NBER WORKING PAPER SERIES

A CROSS-COHORT ANALYSIS OF HUMAN CAPITAL SPECIALIZATION AND THE COLLEGE GENDER WAGE GAP

Carolyn Sloane
Erik Hurst
Dan Black

Working Paper 26348
http://www.nber.org/papers/w26348

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2019, Revised February 2020

We thank seminar participants at the University of Chicago for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap
Carolyn Sloane, Erik Hurst, and Dan Black
NBER Working Paper No. 26348
October 2019, Revised February 2020
JEL No. J16,J24

ABSTRACT

In this paper, we exploit new data to assess gender differences in pre-labor market specialization among the college educated and highlight how those differences have evolved over time. We highlight new results pertaining to gender differences in the mapping between undergraduate major and subsequent occupational sorting. To perform our analysis, we introduce new indices in potential wage space that measure gender differences in major choice and separately the subsequent occupational sorting conditional on major choice. We highlight that women both choose majors with lower potential earnings (based on male wages associated with those majors) and that they then subsequently sort into occupations with lower potential earnings given their major choice. We highlight that these differences have narrowed over time but recent cohorts of women still choose majors and occupations with lower potential earnings. Differences in undergraduate major choice explains a substantive portion of gender wage gaps for the college educated above and beyond simply controlling for occupation. Collectively, our results highlight the importance of understanding gender differences in pre-labor market human capital specialization and the mapping between college major and occupational sorting when studying the evolution of gender differences in labor market outcomes over time.

Carolyn Sloane
Department of Economics
University of California, Riverside
Riverside, CA 92521
c.m.sloane@gmail.com

Erik Hurst
Booth School of Business
University of Chicago
Harper Center
Chicago, IL 60637
and NBER
erik.hurst@chicagobooth.edu

Dan Black
1155 E. 60th Street
Harris School
University of Chicago
Chicago, IL 606037
danblack@uchicago.edu

A data appendix is available at http://www.nber.org/data-appendix/w26348

1 Introduction

In the last several decades, the gender composition of the college-educated workforce in the U.S. has shifted dramatically due to the dual forces of accelerated female labor force participation and growth in the relative supply of college-educated women.[1] At the same time, the specialized skills of college-educated men and women have been evolving. Post-schooling labor market specialization (e.g. occupational choice) has been identified as an important determinant of gender gaps in wages and employment.[2]

Before entering the labor market, US workers with at least a bachelors degree are faced with a menu of potential undergraduate college majors that is nearly as varied as the opportunities to specialize once in the labor market. Until recently, large, multi-cohort data linking detailed college major choice to subsequent labor market outcomes was not available. Due to this empirical constraint, the impact of pre-market specialization (e.g. major choice) has not received as much attention. Further, very little is known about gender differences in the linkages between major and occupation.

In this paper, we use new data to assess gender differences in pre-market human capital specialization and examine the extent to which such gender differences in specialization have converged over time. Furthermore, we fill an important gap in the literature by examining gender gaps in the mapping of college major to subsequent occupational specialization. To go beyond simply documenting gender differences in pre-labor market specialization (college major choice) and occupational sorting conditional on pre-labor market specialization, we introduce new indices that measures specialization in potential wage space. In doing so, we highlight that women both choose majors with lower potential earnings and that they then subsequently sort into occupations with lower potential earnings given their major choice. We then show that gender differences in major choice explains a substantive portion of the gender wage gap among college graduates above and beyond what is explained by gender differences in occupational sorting. Finally, we show evidence that some of the gender differences in occupational sorting conditional on college major can be explained by women choosing occupations with lower potential hours worked. Collectively, our results highlight the importance of understanding gender differences in pre-labor market specialization and

[1]See 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) For literature on overtaking see Becker et al. (2010), Charles and Luoh (2003), DiPrete and Buchmann (2006), Goldin et al. (2006), and Jacob (2002).

[2]The 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).

the mapping between college major and occupational sorting when studying the evolution gender differences in labor market outcomes over time.

The paper exploits 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.[3] These questions are asked regardless of age. Given that major choice is constant for an individual over their life-cycle, this data allows us to explore how pre-market human capital specialization decisions have changed across cohorts. The data also measures an individual's current labor market status allowing us to link undergraduate major to employment status, occupation, and wages for different cohorts of both men and women.

Similar to others in the literature, we document that men and women systematically choose different college majors and those differences have evolved over time. Women have 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.

One innovation of the paper is to measure gender differences in major choice in potential wage space. 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.[4] 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

[3]The ACS covers cohorts from the periods pre-dating and including a nation-wide expansion in the supply of college graduates. The ACS has been used by Altonji et al. (2014) to measure the evolution of wage inequality within college majors and Altonji et al. (2016) to estimate major-specific labor market returns.

[4]Using a similar methodology to rank undergraduate majors in the ACS, (Bertrand, 2017) notes gender convergence at the 90th percentile, 80th percentile, and mean of log earnings across birth cohorts.

the 1950 and 1975 birth cohorts with a modest divergence for recent cohorts.

