Firm Sorting, Field of Study, and the Gender Earnings Gap

Cristobal Huneeus[1], Federico Huneeus[2], Conrad Miller[3], Christopher Neilson[4], and Seth Zimmerman*[5]

[1]Unholster
[2]Yale and Central Bank of Chile
[3]UC Berkeley and NBER
[4]Princeton and NBER
[5]Yale and NBER

January 2021

Abstract

A growing body of evidence shows that differences in firm-specific pay premiums account for a large share of the gender pay gap. This paper asks how college major mediates access to high-paying firms, and what this means for the gender pay gap. Using employer-employee data from Chile matched to educational records, we show that differences in field of study account for more than 60% of the firm contribution to the gender pay gap. Degrees in Technology, which are numerous, male dominated, and associated with high firm premiums, drive these effects.

*We thank Sydnee Caldwell, Pat Kline, Jonathan Leonard, and David Levine for comments.

I Introduction

A growing body of evidence shows that differences in firm pay premiums account for a large share of the gender pay gap overall and for college-educated workers in particular (Card et al., 2016; Sorkin, 2017; Bruns, 2019; Casarico and Lattanzio, 2019; Addario et al., 2020; Morchio and Moser, 2020). Women and men also differ dramatically in their pre-market education. Men are more likely to major in high-earning STEM and business fields, while women are more likely to study lower-paying fields such as education, social sciences, and the humanities (Altonji et al., 2012; Hastings et al., 2013; Rodriguez et al., 2016; Altonji et al., 2016; Kirkeboen et al., 2016; Sloane et al., 2019).

Understanding the connection between these facts is important for two reasons. First, if pre-market major mediates the firm component of the gender pay gap, it may be difficult to address that gap with interventions in the labor market alone. Second, if the returns to field of study depend on firm premiums, traditional models of the returns to major as depending on skill accumulation and skill prices (e.g. Altonji et al. 2012; henceforth ABM) may not accurately capture the productivity effects of education policies that change the composition of college majors, such as tuition policies that cross-subsidize STEM fields (Stange, 2015; Altonji and Zimmerman, 2018) or immigration policies that encourage immigration for STEM degree holders relative to holders of degrees in other fields (U.S. Citizenship and Immigration Services, 2020).

In this paper we use rich data on educational backgrounds linked to employer-employee matched data to study how the returns to major depend on firm premiums, and how this mediates the firm component of the gender earnings gap. Our data come from Chile, a middle-income OECD member country. Records include gender, the degree programs in which students enroll, subject test scores, as well as earnings, industry, and (anonymized) employer identifiers. Our data cover all UI-eligible workers between 2009 and 2016, and all students who take the national college admissions exam between 1977 and 2003. We use these data to estimate wage determination models that allow for worker and employer fixed effects (Abowd et al. 1999; henceforth AKM).

We have two main findings. First, majors where average worker effects are high also have high average firm effects. The covariance between average worker effects and firm premiums explains 42% of cross-major variance in earnings. A leading example is Technology majors, for which firm premiums explain nearly half of the 17.8% earnings premium relative to the average major. In contrast, the covariance between worker effects and firm premiums explains only 13% of the within-major variance in earnings. Hence, assortative matching is substantially stronger *between* major than *within* major. The second is that the firm component of the gender gap is largely a between-major phenomenon, with between-major differences accounting for between 63% and 71% of this gap. In short, a large share of cross-major earnings differences reflects sorting to high-paying firms, and because men are over-represented in majors with high firm premiums, premarket differences in major explain much of the firm component of the gender gap.

We begin by showing that features of the wage distribution observed in other countries are present in Chile as well. Men in our data earn 20% more than women; this gap increases to 25% for students admitted to college. This is similar to the 23% male-female gap reported in Card et al. (2016; henceforth CCK) using Portuguese data, and 23% gap in Sloane et al. (2019; henceforth SHB) report for female college graduates in the US. We also observe large differences in earnings across majors. Breaking down majors into ten coarse, administratively-defined groups, we find that students enrolling in the highest paying group, Technology, earn 77% more than students in the lowest-paying group, Education. This difference is similar to the 70% gap between engineering and education majors that ABM report in the US. Men in our data are 2.5 times more likely to study Technology than women while women are three times more likely to study Education than men.

Our first main finding is that differences in firm premiums play an important role in explaining cross-major earnings differences. Overall, firm effects explain 12.5% of the between-major variance in mean earnings across coarse major categories, while the covariance between firm effects and worker effects explains 42%. This effect is driven by strong assortative matching across popular fields like Technology and Education. The mean firm premium for Technology majors is 0.073, 41% of the 0.178 log point earnings premium for Technology majors relative to the average major. For Education majors, the mean firm premium is -0.131, 33% of the overall major effect.

Our second main finding is that field of study explains a majority of the firm contribution to the gender earnings gap. Among college admits, the male wage premium is 0.252 and the firm effect premium is 0.081, or 32% of the total gap. This is comparable to the firm contribution to gender pay gaps documented in prior studies in Brazil (Morchio and Moser, 2020), Germany (Bruns, 2019), Italy (Casarico and Lattanzio, 2019), Portugal (Card et al., 2016), and the United States (Sorkin, 2017). After conditioning on broad field of study, the gender gap in mean firm premium falls by 63%, from 0.081 to 0.030. This result is closely related to our finding of differences in the firm contribution to major effects. Technology, the field with the highest mean firm premium, is disproportionately male, and Education, the field with the lowest firm pay premium, is disproportionately female.

