reduction in the importance of occupation. Over all, we find that controlling jointly for major choice and occupational choice explains roughly 40 percent of the cross-birth cohort decline in the wage gap.

Finally, we document that undergraduate major choice does not have any effect on extensive margin labor market participation for college graduates. While undergraduate major is informative about gender wage differentials, it is not informative with respect to explaining extensive margin differences in labor supply. However, we document heterogeneous effects by gender on intensive margin participation. Specifically, we find that conditional on college major choice, women select into occupations with lower annual hours worked than men.

Our results complement a literature on gender differences in major choice. This literature has followed two separate strands. First, there is a recent literature documenting gender differences in major choice and how those differences have evolved over time. Dickson (2010) uses administrative data from three public universities in Texas and Zafar (2013) uses administrative data from Northwestern University to document gender differences in majors for one cohort of undergraduates at their respective universities. Using data from the National Center of Education Statistics, England and Li (2006) and Blau et al. (2014) (Chapter 8) document how gender differences in detailed undergraduate majors have diminished over time for a nationally representative sample of undergraduates with detailed measures of field of study. Unfortunately, the data that England and Li (2006) and Blau et al. (2014) use has no information on subsequent labor market outcomes ⁵

Second, there is a separate literature examining how gender differences in major choice affect gender differences in earnings. In a classic reference, Brown and Corcoran (1997) use data from the 1984 Survey of Income and Program Participation (SIPP) and the National Longitudinal Study of High School Class of 1972 (NLS72) to document how course work differences between men and women affect gender wages gaps. For the older cohorts (those born prior to 1960 and, thus, prior to the female overtaking in college completion), they find that the gender wage gap for college graduates falls once controlling for 20 broad major categories. Loury (1997) uses data from the NLS72 and the High School and Beyond Senior Cohort (Class of 1980) to also document that controlling for GPA and four broad major choice categories reduces the gender wage gap.⁶

The goal of our paper is to link and expand on these two strands of the literature. Specifically, we ask how *changes* in the gender gap in major choice explain *changes* in col-

⁵The literature documenting the under-representation of women in STEM fields includes see, for example, Leslie et al. (1998) and the cites within. Ceci et al. (2014) use data from the National Center for Science and Engineering Statistics to highlight the gender convergence in STEM majors. Turner and Bowen (1999) uses data from the College and Beyond database to assess gender differences in major choice within twelve academically selective colleges and universities finding even conditional on SAT scores, gender differences in major choice remain.

⁶Black et al. (2008) use data from the 1993 National Survey of College Graduates to examine the extent to which pre-labor market factors (including undergraduate major) explains differences in wages across various race-gender groups of college graduates. Their data end in 1990, about the time women's field of studying started diverging from men.

lege gender wage gap over the last half century within the United States. We are able to do this by exploiting new data that links undergraduate major and labor market outcomes for many birth cohorts. As a result, we are able to document long-run trends for a nationally representative sample of college graduates. Throughout, we assess the extent to which undergraduate major has predictive power above and beyond occupational choice. In doing so, we present novel results on gender differences in the mapping between undergraduate major and occupational choice. Finally, unlike others in the literature, we examine the extent to which differences in undergraduate major can explain gender differences in extensive and intensive margin labor market participation of college graduates. We end the paper with a discussion of the interpretation of our results.

2 Data

Our primary data source is the 2014-2017 American Community Survey (ACS).⁷ Starting in 2009, the ACS asked all respondents with a bachelor’s degree to report their undergraduate major. Given the possible impact of the Great Recession on undergraduate majors, we restrict our analysis to only include ACS respondents from the 2014-2017 surveys. Specifically, our base sample includes roughly 1.7 million observations of individuals aged 23 to 67 with a bachelors’ degree who reported their undergraduate major.⁸

Respondents are asked to report their undergraduate field of study. For those respondents with a post-bachelor’s degree, no additional information is provided for the field of study of their advanced degree(s). If individuals have more than one bachelor’s degree or more than one major, they are prompted to list multiple majors. The ACS separately records up to two separate majors for each respondent. Only 11% of our sample reports having a second major. For the sample years 2014-2017, the ACS combines major responses into 176 distinct “detailed” majors. The ACS also aggregates these detailed majors into 29 “broad” major categories. Examples of the detailed majors include journalism, economics, chemical engineering, molecular biology, music, and finance while examples of the corresponding broad major fields include communications, social sciences, engineering, biology/life sciences, fine

⁷The ACS is conducted by the U.S. Census Bureau. The ACS samples roughly 1 percent of the U.S. population each year asking detailed questions on demographics, labor market variables and family structure. We downloaded the ACS samples from IPUMS USA database. For additional details, see Ruggles et al. (2019).

⁸Of those with a bachelor’s degree, 91.5% of ACS respondents between the ages of 23 and 67 reported at least one undergraduate major. For those with missing majors, the ACS imputes the majors for these individuals by assigning them a major probabilistically based on the reports of other respondents of similar sex, race and occupation. The fraction with missing major was relatively constant between the ages of 23 and 67 but did increase sharply for respondents over 75. We exclude those with imputed majors. For more details, see the Online Appendix.

arts, and business. We assign one undergraduate major to each individual with at least a bachelor’s degree. For those with only one major, it is the major reported. For those with two majors, we choose the reported major that is associated with the highest median male labor market wage.⁹

Our analysis explores the independent contributions of educational and occupational specialization decisions to the college gender wage gap and explores gender differences in the mapping between undergraduate majors and occupational choice. For the 2014-2017 data, we use the reported occupation for all individuals in our sample with a valid, civilian occupation code who have worked within the previous 5 years.¹⁰ Lastly, we use the data on wages and employment rates from the ACS. We define real wages within the ACS by dividing self reported annual labor income by self reported annual total hours worked and then using the CPI to convert into real 2018 dollars. We classify individuals in our sample as strongly attached to the labor market if the individual reports working at least 30 hours per week for at least 27 weeks in the previous year. See the Online Appendix for additional details about the construction of all variables used in the paper.

3 The Convergence in Major Choice

In this section, we document the presence and evolution of gender gaps in undergraduate major choice and how those gaps have evolved over time. For comparison, we also document trends in relative occupational choice for college-educated men and women.

For each survey year between 2014 and 2017, we begin by assigning each individual a 5-year birth cohort based on current age in the survey year.¹¹ In the U.S., most individuals who ever complete a bachelor’s degree do so by their mid-twenties; this implies that undergraduate major is largely fixed over the life cycle. Figure 1 graphs the ratio of females to males within a broad major category.

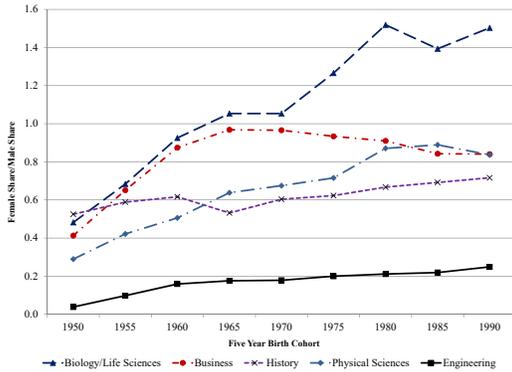
The majors in Figure 1 highlight the heterogeneity with respect to the gender composition of broad major fields. For some majors, there has been substantial gender convergence across birth cohorts. For example, for the 1950 birth cohort, the engineering major contained *twenty* men for every one woman. Today, the engineering major is still much more male-dominated, but that gap has narrowed over time. By the 1990 birth cohort, there were five men for every

⁹In computing potential wage for each major, we restrict the analysis to a sample of native, white men between the ages of 43 and 57 with strong attachment to the labor market in the prior year.

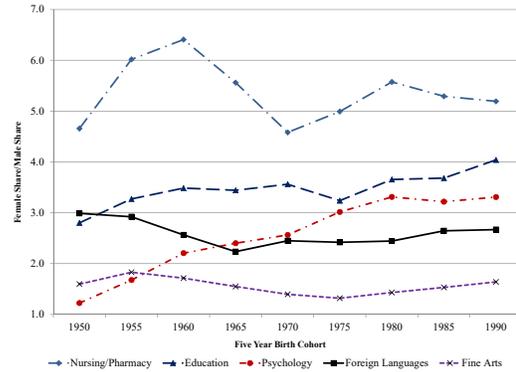
¹⁰We use a balanced panel of detailed occupation codes based on the 1990 Census detailed occupation codes following (Dorn, 2009). This yields 251 detailed occupation codes in our analysis.

¹¹We center the 5-year birth cohorts around years that are multiples of five. For example, what we refer to as the 1950 birth cohort includes all individuals born between 1948 and 1952.

Figure 1: Gender Differences in Selected Majors by Birth Cohort



PANEL A: HISTORICALLY MALE DOMINATED MAJORS



PANEL B: HISTORICALLY FEMALE DOMINATED MAJORS

Notes: These figures plot the ratio of females to males within major category. The left panel shows trends for a set of majors where men outnumber women. The right panel shows trends for a set of majors where women outnumber men. Data from the 2014-2017 ACS and is restricted to those with at least a bachelor’s degree. See text for additional details.

one woman in the engineering major. These patterns are shown in the solid line in Panel A. Similar convergence patterns are seen for the physical sciences (e.g., chemistry, physics, astronomy) and for the biology/life sciences majors (e.g., biology, molecular biology, genetics, ecology). In fact, biology/life sciences switched from being a major field dominated by men (for the 1950-1970 birth cohorts) to one dominated by women (the 1980-1990 birth cohorts). The business major displays a different pattern: women converged toward men between the 1950-1965 birth cohorts and then diverged thereafter. History was male-dominated historically and experienced little convergence or divergence over subsequent birth cohorts.

Similar heterogeneity in trends are seen in historically female-dominated majors (Panel B). There was gender convergence over time in the nursing/pharmacy major. In the 1950 birth cohort, there were five women for every one man in this major. For the 1990 birth cohort, there less than four women for every one man in this major.¹² Notable gender convergence was also seen in both the foreign language and fine arts majors. Like history, the education major saw little convergence or divergence between women and men over the last 50 years. Psychology majors, however, were more likely to be populated by women in the 1950 cohort and became even more female-intensive by the 1990 cohort.

¹²The broad major field referred to as “Nursing/Pharmacy” represents a broad category of health-related majors: nursing, pharmacy, treatment therapy professions, community and public health, and miscellaneous health medical professions.

As seen in Figure 1, some majors moved closer to gender parity and others moved further from parity from the 1950 to 1990 birth cohorts. We perform two exercises to summarize the trend in *overall similarity* in major choice by gender across cohorts. Define $s_{g,c}^m$ as the share of gender group g born in 5-year birth cohort c who matriculated with undergraduate major m . First, we compute a re-normalized Duncan-Duncan segregation index of undergraduate major segregation by gender and cohort (Duncan and Duncan (1955)). Specifically, we compute:

$$I_c^M = 1 - \frac{1}{2} \sum_{m=1}^M |s_{male,c}^m - s_{female,c}^m| \quad (1)$$

where I_c^M is the re-normalized gender segregation index in major choice for cohort c and where M is the total number of *detailed* undergraduate majors reported in the ACS. We re-normalize the segregation index such that perfect major segregation by gender yields an index of 0 and perfect major integration by gender yields a Duncan-Duncan index of 1. The re-normalization implies that an increase in I_c^M indicates increasing convergence in major choice by gender. We similarly define I_c^O , the re-normalized occupational segregation index based on observed gender differences in occupational sorting.