Convergence in major choice is of a similar order of magnitude as the convergence in occupational choice. However, conditional on matriculating with a given major, occupational sorting varies dramatically by gender. For example, among the 1968-1977 birth cohort, 63 percent of female and just 46 percent of male Nursing/Pharmacy majors reported worked in the nursing occupation in the 2014-2017 ACS. In this same birth cohort, men who majored in Nursing/Pharmacy were much more likely than women to sort into Executive/Manager occupations and Sales occupations. Women who majored in Nursing/Pharmacy became nurses. Conversely, men who majored in Engineering were much more likely than women to sort into engineering occupations. For each undergraduate major, we compute Herfindahl-Hirschman Indices (HHI) of occupational choice separately for men and women. We document that for high potential wage majors, men's occupational sorting is systematically more concentrated than women's.

To quantify whether women are systematically sorting into low-wage occupations conditional on their undergraduate major we introduce another new index. Specifically, the index measures gender differences in potential wages based on occupational choice conditional on matriculating with a degree in a given major. Again, we measure an occupation's potential wages based on the wages of older men. We find that on average women born between 1968 and 1977, work in occupations with roughly 10 percent lower potential earnings conditional on major choice. [5] This gap occurs in both higher wage majors (Engineering) and low-wage majors (Education) and occurs in majors dominated by women (Nursing) and majors dominated by men (Business).

While women systematically work in lower pay occupations conditional on major choice, this gap has narrowed somewhat over time. For example, for the 1950 birth cohort, women who majored in Engineering sorted into subsequent occupations with potential wages that were 14 percent lower than the corresponding men who majored in Engineering. For the 1980 birth cohort, however, women who majored in Engineering ended up working in occupations with potential wages that were only 4 percent lower. 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 taken by women who majored in Engineering used to differ markedly from men but now the occupational choice conditional on major has converged.

[5] We primarily focus this exercise on a single cohort in their prime-earnings years. Later, we compare mapping across cohorts with the caveat that these comparisons will necessarily confound age and cohort effects to some extent.

The patterns for Engineering are broadly similar to the patterns for all majors. For the 1950 birth cohort, women work in 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 somewhat for recent cohorts. For example, women from the 1980 birth cohort systematically work in occupations with about 9 percent lower potential earnings conditional on their undergraduate major choice. Interestingly, we find this convergence is driven by movements of women at the top of the major-pay distribution into higher-pay occupations. We show that female sorting into lower hours-worked occupations 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.

Separately, 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 sort 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.[6]

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.[7] Within a broader analysis of factors contributing to the gender wage gap such as psychological attributes and demands for flexibility, Bertrand (2017) uses data from the 2012 to 2015 ACS to briefly document cross-cohort convergence in potential wages based on degree attainment and major. The goal of our paper is to link and expand on these two strands of the literature. Additionally, we add to this literature by providing novel results on gender differences in the mapping between major choice and occupational sorting and how those differences have evolved over time.

[6]The 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.

[7]Black 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. Ironically, their data end in 1990, about the time women's field of studying started diverging from men.

2 Data

Our primary data source is the 2014-2017 American Community Survey (ACS).[8] 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.[9]

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 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.[10]

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.[11] Lastly, we use the data

[8]The 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).

[9]Of 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. We use inverse probability weighting to correct for non-response. See more details in the Online Appendix.

[10]In 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.

[11]We use a balanced panel of detailed occupation codes based on the 1990 Census detailed occupation

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 Gender Differences in College Major Choice

In this section, we document the presence and evolution of gender gaps in undergraduate major choice. For comparison, we also document trends in relative occupational choice for college-educated men and women. In the subsequent section, we discuss gender differences in the mapping between majors and occupations.

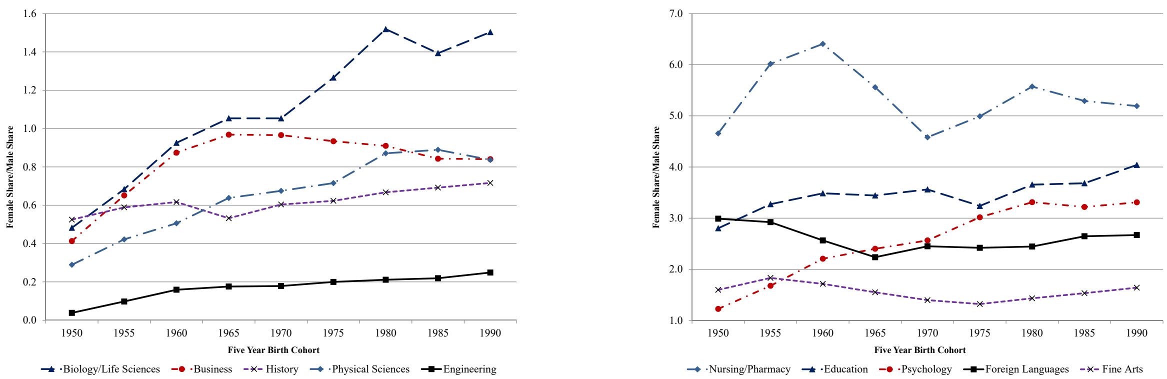
For each survey year between 2014 and 2017, we begin by assigning each individual a 5-year birth cohort based on current age in the survey year.[12] In the U.S., most individuals who ever complete a bachelor's degree do so by their mid-twenties; this implies that undergraduate major is largely fixed 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 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, a period when the Business major itself was expanding. Thereafter, in a period that was marked by a contraction overall in Business majors, women and men once again diverged. History was male-dominated historically and experienced little convergence or divergence over subsequent birth cohorts.

codes following (Dorn, 2009). This yields 251 detailed occupation codes in our analysis.