Several additional analyses supplement our core results. First, we follow CCK and decompose the gender gap in firm premiums into what CCK refer to as *sorting* and *bargaining* components, where the latter measures gender differences in pay premiums within the same firm. To do this decomposition, we augment our baseline AKM specifications to allow for firm premiums to vary by gender and major. We then decompose the overall firm contribution to the gender gap into between-firm and within-firm components by assigning men (women) the value of the typical female-specific (male-specific) premiums at their firm, and recomputing firm wage premiums. Overall, the sorting component explains 80-85% of the firm contribution to the gender earnings gap. Roughly 60% of the variation in sorting effects occurs between coarse major groups and 75-80% of this variation occurs between narrow groups. In contrast, there is almost no between-major variation in the bargaining component of the gender earnings gap.

Finally, we show that the relationship between field of study and firm premiums does not simply reflect sorting on vertical measures of skill, as measured either by college admissions test scores or estimated worker effects. Conditioning on either exam scores or estimated worker effects reduces the firm component of the gender gap by about one quarter. Conditioning on coarse major categories reduces the firm component by 57%, a much larger share. Major retains most of its explanatory power even after conditioning on ability measures. Compared to exam scores, field of study is more closely tied to firm effects. While firm effects and sorting account for only 3% and 27% of the variance of earnings attributable to exam scores, they account for 9% and 36% of the variance in earnings across majors. The role of major in matching students to firms does not simply reflect the sorting of high-ability students to high-paying majors.

We make two contributions to existing work. The first is to show that differences in the firm component of the gender earnings gap for the college-educated are largely determined by what workers studied in school. An implication here is that labor market policies alone may not suffice to reduce the firm component of the gender gap. Framed another way, policies that increase female STEM participation may reduce the gender earnings gap in part by reducing the firm component of that gap. The second is to point out how firm effects mediate cross-field differences in earnings outcomes. Canonical models of the return to major focus on skills and market-wide skill prices, and abstract from the role of firm match (e.g. ABM, AAM). Our findings suggest that understanding major wage premiums requires an understanding of job matching in addition to an understanding of skill formation and skill demand.

We emphasize two claims we do not make in this paper. First, we do not claim that our OLS estimates of the returns to major or the firm component of those returns reflect the causal effects that would accrue if a student were moved at random from one major to another. Our contribution is to provide new evidence on the role firms play in driving major premia in the cross section. Second, that college major can explain a large share of the firm component of the gender gap does not mean that discrimination on the basis of gender is not important. What it does mean is that one cannot eliminate the firm component of the gender gap merely by making firm sorting independent of gender within major; most of that gap would remain if firms did not also change the distribution of hires by major. This is in addition to gender discrimination prior to labor market entry that may affect the gender composition of majors (Price, 2010; Carrell et al., 2010; Zafar, 2013; Buser et al., 2014; Paredes et al., 2020).

II Data and Context

We use data from three sources: college admissions exams, college applications, and earnings records from the unemployment insurance (UI) system.

We use data on college admissions outcomes and exam scores covering applicants from 1977 to 2003. Students apply to nearly all public and private non-profit universities in Chile through a centralized system run by DEMRE (*Departamento de Evaluación, Medición y Registro Educacional*), the Chilean college admissions authority. Students applying through this process submit an ordered list of institution-major pairs, which we will henceforth call 'degree programs'. DEMRE assigns students to degree programs using a deferred acceptance algorithm in which degree programs have a fixed number of open spots and evaluate students using weighted averages of standardized test scores and high school grades. As with the SAT in the US, students may take a variety of subject tests in addition to required math and reading modules. When we use test score data we focus on the required math and reading exams. The standardized admissions exam in Chile was known as PAA until the early 2000s, and as the PSU thereafter. In what follows, we use the name of the current exam, the PSU, to refer to all admissions exams.

Each student is assigned a spot in the most-preferred program willing to accept them. We classify student majors using attributes of the accepted degree program. We do not have access to enrollment data, but evidence reported elsewhere indicates that enrollment rates in the admitted program are very high (HNZ). Students are not allowed to switch programs after enrollment; if they want to join another degree program they typically must drop out and re-apply through the centralized system. Degree programs are classified by field of study based on UNESCO Normalized International Classification of Education standards. There are ten categories: Agriculture, Art and Architecture, Business Administration, Education, Health, Humanities, Law, Natural Science, Social Science, and Technology. We label Agriculture, Business Administration, Health, Natural Science, and Technology as STEM fields. We also consider a narrower set of administrative codes, known as "Futuro Laboral" codes, which divide fields of study into 124 categories. Examples of such codes include Mechanical Engineering and Pedagogy in Elementary Education.

Historically, students were notified of admissions outcomes through a special section of the newspaper. We recover data on admissions outcomes by digitizing the newspaper admissions announcements. We link these data to administrative records of the application process obtained from DEMRE. These records include student gender.

We match our records of college admissions to UI records from 2009 through 2016. These data are maintained by the Ministerio del Trabajo y Prevision Social. This dataset contains matched employer-employee earnings data at the monthly level. These earnings records are subject to three important limitations. First, the data do not include information on hours or weeks worked, so we cannot measure wages. We limit the analysis to earnings observations above half the 2014 full-time minimum wage. In a robustness check described in Section V, we estimate firm pay premiums using only men's earnings to reduce variability in hours worked. Second, monthly earnings are topcoded at the UI maximum. About 3% of earnings observations are topcoded, including 17% among college admits, who are the focus our study. In a robustness check, we repeat our analysis using imputed earnings for topcoded observations as in Card et al. (2013). We transform the monthly earnings data into a panel of job spells, where we average earnings in each job spell at the annual level. More details on the construction of the data are provided in Online Appendix A. Third, the data exclude the public sector.[1]

We limit our analysis to workers between the ages of 20 and 60. We use all workers in that age range to estimate worker- and firm-specific pay premiums. Our main analysis focuses on students admitted to a college program between 1977 to 2003.

[1] This is a common limitation in the literature on the gender gap in firm pay premiums (e.g., Card et al., 2016).