Segregation indices have some notable shortcomings. The first is that these indices are invariant to rank (such as an earnings ordering) of the major field or occupation. In other words, the above segregation index tells us to what extent college-educated men and women have sorted into similar majors (occupations), but would take on the same value if all men were chemical engineering majors and all women were fine arts majors as it would if all men were chemical engineering majors and all women were biomedical engineering majors. Second, the units of the segregation index do not lend themselves easily to an economic interpretation. As an alternative measure, we develop an index of the impact of gender on potential wages based on pre-market educational specialization. In contrast to the segregation index, the units of this index are in wage space allowing for the influence of major rank and thus making it easier to interpret the economic magnitude of gender differences in major choice. Furthermore, the inputs of this index are useful in the ensuing empirical analysis of the college gender wage gap. A crucial input is \bar{Y}_{male}^m , a potential wage based on major that is plausibly unaffected by post-educational factors. Specifically, we define \bar{Y}_{male}^m to be the median within-major labor market log wage of a group we assume faces minimal post-educational frictions in the labor market: native, white men between the ages of 43 and 57 with strong attachment to the labor market.¹³ For example, for anyone (male or female)

¹³Market effects, such as discrimination, can affect equilibrium wages of groups that do not directly face discrimination such as white men. In fact, Hsieh et al. (2019) documents that discrimination against women

who majored in economics, we assign as their potential wage the median log wage of older native white men who majored in economics. We formally define the potential wage index as:

$$I_c^{P,M} = \frac{\sum_{m=1}^M s_{female,c}^m \bar{Y}_{male}^m}{\sum_{m=1}^M s_{male,c}^m \bar{Y}_{male}^m} - 1 \quad (2)$$

where $I_c^{P,M}$ measures the differential “potential” log wage of women of cohort c given that the female distribution of major choice in a given cohort may differ from males in their cohort. A value of $I_c^{P,M} = 0$ means that the major choices of women yield the same potential log wage as their male counterparts. A value of $I_c^{P,M} < 0$ implies that women systematically choose majors associated with lower relative potential log wages. For example, $I_c^{P,M} = -0.35$ implies that women choose majors that reduce their potential wage by 35 percent relative to males from a similar cohort. As with our re-normalized Duncan-Duncan index, an increase in $I_c^{P,M}$ implies that major choice is converging between men and women.

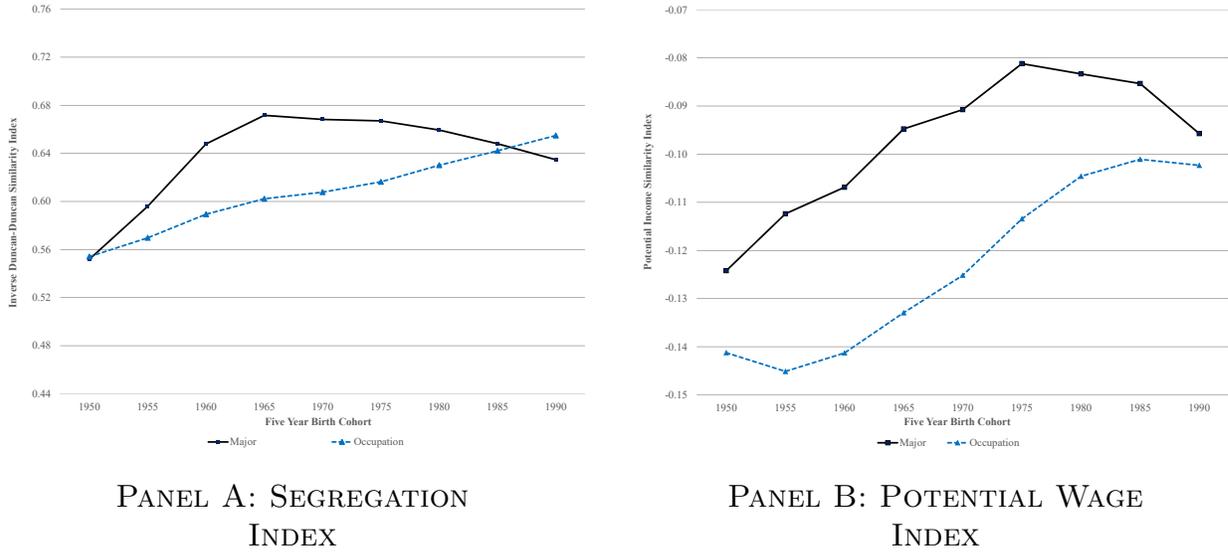
The solid line in Panel A of Figure 2 shows the trend in I_c^M , measuring major similarity across different birth cohorts. For the 1950 birth cohort, $I_c^M = 0.55$. The index increased to 0.64 for the 1990 birth cohort. Over the last half century, the US saw sharp convergence in undergraduate major choice by gender measured by the segregation index. The time series trend in the segregation index however is non-monotonic with the increased convergence occurring between the 1950 and 1965 birth cohorts. Notably, there is a reversal in the index for recent cohorts, and this reversal seems to be increasing among more recent cohorts. As seen in Figure 1, business and psychology majors saw gender divergence for recent cohorts. Nevertheless, the recent divergence is small relative to the convergence that occurred for older cohorts implying that major choice overall converged between men and women between the 1950 and 1990 cohorts. During this time period, the fraction of women with a bachelor’s degree has increased relative to men.¹⁴

The solid line in the Panel B of Figure 2 shows the trend in $I_c^{P,M}$ across cohorts. Like the re-normalized segregation index, the “potential wage” index also shows a strong gender convergence in major choice over the last half century. This index allows us to interpret the economic magnitudes of the gender convergence in college major choice over the last half century. For the 1950 birth cohort, women chose majors that reduced their potential

and blacks in the labor market raises the wages of white men. However, Hsieh et al. (2019) also estimate that these wage effects on men are small.

¹⁴We have performed a series of robustness exercises on the patterns in Figure 2. The hump-shaped pattern in I_c^M across cohorts persists for different levels broad and detailed majors and when we restrict the sample to include only those with strong attachment to the labor market. For more detail, see the Online Appendix.

Figure 2: Gender Similarity in Major Choice and Occupation by Cohort



Notes: Figure plots the segregation index, I_c^M (left panel) and potential wage gender similarity index, $I_c^{P,M}$ (right panel) for different cohorts. The solid line in each panel show the indices for major choice. The dashed line in each panel show the indices for occupation. Data from the 2014-2017 ACS and is restricted to those with at least a bachelor’s degree. See text for additional details.

wage by 12.5 percent relative to their male counterparts. By 1990, women still chose majors that were associated with lower wages relative to men, but the gap narrowed to 9.5 percent. Like the time series pattern in the segregation index, $I_c^{P,M}$ is also non-monotonic with major choice diverging slightly for recent birth cohorts. Overall, the patterns in Figure 2 illustrate a convergence in major choice between men and women particularly between the 1950 and 1970 birth cohorts with a slight divergence for recent cohorts. Moreover, it should be stressed that even for the most recent cohort, women are systematically choosing majors associated with per-hour wages that are 10 percent lower than men. The remaining gender gap in major choice is larger than the convergence in major choice experienced over the last 40 years.

As a way of comparison, Figure 2 also displays the occupation segregation index (Panel A - dashed line) and the potential wage gender segregation index for occupations (Panel B - dashed lines). To make these indices, we use 251 distinct occupation codes reported in the 2014 to 2017 ACS. These indices are the same as above except for the fact that we replace $s_{g,c}^m$ with $s_{g,c}^o$ and replace \bar{Y}_{male}^m with \bar{Y}_{male}^o . $s_{g,c}^o$ measures the share of each gender and cohort that works in a given *occupation* while \bar{Y}_{male}^o is the median labor market per-hour wage of native white men aged 43-57 working full time given that they work in occupation o (regardless of major).

Focusing on Panel A, the occupational segregation index is roughly similar in both level and trend to the major segregation index. Both indices start at a level of around 0.55 for

the 1950 cohort and end at a level of around 0.65 for the 1990 cohort. The occupation index, however, increases monotonically. While there has been a modest divergence in major choice across genders for recent cohorts, occupational choice has continued to converge.¹⁵ Panel B also shows that there was strong convergence in occupational segregation as measured by the potential wage index. College women from the 1950 birth cohort were in occupations that systematically had incomes that were 14 percent lower than the occupations of their male counterparts. The potential wage gap fell to 10 percent lower for the 1990 cohort.

Collectively, the results in this section highlight three facts about gender differences in undergraduate major choice. First, the gender gap in major choice has declined over time. Second, even for the most recent set of college graduates, a large gender gap in major choice still exists. Finally, the convergence in the gender gap in major choice among those with a bachelor’s degree is of the same magnitude as the convergence in the gender gap in occupation choice for college graduates. These patterns suggest that it is important to think about gender differences in pre-labor market specialization alongside gender differences in occupational choice. In the next section, we explore gender differences in the relationship between undergraduate major and occupational choice.

4 Gender Differences in the Mapping of Majors to Occupations

An interesting fact from Panel B of Figure 2 is that $I_c^{P,O}$ (potential wage index based on occupation choice - dashed line) is consistently lower than $I_c^{P,M}$ (potential wage index based on major choice - solid line). This implies that conditional on choosing the same major as men, women systematically work in lower-pay occupations. In this section, we provide empirical evidence that the mapping of major choice to occupation systematically differs between men and women and that this difference has declined over time.

Figure 3 plots $I_c^{P,O|m}$, a summary statistic for gender differences in occupational choice *conditional* on matriculating with a given undergraduate major for various cohorts:

$$I_c^{P,O|m} = \sum_{m=1}^M (s_{female,c}^o|m) \bar{Y}_{male}^o - \sum_{m=1}^M (s_{male,c}^o|m) \bar{Y}_{male}^o \quad (3)$$

where $s_{g,c}^o|m$ is the share of gender g choosing occupation o conditional on being in major

¹⁵We have replicated the gender convergence in occupational choice using the historical U.S. Censuses. This allow us to control for both cohort and age. Even conditional on age, the convergence in occupational choice is nearly identical to what is shown in Figure 2. See the Online Appendix for additional details.

m from cohort c . This figure focuses on the same majors highlighted in Figure 1. In other words, $I_c^{P,O|m}$ measures the gender gap in occupational rank within cohort c conditional on major choice m . Any gender difference in this index reflects gender differences in occupational wage rank conditional on major choice. $I_c^{P,O|m} < 0$ implies that women with major m from cohort c are in occupations with lower potential wages compared to otherwise similar men. $I_c^{P,O|m} > 0$ implies that women with major m from cohort c are in occupations with higher potential wages compared to otherwise similar men. If women and men are in the same occupations conditional on their undergraduate major, $I_c^{P,O|m}$ will be equal to zero by definition.

Figure 3 shows that women are consistently in occupations with lower potential wages conditional on major choice. Interestingly, this is true for both historically male-dominated majors (left panel) and female-dominated majors (right panel). In both panels, all majors have a measure of $I_c^{P,O|m} < 0$ for essentially all cohorts. Consider the engineering major (solid line, Panel A). For the 1950 birth cohort, many women who graduated with an engineering major ended up working in lower-paying occupations than men who also majored in engineering. Over time, women who majored in engineering were more likely to be in occupations that were similar to the men who majored in engineering. For engineering, our index $I_c^{P,O|m}$ increased from -0.14 to -0.02 between the 1950 and 1990 cohorts.