[12]We 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.

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.[13] 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.

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:

[13]The 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.

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

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.[14] 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 $\bar{Y}^m_{male}$, a potential wage based on major that is plausibly unaffected by post-educational factors. Specifically, we define $\bar{Y}^m_{male}$ 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.[15] For example, for anyone (male or female) 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:

[14] This is similar to an index developed in (Bertrand, 2017).

[15] Market 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 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.

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

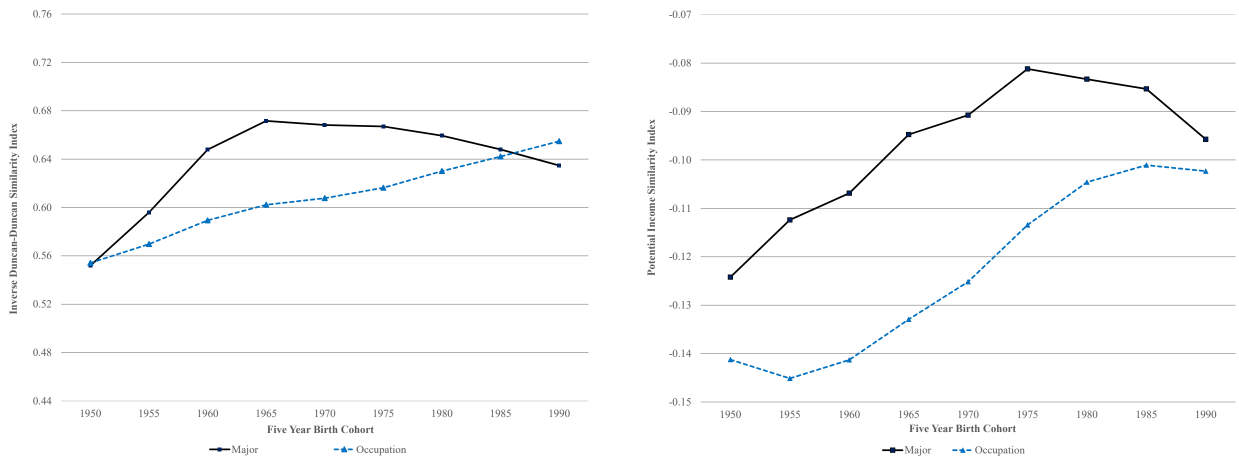
where $I_c^{P,M}$ measures the differential "potential" log wage of women of cohort c given that the female distribution of major choice in a given cohort may differ from males in their cohort. A value of $I_c^{P,M} = 0$ means that the major choices of women yield the same potential log wage as their 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 U.S. 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. We document a surprising new fact that 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.[16]

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 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

[16] We 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



Panel A: Segregation Index

Panel B: Potential Wage Index

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.

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 a been a modest divergence in major choice across genders for recent cohorts, occupational choice has continued to converge.[17] Panel B also shows that there was strong convergence in occupational segregation

[17] We have replicated the gender convergence in occupational choice using the historical U.S. Censuses.

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 somewhat 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 potential wage index based on occupation choice ($I_c^{P,O}$, dashed line) is consistently lower than potential wage index based on major choice ($I_c^{P,M}$, solid line). This implies that conditional on choosing the same major as men, women systematically work in lower-pay occupations. In this section, we explore gender differences in the mapping of undergraduate major choice to subsequent occupation and document how these differences have evolved over time.

Table 1 shows the broad occupational distribution for men and women born between 1968 and 1977 for selected broad major categories. We summarize the occupational distribution by showing the four most common occupations associated with each major. Unlike undergraduate major, occupation may vary over an individual's lifetime. Similar to above, we measure occupations for this birth cohort using data from 2014-2017 ACS. Therefore, this table measures gender differences in occupational choice when these individuals are between the ages of 38 and 47.

The table illustrates that there are sharp differences in occupational sorting between men and women even conditional on major choice. For illustrative purposes, we start by considering patterns among Education majors. In the 1968-1977 birth cohort, 68 percent of all women and just 50 percent of all men who majored in education report their occupation

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.