III Empirical Framework

Our baseline approach relies on a model in which log earnings are a function of additive worker and firm fixed effects (AKM). Monthly earnings w_{it} for worker i at time t are given by

$$\log w_{it} = \alpha_i + \psi_{J(i,t)} + X_{it}'\beta + \epsilon_{it}, \tag{1}$$

where α_i is a worker fixed effect (the *worker effect*), $\psi_{J(i,t)}$ is a firm fixed effect (the *firm effect*), and $J(i,t)$ is a function indicating the employer of worker i at time t. X_{it} is a vector of time-varying controls, including time effects and controls for worker age. The residual ϵ_{it} captures time-varying shocks to wages, including worker-job specific match effects, shocks to human capital, and other factors. Following CHK, in X_{it} we include a third-order polynomial in age and restrict the age profile to be flat at age 40 by omitting the linear age term and re-centering age at 40. We estimate equation (1) using OLS within the largest 'connected set' of firms, i.e., the largest set of firms that can be linked by a path of worker firm-to-firm movements. This connected set includes 98% of all monthly earnings observations.

We estimate over 7 million worker effects and 480,000 firm effects. The standard deviation of worker effects is 0.474, about 50% larger than the standard deviation of firm effects. Higher-paid workers tend to work in higher-paying firms: the correlation between estimated worker effects and firm effects is 0.287. Appendix Table B1 describes the OLS estimates of equation (1) in more detail.

For our approach to yield unbiased estimates for worker and firm effects, the following *exogenous mobility* condition must hold (Card et al., 2018):

$$E[(\epsilon_{it} - \bar{\epsilon})(D_{it}^j - \bar{D}_i^j)] = 0 \ \forall j,$$

where $D_{it}^j \equiv 1_{J(i,t)=j}$ is an indicator for employment at firm j in period t. In words, worker mobility must be uncorrelated with the time-varying residual component of wages. This means that workers that switch firms do not sort based on match effects or worker-specific shocks to wages. We follow the now-standard specification checks developed in CHK and CCK and document empirical evidence consistent with this condition. Job-to-job transitions are associated with abrupt earnings changes, the magnitude of which varies symmetrically and one-to-one with the change in estimated firm effect. This holds for the full sample of all workers, for college admits only, and for both STEM and non-STEM majors. See Online Appendix B for a discussion of this validation exercise.

The AKM model permits a straightforward wage decomposition, where the variance of log wages can be written as

$$\begin{aligned}\text{Var}(\log w_{it}) = &\underbrace{\text{Var}(\alpha_i + X_{it}'\beta)}_{\text{worker component}} + \underbrace{\text{Var}(\psi_{J(i,t)})}_{\text{firm component}} + \underbrace{\text{Var}(\epsilon_{it})}_{\text{residual component}} \\ &+ \underbrace{2\text{Cov}(\alpha_i + X_{it}'\beta, \psi_{J(i,t)})}_{\text{covariance component}}.\end{aligned} \tag{2}$$

The covariance between worker and firm components reflects the degree of assortative matching. If high wage workers are more likely to work at high wage firms, then this covariance term will be positive. Below we will express these variance and covariance terms scaled by $\text{Var}(\log w_{it})$, which can be interpreted as the share of wage variation explained by some component. For example, $\frac{2\text{Cov}(\alpha_i + X'_{it}\beta, \psi_{J(i,t)})}{\text{Var}(\log w_{it})}$ is the share of wage variation attributed to sorting.

The correlation between estimated worker and firm effects, which measures assortative matching between workers and firms, is an object of particular interest in our analysis. These correlations are subject to *limited mobility bias*: they are biased downward in finite samples and the size of the bias is inversely related to the degree of worker mobility among firms (Abowd et al., 2004; Andrews et al., 2008). We address this issue in two ways. First, we evaluate the severity of limited mobility bias in our setting using the approach of Kline et al. (2020) and the the split-sample approach described in Gerard et al. (2018). Second, we consider an alternative wage decomposition based on a firm clustering approach Bonhomme et al. (2019); Lamadon et al. (2019). Interested readers can find a description of each approach and the results in Online Appendix C. In practice, we find that limited mobility bias is negligible in our setting and we find very similar results under these alternative approaches.[2]

Equation (1) restricts firm effects to be the same for all workers. This is a useful benchmark, but it is possible that firm effects may vary by worker type. For example, CCK find that firm pay premiums are smaller for women. In Section IV.C, we consider an augmented model that allows firm pay premiums to vary by gender and field of study.

IV Results

IV.A Earnings Outcomes and Firm Sorting by Field of Study

In this section we describe earnings outcomes and firm sorting by field of study. We compare how firm sorting patterns among workers from the same field of study compare to sorting patterns between fields of study.

The top row of Table I decomposes the earnings of college admits using equation (2). Overall, we find that about 56% of the variance in log earnings among college admits can be attributed to the worker component, 15% to the firm component, and 15% to worker-firm sorting, with 13% remaining as a residual. The results of this decomposition are similar to decompositions for Germany (CHK) and Portugal (CCK) in the 2000s.

We then further decompose variation in earnings into between- and within-major components. We calculate the total share of the variance of earnings that is within major and between major, and the total share of the covariance between worker and firm components that is within major

[2]The importance of limited mobility bias varies across settings (Bonhomme et al., 2020). For example, Lachowska et al. (2019) find limited bias in data from Washington state.

and between major. We calculate the former using the Law of Total Variance, as

$$Var[\log w_{it}] = \underbrace{E_m[Var(\log w_{it}|m(i))]}_{\text{within major}} + \underbrace{Var_m(E[\log w_{it}|m(i)])}_{\text{between major}}$$

and the latter using an analogous decomposition as

$$\begin{aligned} Cov(\alpha_i + X_{it}'\beta, \psi_{J(i,t)}) = & \underbrace{E_m[Cov(\alpha_i + X_{it}'\beta, \psi_{J(i,t)}|m(i))]}_{\text{within major}} \\ & + \underbrace{Cov_m(E[\alpha_i + X_{it}'\beta|m(i)], E[\psi_{J(i,t)}|m(i)])}_{\text{between major}}. \end{aligned}$$

We now come to our first major result. We find that assortative matching is markedly stronger *between* fields of study than *within* fields of study. This can be seen in the share of earnings variation explained by the covariance of worker and firm effects. This share is 13% within majors. By contrast, assortative matching accounts for 42.0% of variation in earnings across majors. Major accounts for only 8% of variance in earnings, but accounts for 22% of the sorting component.