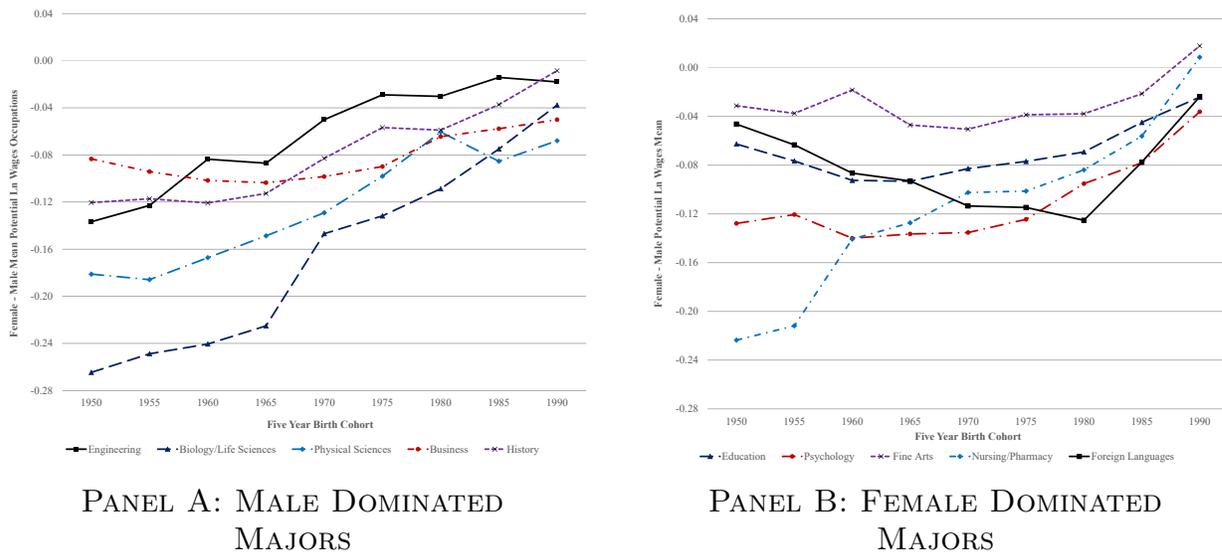
The gender convergence in occupational choice within majors is seen in many but not all of the occupations shown in Figure 3. Many of the historically female-dominated majors like education, foreign languages, and fine arts saw only modest convergence across cohorts in the occupations chosen by women relative to men (in potential wage space) conditional on major choice (Panel B). Collectively, the patterns in Figure 3 highlight the heterogeneity in both gender differences in the mapping of majors to occupation for a given cohort and how that mapping has changed across cohorts.

The information provided by describing $I_c^{P,O|m}$ in Figure 3 is somewhat limited. It only displays results for a few broad majors at a time. Further, it describes majors in a taxonomic fashion by name and gender endowment (Panel A vs. Panel B), but without an economic ranking. Thus, we create a final index, $I_c^{P,O|d,g}$ to summarize heterogeneous patterns of gendered sorting into occupations conditional on economic major rank:

$$I_c^{P,O|d,g} = \sum_{m=1}^M (s_{g,c}^o|d) \bar{Y}_{male}^o \quad (4)$$

where $s_{g,c}^o|d$ is the share of gender g choosing occupation o within major rank decile d . Using $I_c^{P,O|d,g}$ we can quantify major-to-occupation mappings as a relationship between occupation rank and major rank. We previously defined \bar{Y}_{male}^m to be the median within-major

Figure 3: Within-Major Gender Differences in Potential Wage by Occupation, by Gender and Cohort



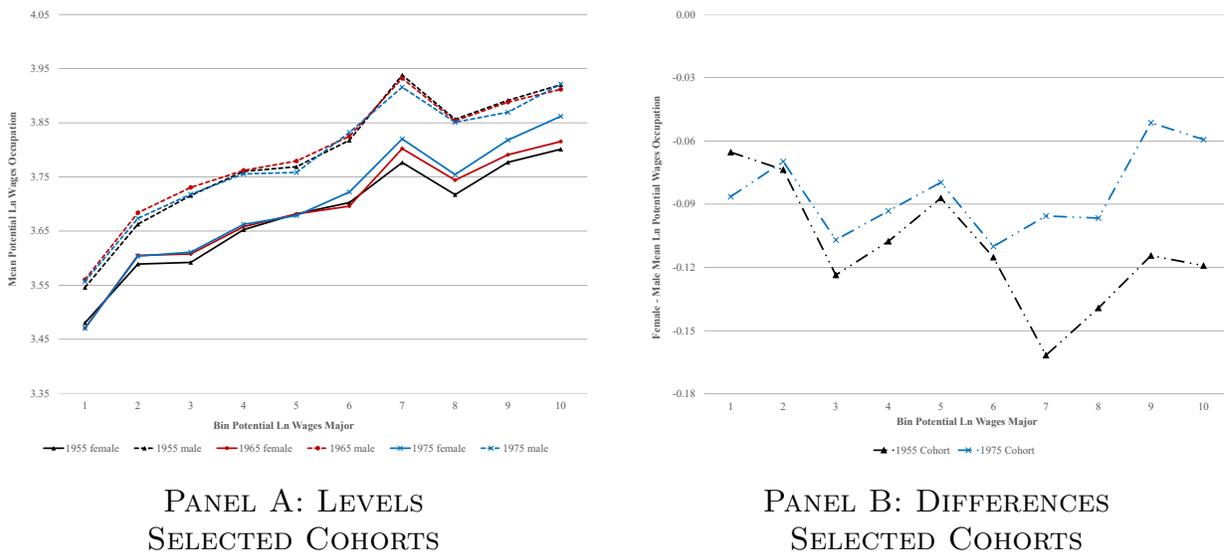
Notes: These figures show the trends in $I_c^{P,O|m}$ conditional on having graduated with major m . Panel A are male-dominated majors. Panel B are female-dominated majors. As with the left panel of Figure 2, occupational potential log wage, \bar{Y}_{male}^o , is computed in the 2014-2017 ACS using the log wages of native-born, white men 43-57 with strong attachment to the labor market.

labor market log wage of a group we assume faces minimal post-educational frictions in the labor market: native, white men between the ages of 43 and 57 with strong attachment to the labor market. In order to rank majors for this exercise, we begin by binning \bar{Y}_{male}^m by decile. Then, we average \bar{Y}_i^o , the potential log wage of individual i with occupation o , within major decile separately by gender and cohort. The index, $I_c^{P,O|d,g}$, is the within-decile mean of occupational potential log wages separately described by gender and cohort.

Our key findings are highlighted in Figure 4. The x-axis of Figure 4 segments majors into deciles based on \bar{Y}_{male}^m . The top decile includes majors like economics, chemical engineering, biochemical sciences, physics and pharmacy. The bottom decile includes majors like communications, elementary education, theology, counseling psychology, and drama and theater arts. Our mapping index, $I_c^{P,O|d,g}$, is on the y-axis. If men and women systematically choose different majors, they will be in different bins on the x-axis. Conditional on major rank, if men and women sort into different occupations, there will be variation in $I_c^{P,O|d,g}$ within a bin reflected as differences on the y-axis. If women are in lower-paying occupations conditional on major choice, the mapping of major choice (x-axis) to occupational choice (y-axis) will be systematically lower for women relative to men.

The mapping of majors to occupations for working men of different cohorts are shown in the dashed lines in Panel A of Figure 4. Each dashed line represents a different 5-year

Figure 4: Mapping of Potential Wage by Major to Potential Wage by Occupation, by Gender and Cohort



Notes: These figures show the mapping between major and occupation. On the x-axes, we have binned majors based on \bar{Y}_{male}^m , the log wage deciles of native, white men age 43 to 57 in major m . On the y-axis in Panel A, we report $I_c^{P,O|d,g}$, the mean log potential occupational wages within these deciles described separately by gender and cohort. In Panel B, the y-axis reports female - male differences in $I_c^{P,O|d,g}$ for two of the cohorts.

birth cohort of men; we highlight the 1955, 1965, and 1975 birth cohorts. The degree of monotonicity in the relationship between binned potential wages based on major on the x-axis and mean potential wages based on occupations on the y-axis reflects match quality between major and occupation *within gender*. For men, monotonicity is almost tautological. The mapping is nearly identical for older men (1955 cohort) as it is for younger men (1975 cohort).¹⁶

For women, monotonicity in Figure 4 can be violated because of gender differences in occupational choice. In fact, what is compelling about Panel A of Figure 4 is the mapping of majors to occupations for women of different cohorts (shown in the solid lines). For all cohorts, working college women are in occupations that have systematically lower wages relative to their male counterparts conditional on major choice. The gap is large. Occupations that women are in – conditional on major choice – have potential wages that are between 10 and 20 log points lower than the occupations chosen by comparable men. This has nothing to do with women earning less than men within an occupation as we only use the within-occupation wages of native, white men in their peak earning years in this figure.

¹⁶As potential wages for both majors and occupations are based on the wages of native, white men in their peak earnings years, deviation from monotonicity within the male series can only arise from race, cohort, or age effects within men.

Panel B of Figure 4 puts the information in Panel A in difference rather than level space for the 1955 (triangles) and 1975 (x's) birth cohorts. The vertical distance between the series for the 1955 cohort and the series for the 1975 cohort reflects gender convergence in the mapping between majors and occupations. In other words, women from the 1975 birth cohort systematically work in occupations that are more similar to men – conditional on major choice – than women from the 1955 birth cohort. The convergence in the gender gap in occupational choice conditional on undergraduate major was particularly strong for the highest-paying majors. Compared to the 1955 cohort, women in the 1975 cohort (1) chose majors that were more similar to men; and (2) work in occupations that are more similar to men conditional on major choice. This convergence is non-trivial. For the highest wage majors (deciles 9 and 10), women from the 1955 birth cohort chose occupations that had log wages that were 12 percent lower than comparable men. Women in these majors from the 1975 cohort now only find themselves in occupations that have log wages that 6 percent lower than men. This change in the mapping of majors to occupations is one of the key findings of the paper.

It is well-known that, in the aggregate, highly-educated women select occupations with lower hours worked (e.g., Goldin and Katz (2011) and Cortez and Pan (2016)). How much of the patterns in Figure 4 can be attributed to women systematically moving to occupations with lower hour requirements? We explore this issue in-depth in the Online Appendix and summarize the results here. We rank the potential work requirements of an occupation by computing the median annual hours worked in that occupation for men aged 43-57. We call this measure \bar{H}_{male}^o . We find that conditional on major choice, women systematically are in occupations with lower hours-worked requirements. Over all majors and across all cohorts, conditional on major choice, women are in occupations that have a work requirement that is about 3 percent less than comparable men. There is little trend in this gap across cohorts. For comparison, over all majors and across all occupations, women are in occupations that have per-hour wages that are about 10 percent less than comparable men (Panel B of Figure 4). Putting the two together, women are in occupations where annual earnings are about 13 percent lower (3% from hours + 10% from per-hour wages) with most of the effect coming from wage differences as opposed to hour differences.