Table 1: Gender Differences in Occupational Choice, Selected Majors, 1968-77 Birth Cohort

Panel A: Education Majors					
	Teachers	Executive/ Manager	Sales	Admin Support	$HHI^m_{g,c}$
Men	0.50	0.18	0.06	0.03	0.29
Women	0.68	0.09	0.03	0.07	0.48

Panel B: Nursing/Pharmacy					
	Nurses/ Health	Executive/ Manager	Sales	Health Technicians	$HHI^m_{g,c}$
Men	0.46	0.15	0.07	0.06	0.25
Women	0.63	0.09	0.03	0.05	0.42

Panel C: Social Sciences					
	Executive/ Manager	Sales	Lawyers/ Judge	Admin Support	$HHI^m_{g,c}$
Men	0.26	0.13	0.11	0.06	0.11
Women	0.20	0.07	0.08	0.13	0.10

Panel D: Business					
	Executive/ Manager	Sales	Accountant/ Underwriter	Admin Support	$HHI^m_{g,c}$
Men	0.31	0.18	0.12	0.07	0.16
Women	0.24	0.11	0.17	0.18	0.14

Panel E: Engineering					
	Executive/ Manager	Engineer	Other Technicians	Architects/ Civil Engin.	$HHI^m_{g,c}$
Men	0.28	0.23	0.09	0.08	0.16
Women	0.27	0.18	0.05	0.07	0.13

Note: Table shows the occupational distribution of men and women born between 1968 and 1977 for different majors. We use both broad major categories and broad occupational categories for this table. Each panel shows a different undergraduate major. The cells of the panel show the fraction of men (women) who majored in that occupation who subsequently worked in different broad occupations in the 2014-2017 ACS. For each major, we show the four largest occupational categories based on where men with that major in that age range worked.

as “teacher” in the 2014-2017 ACS. Such patterns are robust across all birth cohorts. For example, for those majoring in education, 72 percent of women from the 1978-1987 birth cohort, 63 percent of women from the 1958-1967 birth cohort and 52 percent of women from the 1948-1957 birth cohort worked as teachers. The comparable numbers for men from the various birth cohorts were 58 percent, 43 percent, and 28 percent.

In this simple exercise, rank effects are stark. For example, men from the 1968-1977 birth cohort who majored in education are twice as likely as women (18 vs. 9 percent) to be executives or managers – including principals and superintendents. Women who majored in education are more than twice as likely as men (7 vs. 3 percent) to work in administrative support roles– including teachers' aides, administrative assistants, and office supervisors. For all cohorts, men majoring in education were roughly twice as likely as women to be in executive or managerial occupations and women were twice as likely to be in administrative support occupations.

Panels B-E show similar patterns for individuals majoring in Nursing/Pharmacy, Social Sciences, Business, and Engineering. In all the majors we explored in Table 1, conditional on undergraduate major women are less likely to be executives and managers and more likely to work in administrative support occupations. This glass ceiling effect is smallest among Engineering majors where men and women are almost equally likely (28 v. 27 percent) to be executives and managers; however, male Engineering majors are much more likely than female Engineering majors to work as engineers (23 v. 18 percent).

How systematically different is the occupational sorting of men and women conditional on undergraduate major? To answer this question, we define two measures to summarize the gender differences in occupational sorting after matriculating with a given major. First, we create a cross-occupation Herfindahl-Hirschman Index ($HHI^m_{g,c}$) for each gender g and cohort c for every major m.[18] This index measures the occupational concentration for individuals in major m separately by gender. Given we are measuring occupational sorting as shares, our occupational concentration index ($HHI^m_{g,c}$) ranges from 0 to 1. Higher levels of $HHI^m_{g,c}$ imply that occupational sorting is more concentrated.

Referring back to Table 1, column (5) shows the $HHI^m_{g,c}$ for the five broad major categories. Some major categories, such as Education and Nursing/Pharmacy, have more highly concentrated occupational distributions for both men and women. For both majors, women's occupational choices are more concentrated that men's. In the Engineering majors, men's occupations are more concentrated than women's. Appendix Table A2 reports $HHI^m_{g,c}$ sep-

[18]Formally, $HHI^m_{g,c} = \sum_o^O (s^{o|m}_{g,c})^2$ where $s^{o|m}_{g,c}$ is the share of group g from cohort c working in occupation o conditional on having major m. For each gender and cohort who matricate with a given major, $\sum_o^O s^{o|m}_{g,c} = 1$ where O is the total number of potential occupations.

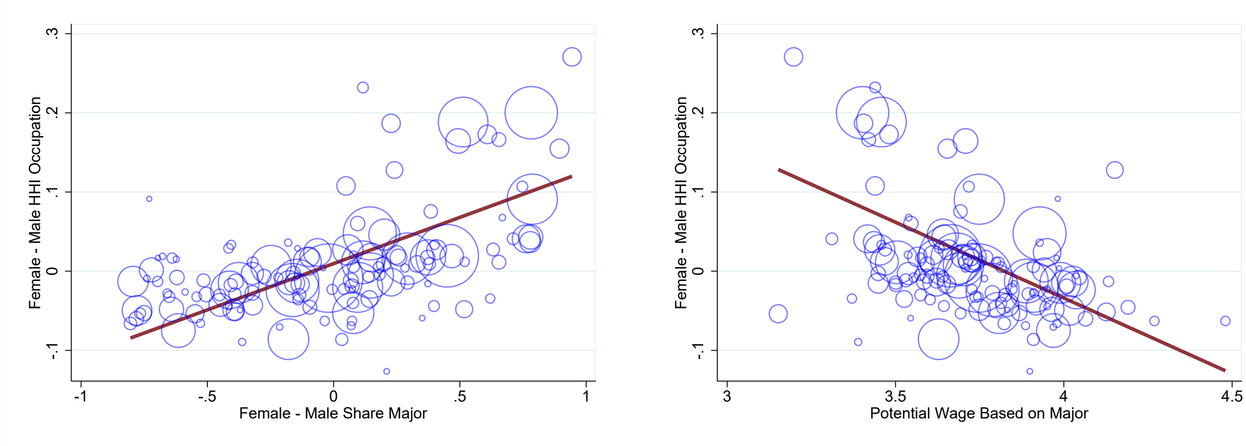
arately for both men and women from the 1968-1977 birth cohort for each broad major category. For many broad major categories, the degree of occupational dispersion is quite similar for men and women at the broad major level. Notable differences arise, however, in some broad majors like Computer and Information Sciences and Policy and Social Work.