Table I also reports means of the following objects for each field of study: log earnings, the worker component, and the firm component. We demean each outcome, so that reported means for each major are differences from the average major-specific wage, firm, or individual effect premium. We also report the shares of male and female students enrolling in each major, and the ratio of the major-specific mean firm effect to the major-specific earnings premium. Finally, we decompose the variance of earnings as described in Section III into components due to variation in worker components, firm components, the covariance between the worker and firm components, and a residual within each major.

[Table 1 about here.]

Consistent with prior work (ABM, AAM, HNZ), we see the highest earnings among workers from STEM fields and the lowest earnings among those studying Education and humanities. Technology majors have log earnings 17.8 log points higher than average, while Education and Humanities majors have earnings 30.0 and 39.2 log points lower than average, respectively. Technology and Education are by far the largest coarse major categories, and men are relatively overrepresented in the former category and underrepresented in the latter. 54.5 percent of all college admitted men are admitted to Technology degrees, compared to 21.7 percent of women; for Education, these shares are 7.4% and 24.8% respectively.

The finding that assortative matching is stronger between fields of study than within fields of study is driven in large part by the popular Technology and Education fields. The mean firm premium for Technology majors is 0.073, 41% of the 0.178 earnings premium for Technology majors relative to the average major. For Education majors, the mean firm premium is -0.131, 33% of the overall major effect. Agriculture, Art, Social Science, and the Humanities follow similar patterns.

In contrast, negative average firm effects for law and health majors partially offset positive average worker effects, and in Business the mean worker component is positive but the firm component is close to zero.

IV.B Field of Study and the Gender Gap in Firm Effects

Our finding that firm premiums account for large shares of the overall earnings premiums in male dominated majors like Technology suggests that college major may play an important role in mediating the firm contribution to the gender gap in earnings. Table II address this question. In the upper two panels, we decompose the gender earnings gap into worker and firm components. In our sample, the overall gender earnings gap is 19.9 log points. The gaps in firm and worker components are 5.7 log points and 15.4 log points. Among college admits, the gender earnings gap is 25.2 log points, with a 8.1 log point gap in the firm component and 16.8 log point gap in the worker component. The firm component of the gender gap is thus 29% of the total gap in the full sample 28% in the non-college sample, and 32% in the sample of college admits. These findings are both quantitatively similar to CCK, who report firm shares of the gender gap as 20%, and also qualitatively similar in the sense that firm contribution to the gender gap is relatively stable across education *levels*.

We further decompose the gender earnings gap among college admits by major using the following decomposition (Kitagawa, 1955; Duncan, 1969; Oaxaca, 1973; Blinder, 1973):

$$\begin{aligned} E[Y_{it}|male] - E[Y_{it}|female] = &\underbrace{E_m[E[Y_{it}|m(i)]|female] - E_m[E[Y_{it}|m(i)]|male]}_{\text{between major}} \\ &+ E_m[E[Y_{it}|m(i); male]] - E_m[E[Y_{it}|m(i)]|male] \\ &- (E_m[E[Y_{it}|m(i); female]] - E_m[E[Y_{it}|m(i)]|female]) \end{aligned} \tag{3}$$

where Y_{it} is either log wages, the worker component, or the firm component and we refer to the remaining terms as the 'residual'.

[Table 2 about here.]

We find that the between-major gender earnings gap is 13.4 points and the residual gender earnings gap is 11.8 log points. That is, differences in average earnings by major explain 53% of the gender gap among college admits. The cross-major component of the gender gap is even larger—0.163 log points, or 65%—when we consider narrow major categories.

This brings us to our second main finding, which is that the contribution of majors to the gender earnings gap comes disproportionately through the *firm* component of wages, and, further, that the firm component of the gender gap is mostly a between-major phenomenon. The between-major component of the firm gap is 0.051 when we use coarse major categories, and 0.064 when we use the finer categories. These values are equal to 63% and 79% of the firm component of the gender gap, respectively. In contrast, broad and narrow major categories explain, respectively, 47% and 58% of the worker component of the gender gap.

Another way to frame this is to observe that the firm component accounts for 38% of the between-major area gender earnings gap, but only 25% of the residual within-major gender earnings gap. The differences are larger still when we replace major area with *specific major* in the decomposition. The firm component accounts for 39% of the between-specific-major gender earnings gap, but only 20% of the residual gender earnings gap.

The bottom panel of Table II reports earnings gaps and mean firm and worker contributions to those gaps for specific broad majors. The male-female earnings gap is largest in STEM fields of Agriculture, Natural Science, and Technology, where it is equal to roughly 0.2. It is also fairly large in Business (0.146). The firm contribution to the gap is largest in Business and Natural Science, but near zero in Business and Health.

One possible explanation for the finding that majors explain a large share of gender differences in firm premiums is that they simply capture vertical differences in worker skill, and high-skill workers match assortatively into high-skill firms. The intuition is similar to Gerard et al. (2018), who consider the hypothesis that racial differences in sorting may be attributable to race-neutral sorting on worker effects. They distinguish between sorting on worker effects and "residual sorting" that persists after controlling for worker effects.

This does not prove to be the case in our setting: majors retain their explanatory power for firm sorting even after conditioning on worker effects. We illustrate this using a regression approach. We estimate regression models of the form

$$Y_{it} = \alpha + \beta \text{Male}_i + X_{it}\delta + \mu_i^{major} + \epsilon_{it}, \tag{4}$$

where Y_{it} are earnings outcomes (log earnings, firm effects, worker effects), X_{it} is a vector of individual characteristics that varies across specifications, and μ_i^{major} are major-specific fixed effects that we include in some specifications.