5 Major Choice, Gender Wage Gaps, and Gender Differences in Participation

Previous sections established that (1) men and women sort differently into undergraduate major, (2) gendered sorting into college major has declined over time, and (3) conditional on major choice, women work in occupations with lower potential wages. In this section, we examine the extent to which these patterns are associated with the college gender wage and employment gaps both within and across cohorts. To shed light on this, we estimate regressions of the following form:

$$L_i = \alpha + \beta Female_i + \delta_m Major_i + \delta_o Occ_i + \Gamma X_i + \epsilon_i \quad (5)$$

where L_i is a measure of various labor market outcomes for individual i and where L is either the individual's log wage or a dummy variable equal to 1 if the individual is employed. $Female_i$ is a dummy variable equal to 1 if the individual is female. Our estimated variable of interest is β which measures the gender gap in either log wages or employment rates (depending on specification). $Major_i$ and Occ_i are summary measures of the individual's chosen undergraduate major and observed occupation. For our base specification, we summarize an individual's major and occupation choice with the potential log wage variables \bar{Y}_i^m and \bar{Y}_i^o .¹⁷ In all specifications, we include a vector of demographic controls summarized in the vector X_i . Specifically, we control for 5-year birth cohort, race, state of residence, educational attainment beyond bachelors, survey year, and marital status. Standard errors are clustered by state of residence.

Table 1 shows the estimates of (5). The top panel of the table pools the estimation across all cohorts and explores two distinct dependent variables. In columns (1) to (4), our measure of L_i is log wages, In columns (5) and (6), our dependent variable is a dummy variable equal to 1 if the individual is employed. As with the analysis throughout the paper, the sample in all columns is restricted to only those with at least a bachelor's degree. The sample in columns 1-4 is restricted further to include those individuals with strong attachment to the labor market.

The coefficient on $Female_i$ in Column 1 shows that college-educated women earn 23.1 log points lower wages than their male counterparts after controlling for demographics.¹⁸

¹⁷In Table A2 of the Online Appendix, we report results from an alternate specification where we do not include demographic controls or time fixed effects. In Table A3 of the Online Appendix, we report results from two alternate specifications where we aggregate majors and occupations to broader categories including including dummies for each broad major and occupation category. These exercises yield results that are very similar to those in Table 1.

¹⁸The estimated gender wage gap for college graduates using the pooled cohorts from the 2014-2017 ACS

Table 1: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment

(a) Log Wage and Employment Rate Regressions, Pooled Cohorts

Variable	Log Wages				Employment Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.233 (0.006)	-0.158 (0.004)	-0.143 (0.004)	-0.114 (0.003)	-0.088 (0.003)	-0.083 (0.003)
\bar{Y}_i^m		0.807 (0.015)		0.408 (0.012)		0.045 (0.003)
\bar{Y}_i^o			0.757 (0.011)	0.677 (0.009)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.27	0.36	0.37	0.13	0.13

(b) Log Wage Regressions, Separately By Cohort

Variable	1958-1967 Birth Cohorts			1978-1987 Birth Cohorts		
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.322 (0.008)	-0.198 (0.005)	-0.168 (0.004)	-0.155 (0.005)	-0.093 (0.004)	-0.065 (0.004)
\bar{Y}_i^m			0.411 (0.016)			0.443 (0.010)
\bar{Y}_i^o		0.909 (0.015)	0.823 (0.012)		0.599 (0.008)	0.513 (0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.13	0.32	0.33	0.13	0.25	0.27

Note: Table shows estimates from regression (5). See text for additional details. Sample size for panel A columns 1-4 is 2,270,392. Sample size for panel A columns 5-6 is 3,428,990. Sample size for panel B columns 1-3 is 533,348. Sample size for panel B columns 4-6 is 614,106.

Column 2 includes a control for the individual’s undergraduate major (as measured by \bar{Y}_i^m). Column 3 includes a control for the individual’s current occupation (as measured by \bar{Y}_i^o). Column 4 includes both the controls for undergraduate major and current occupation. Controlling for major choice (but not occupation choice) reduces the gender gap in wages by about one-third (to 15.9 log points). Controlling for occupation choice (but not major choice) reduces the gender gap by 40 percent (to 14.4 log points). The key finding from Table 1 arises by comparing the results between columns 3 and 4. Controlling for undergraduate major in addition to controlling for current occupation reduces the gender wage gap further by an additional 2.8 percentage points. There is information in undergraduate major that helps to explain gender differences in wage above and beyond current occupational choice.¹⁹ Collectively, controlling for occupation and undergraduate major reduces the gender wage gap for college-educated women by half (11.5 log points).

Columns 5 and 6 of Table 1 evaluates the relationship between major choice and the gender gap in extensive margin labor market participation. In the 2014-2017 ACS, highly educated women were 8.6 percentage points less likely to work than men conditional on demographics (column 5). As seen in column 6, controlling for major choice did not alter the estimated gender gap in employment rates for women with a bachelor’s degree. This specification cannot control for occupational choice given that occupation is often not defined for those who are not working.²⁰ While controlling for undergraduate major choice reduces the estimated wage gap between college-educated men and women, major choice is not important for understanding gender differences in employment rates for this group.

The bottom panel of Table 1 displays estimates from (5) when the sample is restricted to include only those born between 1958-1967 (columns 1-3) and only those born between 1978-1987 (columns 4-6).²¹ For this panel, we only show results where the dependent variable is log wages. Columns 1 and 4 show the gender wage gaps conditional on demographics for the two cohorts. Columns 2 and 5 add controls for occupational choice. Columns 3 and 6 add controls for both occupational choice and major choice. The gender gap in wages conditional on demographics fell from -32 log points to -16 log points between the 1958 and 1987 birth cohorts of college graduates. For each cohort separately, controlling for

is nearly identical to the estimated gender wage gap for young college graduates using historical U.S. Census data spanning the same cohorts. See the Online Appendix for additional details.

¹⁹These findings are complementary with the results in Brown and Corcoran (1997) which finds that major choice has predictive power in explaining gender wage gaps independent of occupation choice in cohorts born in the 1940s and 1950s.

²⁰Occupation is recorded for those who are not working only if they were employed at some point in the prior five years.

²¹To increase power in these regressions, we pool together adjacent 5-year cohorts to make 10-year cohorts. The Online Appendix shows similar regressions for all the 10-year cohorts in our data.

occupation and major reduces the gender wage gap by between 50 and 60 percent. This is similar to the pooled cohort results in the top panel of the table. Also, within each cohort, undergraduate major choice has predictive power above and beyond controlling just for occupational choice. Interestingly, major choice has more predictive power and occupational choice has less predictive power for more recent cohorts relative to older cohorts.

The bottom panel of Table 1 also allows us to assess how controlling for occupational choice and undergraduate major alters the time series patterns in the gender wage gaps. By comparing columns 1 and 4, the baseline gender wage gap fell by 18 percentage points for a sample of individuals with a bachelors degree between cohorts born in the 1960s relative to the 1980s. The gender wage gap after controlling for both occupational choice and major choice fell by 11 percentage points. Therefore, the changing gender difference in both occupational and major choice explain roughly 39 percent of the declining gender gap between these two birth cohorts ($1 - 11/18$).

Panel B of Table 1 provides evidence of large convergence in the college gender wage gap across 10-year birth cohorts. In order to shed light on the power of our explanatory variables within cohort, we conduct a wage decomposition exercise. We report the formal results in Appendix Table A5 of the Online Appendix and briefly discuss those results here. Across all 10-year birth cohorts, occupational specialization explains the largest share of the gender wage gap for college graduates ranging from explaining 43.9% of the gap in the oldest cohort (1948-1957) to explaining 36.9% of the gender wage gap in the youngest cohort (1978-1987). Pre-labor market human capital specialization (major choice) is also important in explaining the college gender wage gap ranging from explaining 17.6% of the gender wage gap in the oldest cohort (1948-1957) to 27.9% of the gender wage gap in the youngest cohort (1978-1987). Notably, human capital attainment above and beyond a bachelors degree (such as a graduate degree) explains considerably less of the college gender wage gap. Occupational specialization has become slightly less important over time; it explains about 7 percentage points less of the gender wage gap in the youngest (1978-1987) compared to the oldest (1948-1957) birth cohort. In contrast, college major has become increasingly important in explaining the gender wage gap for college graduates over time; it explains 10.3 percentage points *more* of the gap in the youngest (1978-1987) compared to the oldest (1948-1957) birth cohort. These results suggest that properly accounting for human capital decisions above and beyond schooling attainment and occupational specialization is centrally important in understanding the causes of the gender wage gap among the highly-skilled.

6 Discussion and Conclusion

In this paper, we document five novel facts: (1) Over the past 40 years, men and women have chosen more similar undergraduate majors. (2) Women have historically chosen college majors associated with lower potential wages than men. Although these gaps have narrowed over time, this is still true for women in the youngest birth cohorts. (3) There are gender differences in the mapping of major to occupation. Women systematically choose lower potential wage and lower potential hours-worked occupations than men even conditional on major. (4) There is a modest convergence between the 1950 and 1990 birth cohorts in the gendered mapping of major to occupation. (5) College major choice has strong predictive power in explaining gender wage gaps independent of occupation choice. These trends arise in an era where both labor force participation of women and the fraction of women that have graduated from universities has increased dramatically, fundamentally changing the composition of the educated workforce.

Attempts to measure detailed differences in pre-market specialization by college workers have been limited by data availability. Researchers have long suspected that pre-market specialization (major choice) should in some way pre-determine labor market opportunities particularly with respect to occupational choice. As specialized knowledge is iterative, a biochemical engineer would likely be ill-served by choosing studio art as her primary undergraduate major. If men and women sort into field of study in systematically different patterns, it follows that major choice should have non-trivial implications for the college gender wage gap. The combined results of this paper support this intuition and underscore the importance of further evaluating gender differences in pre-labor market specialization including college major choice.

As a final thought, this work focuses on one aspect of pre-market specialization—major choice. Yet, there are other ways in which men and women make pre-market investments in human capital that pre-determine labor market opportunities. The results in this paper suggest that an investigation into other avenues of pre-market specialization are important, particularly investments that may happen earlier in the human capital chain. Further, understanding the mechanisms for gendered specialization—whether it is driven by preferences or information- is of first-order concern for researchers.

References

- Altonji, J.G., P. Arcidiacono, and A. Maurel**, “Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects,” in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds.*, Vol. 5 of *Handbook of the Economics of Education*, Elsevier, 2016, pp. 305 – 396.
- Altonji, Joseph G and Rebecca M Blank**, “Race and gender in the labor market,” *Handbook of labor economics*, 1999, 3, 3143–3259.
- Altonji, Joseph G., Lisa B. Kahn, and Jamin D. Speer**, “Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs,” *The American Economic Review*, 2014, 104 (5), 387–393.
- Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske**, “New evidence on sex segregation and sex differences in wages from matched employee-employer data,” *Journal of labor Economics*, 2003, 21 (4), 887–922.
- Becker, Gary S, William HJ Hubbard, and Kevin M Murphy**, “Explaining the worldwide boom in higher education of women,” *Journal of Human Capital*, 2010, 4 (3), 203–241.
- Bertrand, Marianne and Kevin F Hallock**, “The gender gap in top corporate jobs,” *ILR Review*, 2001, 55 (1), 3–21.
- Black, Dan, Amelia Haviland, Seth Sanders, and Lowell Taylor**, “Gender Wage Disparities Among the Highly Educated,” *The Journal of Human Resources*, 2008, 43 (3), 630–59.
- Black, Sandra E and Chinhui Juhn**, “The rise of female professionals: Are women responding to skill demand?,” *American Economic Review*, 2000, 90 (2), 450–455.
- Blau, Francine D and Lawrence M Kahn**, “Swimming upstream: Trends in the gender wage differential in the 1980s,” *Journal of labor Economics*, 1997, 15 (1, Part 1), 1–42.
- and –, “Gender differences in pay,” *Journal of Economic perspectives*, 2000, 14 (4), 75–99.
- Blau, Francine D. and Lawrence M. Kahn**, “The Gender Wage Gap: Extent, Trends, and Explanations,” *Journal of Economic Literature*, September 2017, 55 (3), 789–865.
- , **Marianne Ferber, and Anne Winkler**, *The Economics of Women, Men and Work*, Pearson, 2014.
- Blau, Francine D, Patricia Simpson, and Deborah Anderson**, “Continuing progress? Trends in occupational segregation in the United States over the 1970s and 1980s,” *Feminist economics*, 1998, 4 (3), 29–71.