The concentration indices ($HHI^m_{g,c}$) shown in Table 1 use broad major categories. We can also compute these indices using detailed major categories.[19] Panel A of Figure 3 plots the female-male difference in $HHI^m_{g,c}$ for each detailed major ($HHI^m_{female,c}$-$HHI^m_{male,c}$) against the female-male difference in the share matriculating with that major ($s^m_{female,c}$ - $s^m_{male,c}$). Positive values on the y-axis indicate majors where female occupational sorting is more concentrated than males. An example of such a major is Early Childhood Education where most females become teachers while more men disperse into executive/manager roles. Conversely, negative values on the y-axis indicate majors where male occupational sorting is more concentrated. An example of this type of major is Computer Science where most men become engineers and programmers while women are more likely to disperse into other occupations. For this figure, we again restrict our analysis to the 1968-1977 birth cohort. As seen in Panel A, there is a strong and significant positive gradient between the extent of relative female occupational concentration (conditional on major choice) and the extent to which women are relatively more likely to choose that major. In female-dominated majors, women sort into a narrower set of occupations than men. In male-dominated majors, men sort into a narrower set of occupations than women.

Panel B of Figure 3 plots the female-male difference in $HHI^m_{g,c}$ for each detailed major (on the vertical axis) against $\bar{Y}^m_{male}$ (on the horizontal axis). The vertical axis of Panel B is identical to that of Panel A. Recall, $\bar{Y}^m_{male}$ measures the log wages of native white men aged 43-58 who graduated with undergraduate major m. Detailed undergraduate majors with high potential earnings (like Finance) are on the right of the figure while lower pay majors (like Fine Arts) are on the left of the figure. Panel A suggests a symmetry in the relative occupational dispersion of men and women across gender-dominance of majors. However, Panel B suggests a dimension in which not all dispersion is created equal. Specifically, this figure shows a strong and significant negative gradient between the relative occupational concentration between men and women conditional on major choice and how lucrative that major is. Men tend to have a higher occupational concentration in high earning majors. These patterns suggests that when the potential returns to a major are low, men disperse into a wider set of occupations than women to avoid the wage penalty of a low-return major. However, when the potential returns to a major are high, men sort into a *narrower* set of

[19] When computing the concentration measures for detailed majors, we still use broad occupation categories as an input into $HHI^m_{g,c}$.

Figure 3: Cross Major Variation in Within-Major Gender Differences in Occupational Dispersion, 1968-1977 Birth Cohort



Panel A: Gender Composition of Major

Panel B: Potential Income of Major

Notes: These figures show cross-major variation in $HHI^m_{g,c}$ as a function of how female dominated is the major (panel A) and average major potential income (panel B). See text for additional details. Each observation in both panels is a detailed major. We use broad occupation categories. Data is shown only for the 1968-1977 birth cohort. Both panels include a fitted regression line. The slopes of the regression lines are -0.19 (standard error = 0.027) and 0.12 (standard error = 0.011), respectively.

occupations than women.

A shortcoming of the above concentration index ($HHI^m_{g,c}$) is that, like the Duncan-Duncan index, it does not allow for any rank ordering of occupations. Consider two scenarios. In the first scenario, all men from a major sort into the same high-pay occupation and all women from the major sort into the same low-pay occupation. Contrast that with a second scenario where both men and women from the major sort into the same high-pay occupation. In both scenarios, men and women will have the same measure of occupational concentration ($HHI^m_{g,c}$). The concentration measure ($HHI^m_{g,c}$) does not account for men and women sorting into occupations of different quality. To overcome this limitation, we create a second index based on occupational sorting by potential income. Specifically, we define the following index:

$$I_c^{P,O|m} = \sum_{m=1}^{M} (s^o_{female,c}|m)\bar{Y}^o_{male} - \sum_{m=1}^{M} (s^o_{male,c}|m)\bar{Y}^o_{male} \tag{3}$$

where $s^o_{g,c}|m$ is the share of gender g choosing occupation o conditional on major m from cohort c. In other words, $I_c^{P,O|m}$ measures the gender gap in *potential occupational earnings* for a given cohort c conditional on major choice m. Any gender difference in this index reflects gender differences in occupational wage rank conditional on major choice. $I_c^{P,O|m} < 0$