We present our results in Table III. In the top row the outcome is log earnings; in the middle row the outcome is the firm component; and in the bottom row the outcome is the worker component. All specifications include a full set of indicators for potential work experience (years since last admissions exam). Column (1) does not include any additional controls. We find a gender earnings gap of 22.8 log points of which 8.2 log points (37%) are attributable to differences in firm sorting. In column (2) we add controls for math and verbal exam scores. Adding these controls reduces the earnings gap by 37% to 14.3 log points, and the firm contribution to the gap by 22%, to 6.5 log points. In column (3) we control for a four-piece spline in the estimated worker effect, $\hat{\alpha}_i$. We find a similar residual difference in firm effects to what we saw for test score controls: the firm contribution to the gender gap declines by 24%, to 6.3 log points. After accounting for assortative matching—the covariance between worker and firm effects—most of the firm contribution to the gender gap remains.

Major explains a large share of the sorting that remains. In column (4) we remove controls for exam scores and include major area fixed effects. Controls for major reduce the firm sorting component by 57%, to 3.6 log points. Column (5) includes controls for both exam scores and major. Comparing to column (2), which included controls for exam scores alone, adding college major reduces the firm sorting component by 52%, to 3.1 log points. Again, we find a similar pattern if we replace exam scores with estimated worker effects as controls. Column (6) includes controls for both worker effects and major. Compared to column (3), adding college major reduces the firm sorting component by 56%, to 2.7 log points. Finally, in column (7) we include specific major fixed effects. These controls decrease the earnings gap to 6.9 log points and the unexplained sorting effect to 2.1 log points, 67% less than the specification that includes exam score controls only. College major explains between half and two-thirds of the residual gender gap in firm effects.

[Table 3 about here.]

IV.C Sorting Versus Bargaining Channels

Our analysis thus far constrains firm pay premiums to be constant across workers. However, the pay premium associated with a particular firm may depend on worker gender as in CCK as well as the *major* of the worker in question. We extend our baseline model to account for these possibilities. This extension also allows us to decompose the firm component of the gender gap into what CCK label *sorting* and *bargaining* terms. The former reflects the firm component of the gap if all workers received the male (or female) pay premium for the firm; the latter reflects within-firm differences in pay premiums for men and women.

The more flexible wage model can be written as

$$\log w_{it} = \alpha_i + \psi_{J(i,t)}^{m(i)g(i)} + X_{it}'\beta + \epsilon_{it} \tag{5}$$

where $m(i)$ indexes the major of worker i. There are insufficient worker flows per firm to estimate equation (5) with fully saturated major by gender by firm fixed effects. Instead, we allow for major-gender-sector interactions. We define $g(i)$ to map workers into three categories: non-college admits ('0'); male college admits ('M'); and female college admits ('F'). Concretely, we specify $\psi_{J(i,t)}^{m(i)g(i)}$ as

$$\psi_{J(i,t)}^{m(i)g(i)} = \psi_{J(i,t)}^{0} + \sigma_{s(J(i,t))m(i)g(i))} \tag{6}$$

where $\psi_{J(i,t)}^{0}$ is the firm effect for workers with $m(i) = g(i) = 0$ (no assigned major) and $s(J(i,t))$ maps firm J to one of 11 sectors. Hence, we are allowing for sector by gender by major specific interactions. Following Card et al. (2016), we constrain $\sigma_{s(J(i,t))m(i)g(i)}$ to be mean zero within the restaurant and hotel sector for all major/gender combinations. We refer to this as the *augmented* AKM model.

Following CCK we use equation (5) to decompose the gender difference in firm pay premiums among college admits into a combination of *bargaining* effects and *sorting* effects:

$$
\begin{aligned}
E\left[\psi^{m(i)M}_{J(i,t)}\middle|male\right] - E\left[\psi^{m(i)F}_{J(i,t)}\middle|female\right] =& E\left[\tilde{\psi}^{M}_{J(i,t)}\middle|male\right] - E\left[\tilde{\psi}^{F}_{J(i,t)}\middle|female\right] \\
=& E\left[\tilde{\psi}^{M}_{J(i,t)} - \tilde{\psi}^{F}_{J(i,t)}\middle|female\right] \\
&+ E\left[\tilde{\psi}^{M}_{J(i,t)}\middle|male\right] - E\left[\tilde{\psi}^{M}_{J(i,t)}\middle|female\right] \qquad (7)
\end{aligned}
$$

where

$$
\tilde{\psi}^{M}_{J(i,t)} = \sum_k \alpha^{M,k}_{s(J(i,t))} \psi^{kM}_{J(i,t)} \qquad (8)
$$

and $\alpha^{M,k}_s = P(m(i) = k|male; s(J(i,t)) = s)$. $\tilde{\psi}^M_j$ is a weighted average of $\psi^{m(i)M}_j$ across majors, where weights are determined by the major distribution of men working in sector $s(j)$. Put simply, $\tilde{\psi}^M_j$ is the average pay premium at firm j for men employed in sector $s(j)$. $\tilde{\psi}^F_{J(i,t)}$ is defined analogously for women.

The first term in equation (7) is the average bargaining effect, calculated by comparing $\tilde{\psi}^M_j$ and $\tilde{\psi}^F_j$ across the distribution of jobs held by women. The second line of equation (7) gives the average sorting effect, calculated by comparing the average value of $\tilde{\psi}^M_j$ across the jobs held by women versus men. We also conduct an analogous decomposition where we compute the bargaining effect using the distribution of jobs held by men and the sorting effect using $\tilde{\psi}^F_j$.