- Brown, Charles and Mary Corcoran**, “Sex-Based Differences in School Content and the Male-Female Wage Gap,” *Journal of Labor Economics*, 1997, 15 (3), 431–465.
- Ceci, Stephen, Donna Ginther, Shulamit Kahn, and Wendy Williams**, “Women in Academic Science: A Changing Landscape,” *Psychological Science in the Public Interest*, 2014, 15 (3), 75–141.
- Charles, Kerwin Kofi and Ming-Ching Luoh**, “Gender differences in completed schooling,” *Review of Economics and statistics*, 2003, 85 (3), 559–577.
- Cortes, Patricia and Jessica Pan**, “Occupation and gender,” *The Oxford Handbook of Women and the Economy*, 2018, p. 425.
- Cortez, Patricia and Jessica Pan**, “Prevalence of Long Hours and Skilled Women’s Occupational Choice,” *IZA Discussion Paper 10225*, 2016.
- Dickson, Lisa**, “Race and Gender Differences in College Major Choice,” *The Annals of the American Academy of Political and Social Science*, 2010, 627 (1), 108–124.
- DiPrete, Thomas A and Claudia Buchmann**, “Gender-specific trends in the value of education and the emerging gender gap in college completion,” *Demography*, 2006, 43 (1), 1–24.
- Duncan, Otis Dudley and Beverly Duncan**, “A Methodological Analysis of Segregation Indexes,” *American Sociological Review*, 1955, 20 (2), 210–217.
- England, Paula and Su Li**, “Desegregation Stalled: The Changing Gender Composition of College Majors, 1971-2002,” *Gender and Society*, 2006, 20 (5), 657–676.
- Goldin, Claudia**, “Understanding the gender gap: An economic history of American women,” *OUP Catalogue*, 1992.
- **and Lawrence F. Katz**, “The Cost of Workplace Flexibility for High-Powered Professionals,” *The Annals of the American Academy of Political and Social Science*, 2011, 638 (1), 45–67.
- , – , **and Ilyana Kuziemko**, “The Homecoming of American College Women: The Reversal of the College Gender Gap,” *Journal of Economic Perspectives*, 2006, 20 (4), 133–156.
- Groshen, Erica L et al.**, *The structure of the female/male wage differential: Is it who you are, what you do, or where you work?*, Vol. 44, Federal Reserve Bank of Cleveland, 1987.
- Hsieh, Chang-Tai, Erik Hurst, Chad Jones, and Pete Klenow**, “The Allocation of Talent and Economic Growth,” *Econometrica*, 2019, *forthcoming*.
- Jacob, Brian A**, “Where the boys aren’t: Non-cognitive skills, returns to school and the gender gap in higher education,” *Economics of Education review*, 2002, 21 (6), 589–598.

- Jacobsen, Joyce P, James Wishart Pearce III, and Joshua L Rosenbloom**, “The Effects of Childbearing on Married Women’s Labor Supply and Earnings.,” *Journal of Human Resources*, 1999, 34 (3).
- Leslie, Larry L., Gregory T. McClure, and Ronald L. Oaxaca**, “Women and Minorities in Science and Engineering: A Life Sequence Analysis,” *The Journal of Higher Education*, 1998, 69 (3), 239–76.
- Loury, Linda Datcher**, “The Gender Earnings Gap among College-Educated Workers,” *Industrial and Labor Relations Review*, 1997, 50 (4), 580–593.
- Macpherson, David A and Barry T Hirsch**, “Wages and gender composition: why do women’s jobs pay less?,” *Journal of labor Economics*, 1995, 13 (3), 426–471.
- Pan, Jessica**, “Gender segregation in occupations: The role of tipping and social interactions,” *Journal of Labor Economics*, 2015, 33 (2), 365–408.
- Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek**, “IPUMS USA: Version 9.0 [dataset],” 2019.
- Turner, Sarah E. and William G. Bowen**, “Choice of Major: The Changing (Unchanging) Gender Gap,” *Industrial and Labor Relations Review*, 1999, 52 (2), 289–313.
- Zafar, Basit**, “College Major Choice and the Gender Gap,” *Journal of Human Resources*, 2013, 48 (3), 545–95.

Online Appendix for “Gender Differences in College Major Choice: Not for Publication”

A1 Data Description

A1.1 Main ACS Samples

Our analysis is conducted using the 2014 to 2017 American Community Survey (ACS). The sample is restricted to include those who are not living in institutional group quarters, were born in one of the 50 U.S. states, have attained at least four years of college completion, and are age 23 to 67. We construct 5-year birth cohorts centered around the reported birth cohort. For example, the 1965 birth cohort includes those born between 1963 and 1967 (inclusive). When data for key demographic variables is missing, the ACS imputes values including age, sex, race, place of birth, educational attainment and undergraduate major. In the 2014 to 2017 ACS, 276,448 (2.2% of the 2014-2017 ACS) respondents have imputed educational attainment information and 196,379 respondents (1.6% of the 2014-2017 ACS) have imputed degree field information. We restrict our sample to include only those with non-imputed age, sex, race, origin, educational attainment and undergraduate major field information. We use inverse probability weighting to correct for non-response. In doing so, we preserve the age, sex race, and state of birth joint distribution. In total, our analysis sample of ACS respondents includes 1,718,330 individuals.

In our analysis, we proxy an hourly wage by dividing reported annual labor income by the reported usual hours the respondent worked in the previous year times the reported number of weeks the respondent worked during the previous year. As the weeks worked variable is an intervalled variable in the 2014-2017 ACS, we assign the midpoint of the category as the number of weeks worked. Nominal wages are converted to real 2018\$. In all analyses including wages, we follow the conventional practices in the literature and restrict the sample to a set of people with well-measured wages: those who are employed civilians (excluding the self-employed) with non-missing annual labor income and strong attachment to the labor market defined as usually working at least 30 hours a week for a minimum of 27 weeks in the previous year. In calculating the potential wage indices by occupation and undergraduate major, we restrict the sample to white men in their peak wage years (ages 45 to 55) with well-measured wages. All analyses use log wages.

We use the variables `degfield`, `degfield2`, `degfieldd`, and `degfield2d` in IPUMS to identify both broad and detailed majors. IPUMS provides 176 detailed major codes for the 2014 to 2017 samples. In our analysis, we aggregate to 134 detailed major categories by subsuming very small major categories into larger categories. For example, we combined General Agriculture, Soil Science, and Miscellaneous Agriculture into one detailed agricultural major. Similarly, we combine Mathematics, Actuarial Science, and Mathematics and Computer Science into one detailed mathematics major. Our main analysis uses these detailed major categories. There are 29 broad major categories in our analysis. We use the broad major categories to describe trends in Figure 1 and describe major-to-occupation mappings in Figure 3.

Approximately 11 percent of the observations in our sample have dual majors. Our analysis requires a maximum of one major for each unit of observation. Thus, we assign a primary major to each person based on the maximum median potential wage in the two majors (based on white men aged 43-57 as described above). This assignment process relies on the assumption that agents will present their highest-wage major as their primary major in the labor market.

Figures A1 to A5 include a full listing of our detailed and broad major codes. Our data replication kit provides the code for our combination of majors.

We use a balanced panel of detailed occupation codes based on the 1990 Occupation codes and following the cross-walking strategy outlined by David Dorn which constructs a panel of 330 occupation codes.²² In our analysis, we aggregate to 251 detailed occupation codes by subsuming very small occupation categories into larger categories. For people who are employed, the ACS reports occupation based on primary occupation. For people who are unemployed, the ACS reports occupation based on their most recent primary occupation in the last five years.

A2 Gender Differences in Wages and Employment Rates, Historical U.S. Censuses

As way of background, we measure time series trends in gender wage and employment gaps for young individuals using cross-sectional data from historical U.S. Censuses. We focus on cross-sections from the 1960, 1970, 1980, 1990, and 2000 U.S. Censuses as well as the 2010-2012 American Community Surveys (pooled). Within these data sets, we define the wages of those with at least a bachelor's degree similarly to our wage computation within the 2014-2017 ACS. Employment rates are measured as the individual currently working at least 30 hours per week (full-time).

The solid line in Figure A6 shows the difference in log wages between young working women and young working men with a bachelor's degree at decadal intervals using the historical Census and recent ACS data. We define "young" as individuals between the ages of 25 and 34. Given the decadal frequency and 10-year age range, each point on the lines in Figure A6 represents a different cohort of individuals. Young women with at least a bachelor's degree entering the labor market in 1960 (born between 1926 and 1935) earned average hourly wages that were 31 log points lower than their male counterparts. The gender wage gap for those with a bachelor's degree narrowed to 13 log points for the young cohort entering the labor market in 2010 (born between 1976 and 1985). The wage gap declined monotonically by a total of 18 log points for the cohorts entering the labor market between 1960 and 2010 with most of the decline occurring for cohorts entering the labor market between 1980 and 1990 (those born between 1956 and 1965). The results in Figure A6 show that substantial gender wage convergence occurred within the highly selected sample of those who attained a bachelor's degree.

The dashed line of Figure A6 shows gender convergence in full time employment rates among our sample of college graduates. We define full time employment as those individuals

²²See <https://www.ddorn.net/data.htm>.

who report currently working at least 30 hours per week. Throughout all years, roughly 90 percent of college-educated young men reported working full time. For women, only about 35 percent of 25-34 year old female college graduates worked full time in 1960 – a gender employment rate gap of 35 percentage points. By 2010, roughly 80 percent of young educated women worked full time. For individuals with a bachelor’s degree, there was a strong increase in women’s wages and employment propensities relative to their male counterparts during the last half century. One goal of the paper is to assess how changing differences in undergraduate major choice is associated with changing gender wage and employment gaps across cohorts.

A3 Construction of Potential Wage Indices

In Section 3, we define our potential wage index as:

$$I_c^{P,M} = \frac{\sum_{m=1}^M s_{female,c}^m \bar{Y}_{male}^m}{\sum_{m=1}^M s_{male,c}^m \bar{Y}_{male}^m} - 1 \quad (6)$$

In practice, we compute this index by running the following regression locally within 5-year birth cohort c :

$$\bar{Y}_{male_i}^m = \alpha + \beta Female_i + \Gamma X_i + \epsilon_i \quad (7)$$

where $Female_i$ is a dummy variable indicating whether the respondent i has self-reported as female and X_i is a vector of demographic characteristics: race, state of birth, masters attainment, doctorate attainment, and marital status. Our potential wage index, $I_c^{P,M}$, is defined as β from each local regression by 5-year birth cohort and therefore, measures the differential “potential” wage of women of cohort c given that the female distribution of major choice in a given cohort may differ from males in their cohort. The units of this index are differential potential log wage based on major choice. In computing the comparable index for occupation, we substitute detailed occupation code, o , for detailed major code, m .