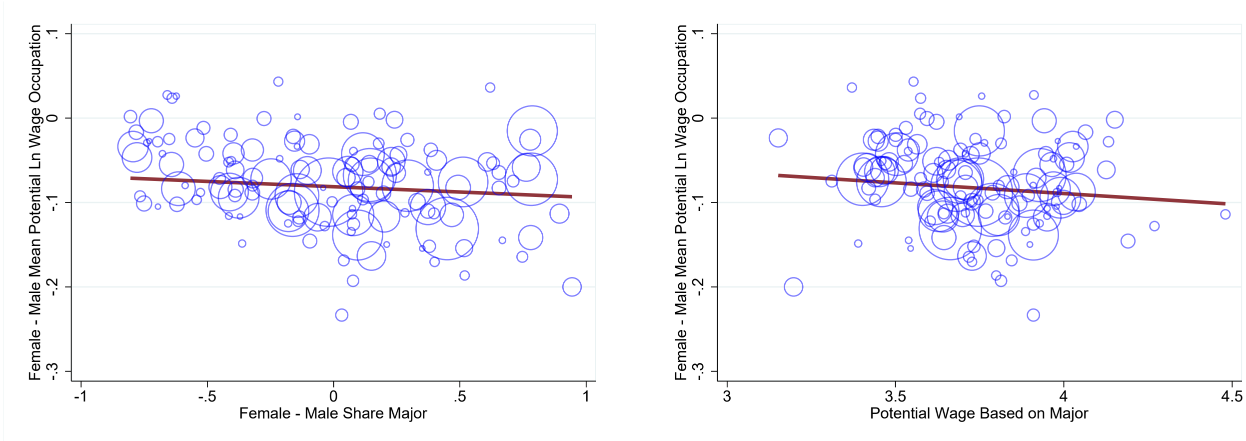
implies that women with major m from cohort c are systematically in occupations with lower potential wages compared to men from the same major and cohort. For example, if $I_c^{P,O|m} = -0.1$, it implies women in detailed major m choose occupations that are associated with 10 percent lower earnings, on average. To reiterate, as in Figure 2, average potential wages of an occupation ($\bar{Y}^o_{male}$) is based on the wages of men aged 43 to 57 with strong attachment to the labor market. 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 4 is the analog of Figure 3 except the vertical axis now being $I_c^{P,O|m}$.[20] Specifically, Panel A of Figure 4 plots $I_c^{P,O|m}$ against $s^m_{female,c} - s^m_{male,c}$ (on the horizontal axis). Recall, $s^m_{female,c} - s^m_{male,c}$ measures the extent to which the major is female-dominated. There are two things of note from Panel A of Figure 4. First, in essentially all detailed majors, women sort into occupations where potential earnings are lower relative to men. Second, there is no strong, significant linear relationship between whether a major is female-dominated and the propensity for women to sort into lower-pay occupations. Women from the 1968-1977 birth cohort earn roughly 9 percent less than their male counterparts, conditional on their undergraduate major, regardless of their relative propensity to choose that major relative to me. Likewise, panel B of Figure 4 plots $I_c^{P,O|m}$ against $\bar{Y}^m_{male}$ (on the horizontal axis). Again, there is no strong, significant linear relationship between the potential wage of the major and the propensity for women to sort into lower-pay occupations conditional on major choice. These patterns stand in contrast to the pattern shown in Panels A and B of Figure 3 and are consistent with the fact that female-male differences in occupational concentration conditional on major choice occurs because women are more likely to sort into lower earning occupations. Appendix Tables A3-A6 rank detailed major categories for the 1968-1977 cohort by $I_c^{P,O|m}$.

The above figures highlighted the cross-major variation in occupational sorting between men and women for a given cohort - those born between 1968 and 1977. How has occupational sorting conditional on major choice evolved across cohorts? Figure 5 shows the cross-cohort evolution of occupational differences conditional on major choice between men and women. We again measure occupational differences using $I_c^{P,O|m}$. For consistency, the figure focuses on the same majors highlighted in Figure 1. Figure 5 highlights 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,

[20] As with Figure 3, Figure 4 uses detailed majors and broad occupations to make the index and restricts the analysis to the 1968-1977 birth cohort.

Figure 4: Cross Major Variation in Within-Major Gender Differences in Potential Wage by Occupation, 1968-1977 Birth Cohort



Panel A: Gender Composition of Major

Panel B: Potential Income of Major

Notes: These figures show cross-major variation in $I_c^{P,O|m}$ as a function of how female dominated is the major (panel A) and average major potential income (panel B). See text for additional details. Each observation in both panels is a detailed major. Data is shown only for the 1968-1977 birth cohort. Both panels include a fitted regression line. The slopes of the regression lines are -0.013 (standard error = 0.009) and -0.025 (standard error = 0.017), respectively.