In Table IV we decompose the gender earnings gap into a worker and firm component using the augmented model. The average male firm premium among men is 32 log points, and the average female firm premium among women is 23.8 log points. The total firm contribution to the gender gap in the augmented model is the difference between these numbers: 8.2 log points. This is nearly identical to the 8.1 log point contribution we observed in the baseline model. Most of the firm contribution to the gender gap comes through the sorting channel. Sorting accounts for 7.0 log points (85%) of the firm gender gap when computed using firm effects for men, and 6.6 log points (80%) when computed using firm effects for women.

Also as in the baseline AKM model, cross-major variation accounts for most of the firm component of the gender gap. Out of the 8.2 log point gap in firm effects, 4.2 log points (51%) comes from variation between broad majors and 5.4 log points (66%) comes from variation between specific majors. All of this is through the sorting channel. Measured using male firm effects, major explains 4.3 log points (62%) of the sorting component of the firm gender gap, while the cross-major component of the bargaining share is slightly negative.

We also decompose the gender earnings gap within major. For this exercise we use a different decomposition to limit the comparisons to men and women in the same major. For fixed major m, we decompose the gender earnings gap as follows:

$$
\begin{aligned}
E\left[\psi^{mF}_{J(i,t)}\middle|female, m\right] - E\left[\psi^{mM}_{J(i,t)}\middle|male, m\right] =& E\left[\psi^{mF}_{J(i,t)} - \psi^{mM}_{J(i,t)}\middle|female, m\right] \\
&+ E\left[\psi^{mM}_{J(i,t)}\middle|female, m\right] - E\left[\psi^{mM}_{J(i,t)}\middle|male, m\right]. \qquad (9)
\end{aligned}
$$

The bargaining effect in equation (9) is the average gender difference in firm pay premiums among m majors across the distribution of jobs held by women from major m. The sorting effect is the difference between the average pay premium earned by men from major m across the jobs held by women and the same average across jobs held by men.

As in the baseline model, the firm contribution to the gender gap is large in Technology and Natural Science. Sorting components in the augmented model are similar those we observe in the baseline model, with the total firm effect growing due to bargaining components in Natural Science. In contrast to the baseline model, the augmented model suggests substantial firm components to gender gap in Law, Humanities, and Art/Architecture, operating through the bargaining channel. These fields account for a relatively small share of college admits students– 6.4% of men and 10.3% of women, combined. Bargaining effects are close to zero in Technology and Business and negative in other fields.

We draw two conclusions from the augmented model. First, the baseline model does a good job capturing the sorting component of the firm contribution to gender gaps and its interactions with field of study. Using a more flexible model does not change the conclusions we draw. Second, though the bargaining channel is important in some fields, the explanatory power of major for the firm component of the gender gap comes mostly through the sorting channel.

[Table 4 about here.]

V Additional Analyses and Robustness Tests

Online Appendix C addresses possible concerns about our data and estimation approach. We evaluate and account for limited mobility bias using approaches developed in Kline et al. (2020), Gerard et al. (2018), Bonhomme et al. (2019), and Lamadon et al. (2019). We impute topcoded earnings to account for topcoding in the earnings data (Card et al., 2013) and we estimate firm effects using only male workers to reduce variation in hours worked. Imputing topcoded earnings increases the gender pay gap among college admits from 25 log points to 32 log points, with most of that increase coming from the worker component. None of these analyses alter our main findings. We also consider the role of industry in mediating major effects.

VI Conclusion

This paper used worker-firm data matched to education records from Chile to describe how the returns to major depend on how workers match to firms, and how this affects our interpretation of the firm component of the gender earnings gap. We show that 42% of earnings variation across majors is due to the fact that students in high-paying majors match assortatively to high-paying firms. For students in Technology majors, the vast majority of whom are male, firm premiums account for 41% of the mean earnings premium relative to the average over all majors. Overall, differences in field of study explain the majority– roughly 60-80%, depending on how we measure–

of the firm component of the gender gap in earnings; the explanatory power of major for the firm component of the gender gap persists even after controlling for measures of individual skill such as test scores or estimated worker effects.

Our findings suggest several avenues for future work. First, they indicate that research seeking to understand where the *private* returns to field of study come from may want to investigate how students find jobs in addition to, or as a function of, the skills they learn in the classroom. Second, if one is willing to interpret firm premiums as reflecting rents to firm match, as in CCHK, our results raise the possibility that there exists a substantial wedge between the public and private returns to non-health STEM fields. One benefit of programs aimed at increasing female STEM representation through recruitment, mentorship, or anti-discrimination policies within the education system may be to reduce the firm component of the gender earnings gap.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–334.

_ , _ , **Paul A. Lengermann, and Sébasten Pérez-Duarte**, "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States," 2004. Unpublished manuscript, Center de Recherche en Economie et en Statistique.

Addario, Sabrina Di, Patrick Kline, Raffaele Saggio, and Mikkel Solvsten, "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity and wages," August 2020. IRLE Working Paper 104-20.

Altonji, Joseph G. and Seth D. Zimmerman, "The Costs of and Net Returns to College Major," in "Productivity in Higher Education" NBER Chapters, National Bureau of Economic Research, Inc, May 2018, pp. 133–176.

_ , **Erica Blom, and Costas Meghir**, "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, 2012, *4*, 185–223.

_ , **Peter Arcidiacono, and Arnaud Maurel**, "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016, chapter 7, pp. 305–396.

Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward, "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?," *Journal of the Royal Statistical Society: Series A*, 2008, *171* (3), 673–697.

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 1973, *8* (4), 436–455.

Bonhomme, Stephane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler, "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?," August 2020. Unpublished manuscript.

_ , **Thibaut Lamadon, and Elena Manresa**, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, May 2019, *87* (3), 699–738.