A4 Robustness: Key Results

A4.1 Robustness Figure 2

In Figure 2 of the main text, we restrict the sample on which our gender similarity indices are built to all individuals with reported majors (for $I_c^{DD,M}$ and $I_c^{P,M}$) or to all individuals with reported occupations (for $I_c^{DD,O}$ and $I_c^{P,O}$). Some individuals with reported majors are not working during the 2014-2017 period. Likewise, some individuals with reported occupations are not currently working (given the ACS asks occupations for people who are currently not working but may have worked at some point in the prior 5 years). To see if including those who are currently not working bias our indices, we perform a robustness exercise by creating the respective indices restricted to a sample of individuals with strong attachment to the labor market as defined throughout our analysis (civilians who are not self-employed and report working for at least 30 hours a week for at least 27 weeks in the previous year). The results of this exercise are shown in Appendix Figure A7. The results in Appendix Figure

A7 are nearly identical to the results in Figure 2 of the main text. This suggests that our results are insensitive to whether we include individuals with strong attachment to the labor market or all individuals when describing patterns of gender sorting in major (occupation) choice.

Another potential issue with the results in Figure 2 stems from the fact that the ACS only asks undergraduate major in recent years (from 2009 to 2017). When we compare patterns for different birth cohorts, we risk confounding cohort and age effects. It is unlikely that this is problematic for our results about the convergence of undergraduate majors given that major is likely fixed over an individual’s life cycle. Occupations are not fixed over the lifecycle, so this presents a potential problem with respect to how we have described occupational segregation by gender in Figure 2. To address this and separate age and cohort effects on occupational segregation by gender, we use data from the 1980, 1990, and 2000 U.S. Censuses along with multiple waves of the American Community Survey to measure $I_c^{DD,O}$ for different birth cohorts at a constant age. The results are shown in Appendix Table A1. As with the results in the main text, birth cohort refers to 5-year birth cohorts centered around the birth year listed. Similarly, age refers to 5-year age ranges centered on the age listed. As seen in Appendix Table A1, age effects are not substantively biasing the main results shown in Figure 2 of the main text. Within each age range, we see large convergence in the occupation similarity between men and women across birth cohorts.

A4.2 Robustness Table 1, Panel A

Appendix Tables A2 and A3 show a series of robustness checks on the results shown in Panel A of Table 1 of the main text. Appendix Table A2 shows our key regression results without including our vector of demographic and time controls. Focusing on column 1 of the Table, the raw gender gap in wages among individuals without a bachelor’s degree for our pooled sample was 26.8 log points. Including demographic controls as in column 1 of Table 1 of the main text, the gender gap only fell to 23.3 log points. The demographic controls only explain a small fraction of the gender wage gap among college graduates.

Appendix Table A3 shows the robustness results for Panel (a) of Figure 1 of the main text to the alternate classification of majors and occupations. In our base specification in the main text, we used detailed occupation and major codes when defining the potential wage variables Y_i^m and Y_i^o . In the top panel of Appendix Table A3, we use the broad occupation and major codes to define Y_i^m and Y_i^o . In the bottom panel, we omit Y_i^m and Y_i^o altogether from the regression and instead include a vector of dummy variables for each broad major (occupation). The results of these alternate specifications are nearly identical to the results shown in Table 1 of the main text. This suggests that most of the variation in explaining gender wage gaps arises from differences across (as opposed to within) the broad major and occupation controls.

A4.3 Table 1, Panel B

In Panel B of Table 1, we show the importance of occupational choice and undergraduate major in explaining gender wage gaps for the 1958-1967 and the 1978-1987 birth cohorts. Appendix Table A4 shows the same results for the 1948-1957 and the 1968-1977 birth cohorts.

A5 Hours Differences Across Occupations

As discussed in the main text, there is a large literature highlighting the fact that women are in occupations with lower annual hours worked relative to men. In this section, we explore these results in the context of our methodology. To guide our empirical work, we define the following two variables: \bar{H}_{male}^m and \bar{H}_{male}^o . \bar{H}_{male}^m is defined as the median log annual hours worked for native-born white men between the ages of 43 and 57 who graduated with major m (regardless of subsequent occupation in which they worked). This is the potential hours associated with a given major based on older male hours. \bar{H}_{male}^o is defined as the median log annual hours worked for native-born, white men between the ages of 43 and 57 who currently work in occupation o (regardless of undergraduate major). This is the potential hours associated with a given occupation based on older male hours. We refer to these variables as our potential annual hours worked indices. Majors (occupations) where men work more on average will have higher levels of \bar{H}_{male}^m (\bar{H}_{male}^o).

How similar are men and women with respect to their occupational choices based on potential annual hours worked? Appendix Figure A8 displays $I_c^{H,M}$ and $I_c^{H,O}$ for different cohorts. $I_c^{H,M}$ is our potential hours index based on male annual hours worked in different majors and is defined as $I_c^{H,M} = \frac{\sum_{m=1}^M s_{female,c}^m \bar{H}_{male}^m}{\sum_{m=1}^M s_{male,c}^m \bar{H}_{male}^m} - 1$. Like our potential wage indices in the main text, the only reason $I_c^{H,M}$ only differs from 0 if men and women inhabit different majors. Likewise, $I_c^{H,O}$ is our potential hours index based on male annual hours worked in different occupations and is defined as $I_c^{H,O} = \frac{\sum_{o=1}^O s_{female,c}^o \bar{H}_{male}^o}{\sum_{o=1}^O s_{male,c}^o \bar{H}_{male}^o} - 1$.

As seen from Appendix Figure A8, women choose majors and occupations associated with lower potential annual hours worked. The major and occupational choice of women have converged to that of men over time in a way that implies women and men are choosing majors and occupations with more similar hours requirements. For the most recent cohorts, women are choosing both majors and occupations where potential annual hours worked are roughly 2% lower than men. Consistent with the literature, we find that college-educated women are choosing occupations with lower annual hours worked. We contribute to the hours literature by introducing the fact that college-educated women are choosing undergraduate majors associated with lower annual hours worked. We also show that the gender similarity of occupations and majors based on potential hours has been converging over time.

Appendix Figure A9 shows gender differences in the mapping of majors to occupations where we measure occupations in units of potential annual hours worked \bar{H}_{male}^o . Appendix Figure A9 is otherwise analogous to Figure 3 of the main text. To measure gender differences in occupational choice in hours units conditional on undergraduate major, we define $I_c^{H,O}|m$ which just recalculates $I_c^{H,O}$ (as defined above) restricting the sample to those individuals that chose major m . Consider individuals who choose to major in engineering (Panel A, solid line). Women from the 1950 birth cohort who majored in engineering subsequently chose occupations that had potential hours worked that were 2% lower than otherwise similar males. That gap disappeared for women who majored in engineering after the 1975 birth cohort. For all majors, the gender gap in potential annual hours worked of occupational choice conditional on major has fallen over time. Women are now choosing occupations that are more similar in hours worked to men, conditional on occupational choice.

Appendix Figure A10 summarizes the mapping of majors to occupations where we measure occupations in potential hours space. This figure is otherwise analogous to Figure 4 in the main text. Women from the 1975 birth cohort are in occupations – conditional on major choice – that have annual hours worked that are three percent lower than comparable men. As a reminder, occupational potential wage differences, conditional on major choice, were about 9 percent for this cohort. Some of the reason that women may be choosing occupations with lower wages is that those occupations also have lower annual hours worked.

A6 Wage Gap Decompositions

In this section, we discuss the findings of a decomposition exercise where we assess the contribution of the independent variables in our main estimation to the college gender wage gap. As with the estimations in Table 1 and Table A4, the sample is restricted to include those with strong attachment to the labor market. We begin by estimating locally within birth cohort log wage equations for men only where race, state of residence, and marital status are categorical variables. As with all other specifications, the independent variable for *Major* is the potential log wage from major choice, \bar{Y}_i^m , and the independent variable for *Occupation* is the potential log wage from occupation choice, \bar{Y}_i^o . Entries in the "Log Points" column are the within-cohort *male – female* differences in the mean of the corresponding variable multiplied by the within-cohort *male* log wage coefficients of the corresponding variable. Entries in the "% Explained" column are the "Log Points" entries divided by the within-cohort *Total Raw Gap*.

In our model specification, occupational specialization plays the largest role in explaining the college gender wage gap. This is true for all 10-year birth cohorts. In the oldest birth cohort (1948-1957), occupation explains 43.9% of the gender wage gap. For the youngest birth cohort (1978-1987), the importance of occupation declines by 7 percentage points explaining 36.9% of the gender wage gap.

The results in Table 1 and Table A4 show that major choice and occupation choice are *independently* related to the college gender wage gap. This finding is a contribution to the literature on the college gender wage gap and the role of pre-market specialization. In our decomposition exercise, we formally show that pre-labor market human capital specialization (major choice) has non-trivial importance in explaining the college gender wage gap. For the oldest birth cohort (1948-1957), major choice explains 17.6% of the college gender wage gap. For the youngest birth cohort (1978-1987), major choice explains 27.9% of the college gender wage gap. While much of the existing literature has focused on the role of human capital attainment with respect to the gender wage gap, our decomposition shows that human capital attainment *above and beyond a bachelors degree* (such as a graduate degree) explains considerably less of the college gender wage gap than both pre-market and market human capital specialization.

Finally, in thinking about the time series patterns, two findings are of particular interest. First, occupational specialization has become less important between the 1948-1957 and 1958-1967 birth cohorts and then mostly stabilized. For the 1948-1957 birth cohort, occupation explained 43.9% of the college gender wage gap. This fell to 38.0% for the 1958-1967 birth and was 37.5% and 36.9% for the 1968-1977 and 1978-1987 birth cohorts respectively.

Second, college major has become increasingly important in explaining the gender wage gap for college graduates over time. It explained 10.3 percentage points *more* of the gap in the youngest (1978-1987) compared to the oldest (1948-1957) birth cohort.

Figure A1: List of Detailed and Broad Majors

<u>Detailed Major</u>	<u>Broad Major</u>
General Agriculture, Soil Science, Misc. Agriculture	Agriculture
Agriculture Production and Management	Agriculture
Animal Sciences	Agriculture
Food Science	Agriculture
Plant Science and Agronomy	Agriculture
Environmental Science	Environment and Natural Resources
Forestry	Environment and Natural Resources
Natural Resources Management	Environment and Natural Resources
Architecture	Architecture
Area, Ethnic, and Civilization Studies	Area, Ethnic, and Civilization Studies
Communications	Communications
Journalism	Communications
Mass Media	Communications
Advertising and Public Relations	Communications
Communication Technologies	Engineering
Computer and Information Systems	Computer and Information Systems
Computer Programming and Data Processing	Computer and Information Systems
Computer Science	Computer and Information Systems
Information Sciences	Computer and Information Systems
Computer Information Management and Security	Computer and Information Systems
Computer Networking and Telecommunications	Computer and Information Systems
Cosmetology Services and Culinary Arts	Cosmetology and Physical Fitness
General Education, School Counseling, Educational Administration and Supervision	Education Administration and Teaching
Elementary Education	Education Administration and Teaching
Mathematics Teacher Education	Education Administration and Teaching
Physical and Health Education Teaching	Education Administration and Teaching
Early Childhood Education	Education Administration and Teaching
Science and Computer Teacher Education	Education Administration and Teaching
Secondary Teacher Education	Education Administration and Teaching
Special Needs Education	Education Administration and Teaching
Social Science or History Teacher Education	Education Administration and Teaching
Teacher Education: Multiple Levels	Education Administration and Teaching
Language and Drama Education	Education Administration and Teaching
Art and Music Education	Education Administration and Teaching
Miscellaneous Education	Education Administration and Teaching

Figure A2: List of Detailed and Broad Majors (continued)