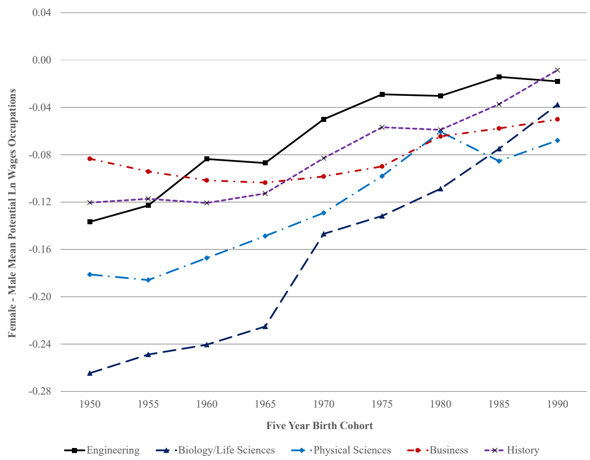
many women who graduated with an Engineering major ended up working in lower-paying occupations than men who also majored in Engineering. Across cohorts, women who majored in Engineering were more likely to be in occupations that were similar to the men who majored in Engineering. For Engineering, our index $I_c^{P,O|m}$ increased from -0.14 to -0.02 between the 1950 and 1990 cohorts.[21]

The gender convergence in occupational choice within majors is seen in many but not all of the occupations shown in Figure 5. Many of the historically female-dominated majors like Education, Foreign Languages, and Fine Arts saw only modest convergence across cohorts in the occupations taken by women relative to men (in potential wage space) conditional on major choice (Panel B). Collectively, the patterns in Figure 5 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.

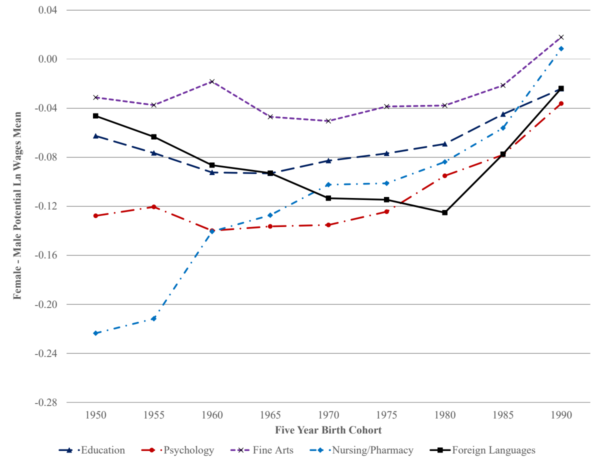
The time series trends provided $I_c^{P,O|m}$ in Figure 5 are somewhat limited. The figure only

[21] As our data is from the ACS, we observe birth cohorts at different points in the life cycle. For majors, this will be less important as major at graduation is sticky over an individual's life cycle. For occupations, this issue is non-trivial. For example, we observe some cohorts at the beginning of their working careers while others are observed at the end. To the extent to which occupation is dynamic over the lifecycle and increasingly dynamic across generations, this will complicate the interpretation of cross-cohort differences in occupation-based results. This limitation should be kept in mind when we highlight the evolution of occupational sorting across cohorts.

Figure 5: Within-Major Gender Differences in Potential Wage by Occupation, by Gender and Cohort



PANEL A: MALE DOMINATED MAJORS



PANEL B: FEMALE DOMINATED MAJORS

Notes: These figures show the trends in $I_c^{P,O|m}$ 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, $\bar{Y}^o_{male}$, 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.

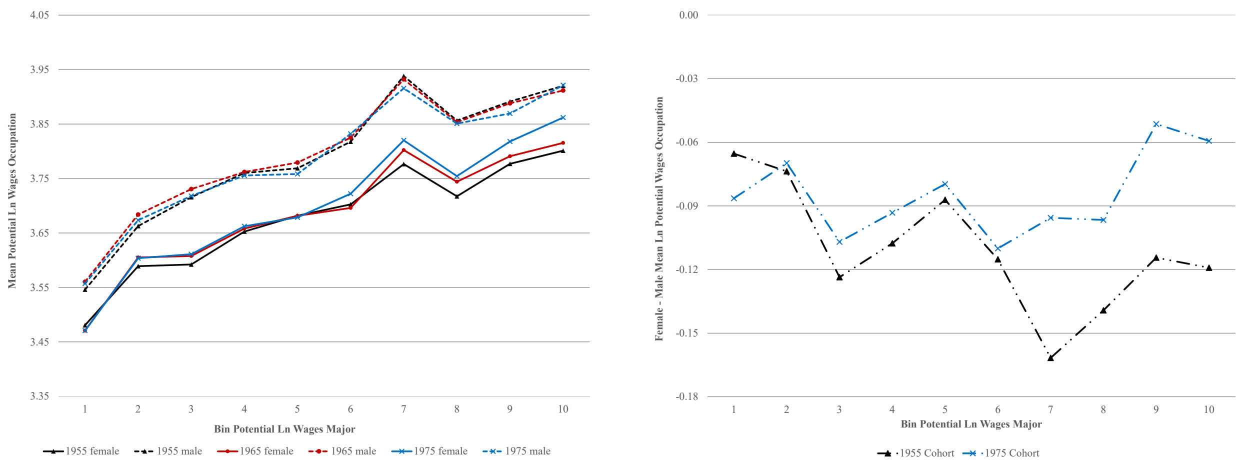
displays results for a few broad majors at a time. Further, it describes majors in a taxonomic fashion by name and gender endowment (Panel A vs. Panel B), but without an economic ranking. Thus, we create one final index, $I_c^{P,O|d,g}$ to summarize heterogeneous patterns of gendered sorting into occupations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