Bruns, Benjamin, "Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap," *American Economic Journal: Applied Economics*, April 2019, *11* (2), 74–113.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, competitiveness, and career choices," *The Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Card, David, Ana Rute Cardoso, and Patrick Kline, "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women," *Quarterly Journal of Economics*, May 2016, *131* (2), 633–686.

_ , _ , Joerg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, January 2018, *36* (S1), S13–S70.

_ , Jorg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, August 2013, *128* (3), 967–1015.

Carrell, Scott E, Marianne E Page, and James E West, "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Casarico, Alessandra and Salvatore Lattanzio, "What Firms Do: Gender Inequality in Linked Employer-Employee Data," June 2019. Cambridge Working Papers in Economics: 1966.

Duncan, Otis D., "Inheritance of Poverty or Inheritance of Race," in Daniel P. Moynihan, ed., *On Understanding Poverty: Perspectives from the Social Sciences*, Basic Books, 1969, pp. 85–110.

Gerard, Francis, Lorenzo Lagos, Edson Severini, and David Card, "Assortative Matching or Exclusionary Hiring? The Impact of Firm Policies on Racial Wage Differences in Brazil," October 2018. Unpublished manuscript.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman, "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," Technical Report, National Bureau of Economic Research 2013.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad, "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, August 2016, *131* (3), 1057–1111.

Kitagawa, Evelyn M., "Components of a Difference Between Two Rates," *Journal of the American Statistical Association*, 1955, *50* (272), 1168–1194.

Kline, Patrick, Mikkel Sølvsten, and Raffaele Saggio, "Leave Out Estimation of Variance Components," *Econometrica*, 2020.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury, "Do firm effects drift? Evidence from Washington Administrative," November 2019. Unpublished manuscript.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler, "Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market," May 2019. Unpublished manuscript.

Morchio, Iacopo and Christian Moser, "The Gender Pay Gap: Micro Sources and Macro Consequences," April 2020. Unpublished manuscript.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 1973, *14*, 693–709.

Paredes, Valentina A, M Daniele Paserman, and Francisco Pino, "Does Economics Make You Sexist?," Technical Report, National Bureau of Economic Research 2020.

Price, Joshua, "The effect of instructor race and gender on student persistence in STEM fields," *Economics of Education Review*, 2010, *29* (6), 901–910.

Rodriguez, Jorge, Sergio Urzua, and Loreto Reyes, "Heterogeneous Economic Returns to Post-Secondary Degrees: Evidence from Chile," *Journal of Human Resources*, 2016, *51* (2), 416–460.

Sloane, Carolyn, Erik Hurst, and Dan Black, "A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap," October 2019. Unpublished manuscript.

Sorkin, Isaac, "The Role of Firms in Gender Earnings Inequality: Evidence from the United States," *American Economic Review: Papers and Proceedings*, 2017, *107* (5), 384–387.

Stange, Kevin, "Differential pricing in undergraduate education: Effects on degree production by field," *Journal of Policy Analysis and Management*, 2015, *34* (1), 107–135.

U.S. Citizenship and Immigration Services, "Optional Practical Training Extension for STEM Students (STEM OPT)," 2020.

Zafar, Basit, "College major choice and the gender gap," *Journal of Human Resources*, 2013, *48* (3), 545–595.

TABLE I
EARNINGS OUTCOMES BY MAJOR

							Variance Decomposition			
Major	Male Share	Female Share	Log Earnings	Worker Component	Firm Component	Firm Ratio	Worker Share	Firm Share	Covariance Share	Residual Share
Overall							0.559	0.154	0.154	0.133
Between Major							0.429	0.125	0.420	0.026
Within Major							0.572	0.156	0.130	0.142
Business	11.4	14.6	0.109	0.093	0.014	0.131	0.600	0.144	0.126	0.130
Agriculture	9.3	6.7	-0.115	-0.052	-0.057	0.502	0.611	0.143	0.102	0.144
Architecture and Art	2.2	3.2	-0.242	-0.155	-0.079	0.325	0.574	0.164	0.110	0.152
Natural Science	5.4	7.2	-0.079	-0.077	-0.000	0.005	0.565	0.150	0.150	0.135
Social Science	2.3	5.5	-0.186	-0.113	-0.070	0.378	0.585	0.142	0.125	0.148
Law	2.3	2.4	-0.027	0.028	-0.042	1.546	0.582	0.133	0.156	0.129
Education	7.4	24.8	-0.392	-0.260	-0.131	0.333	0.602	0.148	0.101	0.149
Humanities	1.9	4.7	-0.300	-0.229	-0.066	0.220	0.590	0.144	0.115	0.151
Health	3.3	9.3	0.074	0.100	-0.029	-0.389	0.620	0.122	0.105	0.153
Technology	54.5	21.7	0.178	0.102	0.073	0.419	0.536	0.174	0.148	0.142

Firm effects and worker effects are estimates from equation (1), which is described in more detail in Section III. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in III. The between-field decomposition weights fields by their number of earnings observations. 'Male Share' and 'Female Share' refer to the percentage of male and female college admits in each major. 'Firm Ratio' is the ratio of the demeaned firm component to demeaned log earnings.

Table II
Contribution of Firm-Specific Pay Premiums and
Major to the Gender Earnings Gap

	Gender Earnings Gap	Firm Component	Worker Component
Overall	0.199	0.057	0.154
		(0.286)	(0.774)
Non-College Admits	0.202	0.057	0.158
		(0.282)	(0.782)
College Admits	0.252	0.081	0.168
		(0.321)	(0.667)
Between Major Area	0.134	0.051	0.080
		(0.381)	(0.597)
Residual	0.118	0.030	0.088
		(0.254)	(0.746)
By Major Area:			
Business	0.146	0.006	0.141
Agriculture	0.221	0.032	0.182
Architecture and Art	0.065	0.016	0.047
Natural Science	0.190	0.054	0.135
Social Science	0.049	0.014	0.030
Law	0.048	0.006	0.026
Education	0.029	0.036	0.003
Humanities	-0.007	-0.006	0.002
Health	0.041	0.017	0.020
Technology	0.210	0.059	0.149
Between Specific Major	0.163	0.064	0.097
		(0.393)	(0.595)
Residual	0.089	0.018	0.071
		(0.202)	(0.798)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker earnings components as described in Section IV.B. Column 1 reports the difference between male and female workers in the subset of workers indicated by the row heading. Columns 2 and 3 report gender differences in the worker component, $\alpha_i + X'_{it}\beta$, and firm-specific pay premiums, $\psi_{J(i,t)}$. 'Between Major Area' ('Between Specific Major') and 'Residual' reports for each component of the gender earnings gap the decomposition described in equation (3). Entries in parentheses represent the percent of the overall male-female earnings gap (in column 1) that is explained by the source described in column heading.