<u>Detailed Major</u>	<u>Broad Major</u>
Engineering: General, Military Technologies, Metallurgical, Biomedical, Geological and Geophysical, Mining and Mineral, Naval Architecture and Marine, Nuclear, Petroleum	Engineering
Aerospace Engineering	Engineering
Biological Engineering	Engineering
Chemical Engineering	Engineering
Civil and Architectural Engineering	Engineering
Computer Engineering	Engineering
Electrical Engineering, Electrical Engineering Technology, Electrical and Mechanic Repairs and Technologies	Engineering
Engineering Mechanics, Physics, and Science	Engineering
Environmental Engineering	Engineering
Industrial and Manufacturing Engineering, Precision Production and Industrial Arts	Engineering
Mechanical Engineering	Engineering
Miscellaneous Engineering	Engineering
Engineering Technologies	Engineering
Engineering and Industrial Management	Engineering
Industrial Production Technologies	Engineering
Mechanical Engineering Related Technologies	Engineering
Miscellaneous Engineering Technologies	Engineering
Linguistics, and Comparative Language and Literature	Linguistics and Foreign Languages
French, German, Latin and Other Common Foreign Languages	Linguistics and Foreign Languages
Other Foreign Languages	Linguistics and Foreign Languages
Family and Consumer Sciences	Family and Consumer Sciences
Pre-Law and Legal Studies, Court Reporting	Law
English Language and Literature	English and Literature
Composition and Speech	English and Literature
Liberal Arts	Liberal Arts and Humanities
Humanities	Liberal Arts and Humanities
Library Science	Education

Figure A3: List of Detailed and Broad Majors (continued)

<u>Detailed Major</u>	<u>Broad Major</u>
Biology, Misc. Biology, Pharmacology, Botany, Neuroscience, Genetics	Biology and Life Sciences
Biochemical Sciences	Biology and Life Sciences
Molecular Biology	Biology and Life Sciences
Ecology	Biology and Life Sciences
Microbiology	Biology and Life Sciences
Physiology	Biology and Life Sciences
Zoology	Biology and Life Sciences
Mathematics, Actuarial Science, Mathematics and Computer Science	Math and Statistics
Applied Mathematics	Math and Statistics
Statistics and Decision Science	Math and Statistics
Interdisciplinary, Multi-Disciplinary, Intercultural and International Studies	Multi-Disciplinary Studies (General)
Nutrition Sciences	Multi-Disciplinary Studies (General)
Physical Fitness, Parks, Recreation, and Leisure	Cosmetology and Physical Fitness
Philosophy and Religious Studies	Philosophy and Theology
Theology and Religious Vocations	Philosophy and Theology
Physical Sciences, Astronomy and Astrophysics, Geosciences, Nuclear, Industrial Radiology, and Biological Technologies	Physical Sciences
Atmospheric Sciences and Meteorology	Physical Sciences
Chemistry	Physical Sciences
Geology and Earth Science	Physical Sciences
Oceanography	Physical Sciences
Physics	Physical Sciences
Materials Science and Materials Engineering	Engineering
Multi-disciplinary or General Science	Physical Sciences
Psychology, Cognitive Science and Biopsychology, Social Psychology	Psychology
Educational Psychology	Psychology
Clinical Psychology	Psychology
Counseling Psychology	Psychology
Industrial and Organizational Psychology	Psychology
Miscellaneous Psychology	Psychology

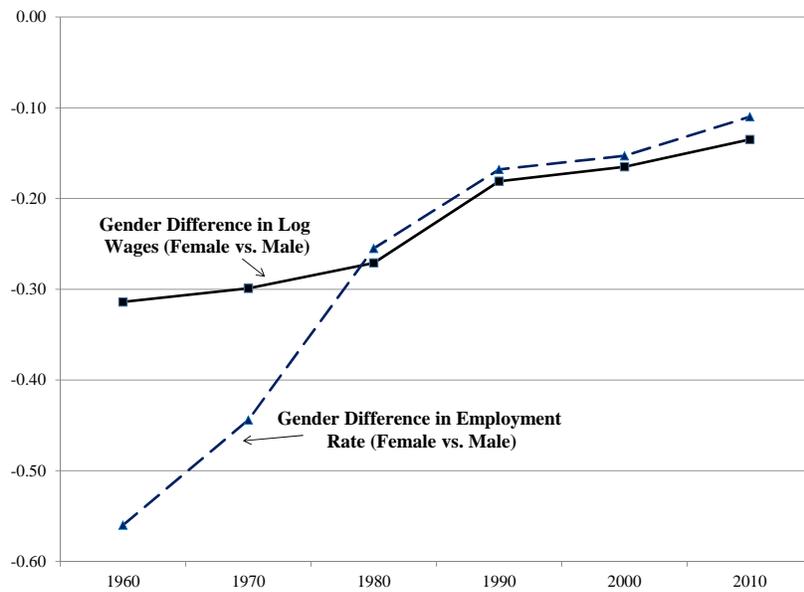
Figure A4: List of Detailed and Broad Majors (continued)

<u>Detailed Major</u>	<u>Broad Major</u>
Criminal Justice and Fire Protection	Criminal Justice and Fire Protection
Public Administration	Public Affairs, Policy, and Social Work
Public Policy	Public Affairs, Policy, and Social Work
Human Services and Community Organization	Public Affairs, Policy, and Social Work
Social Work	Public Affairs, Policy, and Social Work
General Social Sciences	Social Sciences
Economics, Agricultural Economics, Business Economics	Social Sciences
Anthropology and Archeology	Social Sciences
Criminology	Social Sciences
Geography	Social Sciences
International Relations	Social Sciences
Political Science and Government	Social Sciences
Sociology	Social Sciences
Miscellaneous Social Sciences	Social Sciences
Construction Services	Construction Services
Transportation Sciences and Technologies	Construction Services
Fine Arts, Commercial Art and Graphic Design, Film, Video and Photographic Arts, Studio Arts, Miscellaneous Fine Arts	Fine Arts
Drama and Theater Arts, Music, Visual and Performing Arts	Fine Arts
Art History and Criticism	Fine Arts
General Medical and Health Services	Nursing, Medical and Health Sciences
Communication Disorders Sciences and Services	Nursing, Medical and Health Sciences
Health and Medical Administrative Services	Nursing, Medical and Health Sciences
Medical Assisting Services	Nursing, Medical and Health Sciences
Medical Technologies Technicians	Nursing, Medical and Health Sciences
Health and Medical Preparatory Programs	Nursing, Medical and Health Sciences
Nursing	Nursing, Medical and Health Sciences
Pharmacy, Pharmaceutical Sciences, and Administration	Nursing, Medical and Health Sciences
Treatment Therapy Professions	Nursing, Medical and Health Sciences
Community and Public Health	Nursing, Medical and Health Sciences
Miscellaneous Health Medical Professions	Nursing, Medical and Health Sciences

Figure A5: List of Detailed and Broad Majors (continued)

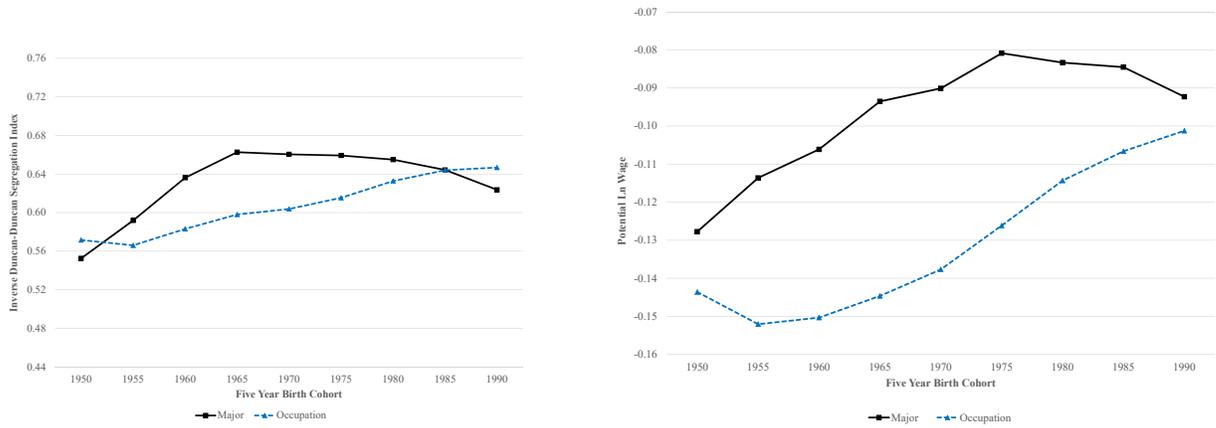
<u>Detailed Major</u>	<u>Broad Major</u>
General Business	Business
Accounting	Business
Business Management and Administration	Business
Operations, Logistics and E-Commerce	Business
Marketing and Marketing Research	Business
Finance	Business
Human Resources and Personnel Management	Business
International Business	Business
Hospitality Management	Business
Management Information Systems and Statistics	Business
Miscellaneous Business and Medical Administration	Business
History	History
United States History	History

Figure A6: Gender Differences in Log Wages and Employment Rates, Individuals Aged 25-34 with Bachelor's Degree



Notes: Figure shows the gender gap in log wages (solid line) and the gender gap in employment rates (dashed line) for individuals aged 25-34 with a bachelor's degree in different Census years. The 2010 data refers to a pooled sample of ACS respondents from sample years 2010, 2011, and 2012. Individual wages are self-reported annual labor income divided by self-reported annual hours worked. The employment rate is the fraction of respondents who report working full-time. Differences are measured as female data minus male data. See text for additional details.

Figure A7: Gender Similarity in Major Choice and Occupation by Cohort, Strongly Attached Sample

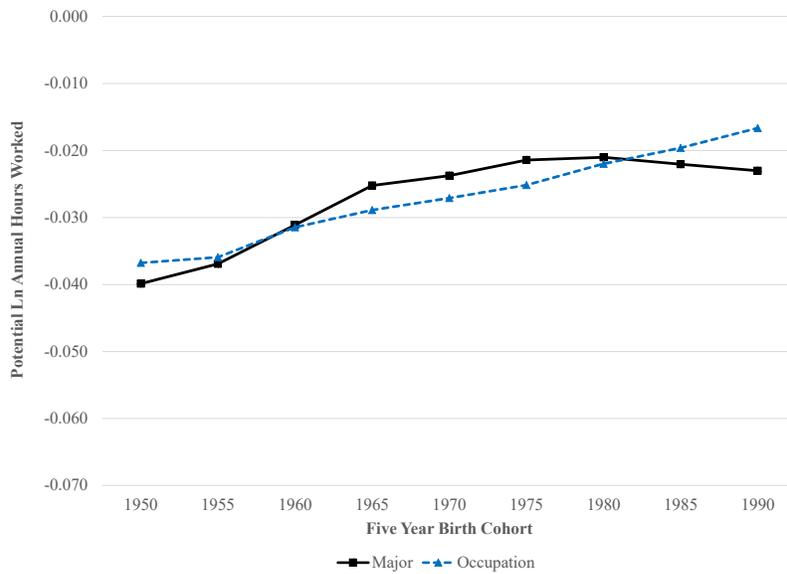


PANEL A: SEGREGATION INDEX INDEX

PANEL B: POTENTIAL WAGE INDEX INDEX

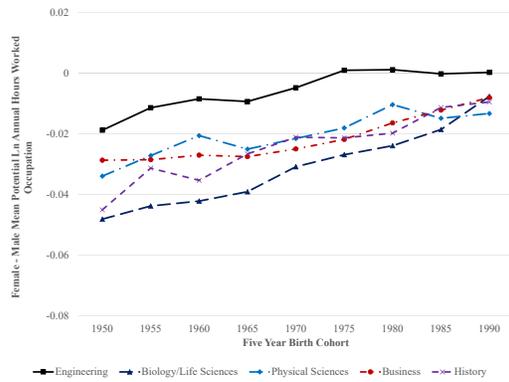
Notes: Figure plots the inverse segregation index (left panel) and potential wage index (right panel) for different cohorts conditioning on strong attachment to the labor market. The solid line in each panel show the indices for major choice. The dashed line in each panel show the indices for occupation. Data from the 2014-2017 ACS and is restricted to those with at least a bachelor's degree. See text for additional details.