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

where $s^o_{g,c}|d$ 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 $\bar{Y}^m_{male}$ 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. In order to rank majors for this exercise, we begin by dividing $\bar{Y}^m_{male}$ by decile. Then, we average $\bar{Y}^o_i$, 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 6. The x-axis of Figure 6 segments majors into deciles based on $\bar{Y}^m_{male}$. The top decile includes majors like Economics, Chemical Engi-

Figure 6: Mapping of Potential Wage by Major to Potential Wage by Occupation, by Gender and Cohort



Panel A: Levels
Selected Cohorts

Panel B: Differences
Selected Cohorts

Notes: These figures show the mapping between major and occupation. On the x-axes, we have binned majors based on $\bar{Y}^m_{male}$, 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.

neering, 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 6. Each dashed line represents a different 5-year 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).[22]

[22] As potential wages for both majors and occupations are based on the wages of native, white men in their

For women, monotonocity in Figure 6 can be violated because of gender differences in occupational choice. In fact, what is compelling about Panel A of Figure 6 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 taken by comparable men. Again, 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.

Panel B of Figure 6 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 worked in 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 sort into occupations with lower hours worked (e.g., Goldin and Katz (2011) and Cortez and Pan (2016)). How much of the patterns in Figure 6 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 $\bar{H}^o_{male}$. 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.

peak earnings years, deviation from monotonicity within the male series can only arise from race, cohort, or age effects within men.

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 6). 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 $\bar{Y}_i^o$.[23] 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 2 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

[23] In Table A7 of the Online Appendix, we report results from an alternate specification where we do not include demographic controls or time fixed effects. In Table A8 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 2.

Table 2: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment

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

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

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

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.[24] 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 2 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.[25] 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 2 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 substantially alter the estimated gender gap in employment rates for women with a bachelor's degree. [26] 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. Given the effect of major choice on the gender gap in wages, the extensive margin employment result is surprising. It points to the importance of potential effects of specialization on gender gaps in the intensive margin of employment which we discuss later in the paper.

[24]The estimated gender wage gap for college graduates using the pooled cohorts from the 2014-2017 ACS 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.

[25]These 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.

[26]This specification cannot control for occupational choice given that occupation is often not defined for those who are not working. Occupation is recorded for those who are not working only if they were employed at some point in the prior five years.

The bottom panel of Table 2 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).[27] 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 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 2 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.

Panel B of Table 2 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 A10 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

[27] To 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.

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

Do changes in the range of specialization decisions made by college graduates– namely college major and subsequent occupational choice conditional on college major– explain changes in the college gender wage gap over the last half century within the U.S.? In our analysis of this broad question, we exploited new data that links undergraduate major and labor market outcomes for many birth cohorts, and documented long-run trends for a nationally representative sample of college graduates. Over the past 40 years, men and women have chosen more similar undergraduate majors. 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.

Our analysis shows that there are gender differences in the mapping of undergraduate major to occupation. In many broad major categories, we show substantial rank effects in occupational choice with men being more likely to work in Executive/Managerial occupations than women. We document that in majors that are female-dominated, women sort into a narrower set of occupations than men, and in majors that are male-dominated men sort into a narrower set of occupations. However, we also show that there is an asymmetry in the relationship between potential earnings of a major and occupational dispersion by gender. We find that men disperse into a wider range of occupations than women when market returns to a major are low and disperse into a narrower range of occupations than women when market returns to a major are high. Using an innovative new metric to account for the wage-cost of gender differences in major-to-occupation mapping, we document a systematic sorting of women into lower potential wage and lower potential hours-worked occupations than men conditional on having made the same major choice. There is a modest convergence between the 1950 and 1990 birth cohorts in the gendered mapping of major to occupation. This set of findings on mapping between major and occupation marks a unique contribution to the literature.

Throughout the paper, we assessed the extent to which undergraduate major has predictive power above and beyond occupational choice, and conclude that college major choice has strong predictive power independent of occupation choice. In fact, in the analysis in our pooled sample, we find that collectively controlling for occupation and major reduces the gender wage gap among the college-educated by half. With respect to independent predictive power, adding major to our model reduces the gender wage gap by an additional 2.8 percentage points (on a base of 0.23 log point gap in wages between college men and women) beyond a model that only accounts for occupational specialization.

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. Although, we do not find any effects of specialization on the extensive margin of participation, we document interesting findings with respect to the intensive margin. 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.

The above analysis and subsequent findings are set in an era in which both the 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, "The glass ceiling," *Becker Friedman Institute for Research in Economics Working Paper*, 2017, (2018-38).

_ **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.