Figure A8: Gender Similarity in Major and Occupation by Cohort, Measuring Occupations by Potential Hours Worked

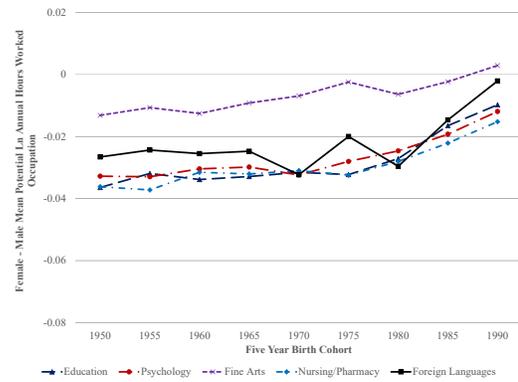


Notes: Figure plots the gender similarity index for different cohorts based on potential hours worked. The solid line in each panel shows the indices based on major choice. The dashed line shows the indices based on subsequent occupation. Data from the 2014-2017 ACS and is restricted to those with at least a bachelor's degree. See text for additional details.

Figure A9: Within-Major Gender Differences in Potential Hours by Occupation, by Gender and Cohort



PANEL A: MALE DOMINATED MAJORS



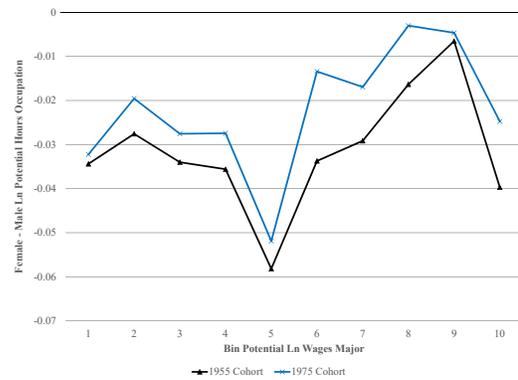
PANEL B: FEMALE DOMINATED MAJORS

Notes: These figures show the trends in $I_c^{H,O|m}$ conditional on having graduated with major m . Panel A are male-dominated majors. Panel B are female-dominated majors. As with the left panel of Figure 2, potential wage in an occupation, \bar{H}_{male}^o , is computed using only annual hours worked of native white males 43-57 who are working full time in the 2014-2017 ACS.

Figure A10: Mapping of Potential Wages by Major to Potential Hours by Occupation, by Gender and Cohort



PANEL A: LEVELS



PANEL B: DIFFERENCES

Notes: These figures show the mapping between major and occupation. On the x-axes, we have binned majors based on \bar{Y}_{male}^m , the log wage deciles of native, white men age 43 to 57. On the y-axis in Panel A, we report $I_c^{H,O|d,m}$, the mean log potential occupational hours worked within these deciles described separately by gender and cohort. In Panel B, the y-axis reports female - male differences in $I_c^{H,O|d,m}$ for two of the cohorts.

Table A1: Robustness of Trends in Inverse Duncan-Duncan Index for Gender Occupational Similarity, Controlling for Age Effects

Cohort	30	35	40	45	50	55
1925						0.358
1930					0.368	
1935				0.380		0.439
1940			0.405		0.475	
1945		0.435		0.499		0.517
1950	0.496		0.503		0.517	
1955		0.522		0.520		0.551
1960	0.609		0.555		0.559	0.576
1965		0.592		0.565	0.583	
1970	0.621		0.581	0.582		
1975		0.605	0.591			
1980	0.625	0.607				
1985	0.632					

Note: This table computes the inverse Duncan-Duncan index for gender similarity in occupational sorting ($I_c^{DD,O}$) for different birth cohorts and age ranges. See main text for construction of the index. Cohorts are five year birth cohorts centered around the birth cohort listed. Age are five year age ranges centered around the age listed. Data come from the 1980, 1990, and 2000 U.S. Censuses as well as various years of the American Community Survey.

Table A2: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment

(a) Log Wage Regressions, No Controls

Variable	Log Wages				Employment Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.268 (0.007)	-0.189 (0.006)	-0.160 (0.005)	-0.135 (0.004)	-0.075 (0.003)	-0.066 (0.003)
\bar{Y}_i^m		0.827 (0.017)		0.355 (0.014)		0.091 (0.005)
\bar{Y}_i^o			0.862 (0.011)	0.793 (0.009)		
Controls	No	No	No	No	No	No
R^2	0.04	0.10	0.23	0.23	0.01	0.01

Note: This table is a robustness check on panel (a) of Table 1 with no demographic or time controls. Sample size for columns 1-4 is 2,270,392. Sample size for columns 5-6 is 3,428,990.

Table A3: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment, Alternative Specifications Using Broad Major and Occupations

(a) Log Wage and Employment Rate Regressions with Broad Major and Occupation Potential Wage Indices, Pooled Cohorts

Variable	Log Wages				Employment Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.232 (0.006)	-0.172 (0.004)	-0.167 (0.004)	-0.141 (0.004)	-0.088 (0.003)	-0.083 (0.003)
\bar{Y}_i^m		0.832 (0.022)		0.461 (0.015)		0.061 (0.004)
\bar{Y}_i^o			0.750 (0.013)	0.668 (0.010)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.26	0.33	0.34	0.13	0.13

(b) Log Wage and Employment Rate Regressions with Flexible, Broad Major and Occupation Dummies, Pooled Cohorts

Variable	Log Wages				Employment Rate	
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.232 (0.006)	-0.169 (0.005)	-0.168 (0.004)	-0.143 (0.004)	-0.088 (0.003)	-0.083 (0.003)
Major dummies	No	Yes	No	Yes	No	Yes
Occupation dummies	No	No	Yes	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.27	0.34	0.35	0.13	0.13

Note: This table is a robustness check on the main results in Panel (a) of Table 1 using two alternate ways to control for occupation and major choice. In Panel (a) of this table, we include as independent variables measures of potential wages determined by the broad major and occupation choice instead of detailed major and occupation choice. In panel (b), we include as independent variables vectors of broad major dummies and occupation dummies instead of our potential wage controls. Sample size for panel A columns 1-4 is 2,256,630. Sample size for panel A columns 5-6 is 3,428,990. Sample size for panel B columns 1-4 is 2,256,630. Sample size for panel B columns 5-6 is 3,428,990.

Table A4: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment

(a) Log Wage Regressions, Older Cohorts

Variable	1948-1957 Birth Cohorts			1958-1967 Birth Cohorts		
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.291 (0.009)	-0.163 (0.005)	-0.130 (0.005)	-0.322 (0.008)	-0.198 (0.005)	-0.168 (0.004)
\bar{Y}_i^m			0.366 (0.018)			0.411 (0.016)
\bar{Y}_i^o		0.886 (0.016)	0.819 (0.013)		0.909 (0.015)	0.823 (0.012)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.12	0.30	0.31	0.13	0.32	0.33

(b) Log Wage Regressions, Younger Cohorts

Variable	1968-1977 Birth Cohorts			1978-1987 Birth Cohorts		
	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.271 (0.008)	-0.169 (0.005)	-0.144 (0.005)	-0.155 (0.005)	-0.093 (0.004)	-0.065 (0.004)
\bar{Y}_i^m			0.410 (0.015)			0.443 (0.010)
\bar{Y}_i^o		0.850 (0.014)	0.766 (0.011)		0.599 (0.008)	0.513 (0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.14	0.31	0.33	0.13	0.25	0.27

Note: The specifications in this table are the same as the specifications shown in Panel (b) of Table 1. Columns 4-6 from this table are exactly the same as the results in Panel (b) of Table 1. The new results in this table are in columns 1-3 of both panels which show the results for alternate birth cohorts. Sample size for panel (a) columns 1-3 is 331,678. Sample size for panel (a) columns 4-6 is 533,348. Sample size for panel (b) columns 1-3 is 543,452. Sample size for panel (b) columns 4-6 is 614,106.

Table A5: Wage Decompositions: Explanatory Variables

(a) Older Cohorts

Variable	1948-1957 Birth Cohort		1958-1967 Birth Cohort	
	(Log Points)	(% Explained)	(Log Points)	(% Explained)
<i>Race</i>	0.0008	0.25%	0.0011	0.32%
<i>State</i>	0.0004	0.14%	0.0002	0.06%
<i>Marital Status</i>	0.0287	9.19%	0.0164	4.74%
<i>Masters</i>	-0.0048	-1.55%	-0.0034	-0.99%
<i>Doctorate</i>	0.0056	1.80%	0.0022	0.63%
<i>Major</i>	0.0549	17.57%	0.0547	15.81%
<i>Occupation</i>	0.1371	43.86%	0.1314	37.97%
<i>Year</i>	-0.0006	-0.19%	-0.0001	-0.04%
<i>Explained</i>	0.2221	71.07%	0.2025	58.51%
<i>Unexplained</i>	0.0904	28.93%	0.1436	41.49%
<i>Total Raw Gap</i>	0.31		0.35	

(b) Younger Cohorts

Variable	1968-1977 Birth Cohort		1978-1987 Birth Cohort	
	(Log Points)	(% Explained)	(Log Points)	(% Explained)
<i>Race</i>	0.0004	0.13%	0.0001	0.04%
<i>State</i>	0.0001	0.03%	0.0001	0.05%
<i>Marital Status</i>	0.0146	5.03%	0.0002	0.11%
<i>Masters</i>	-0.0080	-2.76%	-0.0064	-4.09%
<i>Doctorate</i>	0.0007	0.25%	0.0000	0.03%
<i>Major</i>	0.0444	15.27%	0.0438	27.85%
<i>Occupation</i>	0.1090	37.53%	0.0579	36.86%
<i>Year</i>	-0.0003	-0.09%	0.0001	0.03%
<i>Explained</i>	0.1609	55.39%	0.0956	60.87%
<i>Unexplained</i>	0.1296	44.61%	-0.0615	39.13%
<i>Total Gap</i>	0.29		0.16	

Note: Sample restrictions and cohorts consistent with Table A4. In these estimations, race, state of residence, and marital status are categorical variables instead of flexible dummies. This does not affect our main results and is only for ease in decomposition and display. As with all other specifications, the independent variable for *Major* is \bar{Y}_i^m , and the independent variable for *Occupation* is \bar{Y}_i^o . Entries in the "Log Points" column are the within-cohort *male* – *female* differences in the mean of the corresponding variable multiplied by the within-cohort *male* log wage coefficients of the corresponding variable. Entries in the "% Explained" column are the "Log Points" entries divided by the within-cohort Total Raw Gap. The Total Raw Gap differs from the $Female_i$ in Column (1) of Table A4 in that it is the raw gender wage gap with no controls and the gender wage gap displayed as the coefficient for $Female_i$ in Column (1) of Table A4 includes demographic controls.