Table III
Gender Earnings Gap and Firm Sorting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Log Earnings							
Male	0.228	0.143	0.060	0.114	0.074	0.023	0.069
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Outcome: Firm Effect							
Male	0.082	0.064	0.062	0.036	0.031	0.027	0.021
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome: Worker Effect							
Male	0.149	0.083		0.083	0.049		0.052
	(0.002)	(0.002)		(0.002)	(0.002)		(0.002)
Exam Scores		✓			✓		✓
Worker Effects			✓			✓	
Major Area				✓	✓	✓	
Specific Major							✓

This table presents estimates of equation (4), a regression model for the male-female differences in earnings outcomes. All specifications include a full set of indicators for potential work experience (years since last admissions exam). Columns (3) and (6) include a four-piece spline in the estimated worker effect, $\hat{\alpha}_i$, as controls. There are 2,791,187 observations for 420,435 workers. Standard errors are clustered at the worker level.

Table IV
Augmented Model Decomposition of the Gender Earnings Gap

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Means of Firm Premiums			Decompositions of Contribution of Firm Component			
					Sorting		Bargaining	
	Gender Earnings Gap	Male Premium among Men	Female Premium among Women	Total Contribution of Firm Components	Using Male Effects	Using Female Effects	Using Male Distribution	Using Female Distribution
Overall	0.252	0.320	0.238	0.082	0.070	0.066	0.012	0.016
				(0.325)	(0.278)	(0.262)	(0.048)	(0.063)
Between Major Area	0.134			0.042	0.043	0.040	-0.001	0.002
				(0.313)	(0.328)	(0.291)	(-0.007)	(0.022)
Remainder	0.118			0.040	0.027	0.026	0.013	0.014
				(0.339)	(0.229)	(0.229)	(0.110)	(0.110)
Between Specific Major	0.163			0.054	0.054	0.051	-0.000	0.002
				(0.331)	(0.331)	(0.307)	(-0.000)	(0.018)
Remainder	0.089			0.028	0.016	0.015	0.012	0.014
				(0.314)	(0.180)	(0.180)	(0.135)	(0.146)
By Major Area:								
Business	0.146	0.266	0.257	0.009	0.005	0.004	0.004	0.005
Agriculture	0.221	0.308	0.242	0.066	0.036	0.036	0.030	0.030
Architecture and Art	0.065	0.310	0.207	0.102	0.013	0.011	0.089	0.091
Natural Science	0.190	0.368	0.258	0.110	0.051	0.051	0.059	0.059
Social Science	0.049	0.248	0.269	-0.021	0.006	0.009	-0.027	-0.030
Law	0.048	0.311	0.260	0.051	0.003	0.005	0.048	0.047
Education	0.029	0.184	0.164	0.020	0.025	0.027	-0.004	-0.007
Humanities	-0.007	0.277	0.189	0.088	-0.008	-0.007	0.096	0.095
Health	0.041	0.266	0.284	-0.018	0.019	0.015	-0.036	-0.032
Technology	0.210	0.356	0.287	0.069	0.056	0.056	0.012	0.012

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section IV.C. Column 1 reports the difference between male and female college admits in the subset of workers indicated by the row heading. Columns 2 and 3 report firm-specific pay premiums for male and female workers described in Section IV.C . Column 4 reports the total contribution of firm-specific wage premiums to the gender earnings gap reported in column 1. Columns 5 through 8 report the contributions of sorting and bargaining components to gender earnings gap described in Section IV.C. 'Between Major Area' ('Between Specific Major') and 'Residual' reports for each component of the gender earnings gap the decomposition described in equation (3). Entries in parentheses represent the percent of the overall male-female earnings gap (in column 1) that is explained by the source described in column heading.

A Data Construction

Each observation in the data corresponds to a worker, employer, and month combination. We limit observations to a worker's highest paying job in a given month. To minimize the role of variation in hours worked, we make two restrictions. First, we only include observations where a worker earns at least half the 2014 minimum wage for the month full-time. Second, for workers that transition from one establishment to another, we drop the month corresponding to the transition and the preceding month. We do this to exclude observations where the worker was not employed at the establishment for the full month. We then collapse monthly earnings observations for a worker-establishment combination to the annual level.

B AKM Specification Checks

To validate the AKM wage model, equation (1), in our setting we apply specification checks similar to those developed in Card et al. (2013) and Card et al. (2016). In particular, we document evidence that job-to-job transitions are associated with abrupt earnings gains and losses in a manner consistent with exogenous mobility.

Figure B1 plots wage changes for people moving from firm to firm on the vertical axis against the change in coworker mean wages associated with the move on the horizontal axis. Each point corresponds to a decile of the wage increase (or wage decrease) distribution. Panel A is constructed using all workers that change firms; Panel B is constructed using only (1) workers that are admitted to degree programs, (2) workers that are admitted to STEM fields, and (3) workers that are admitted to non-STEM fields. The relationship we observe is linear and approximately symmetric, consistent with the hypothesis that workers do not sort on match effects. The relationship is similar for both STEM and non-STEM majors.

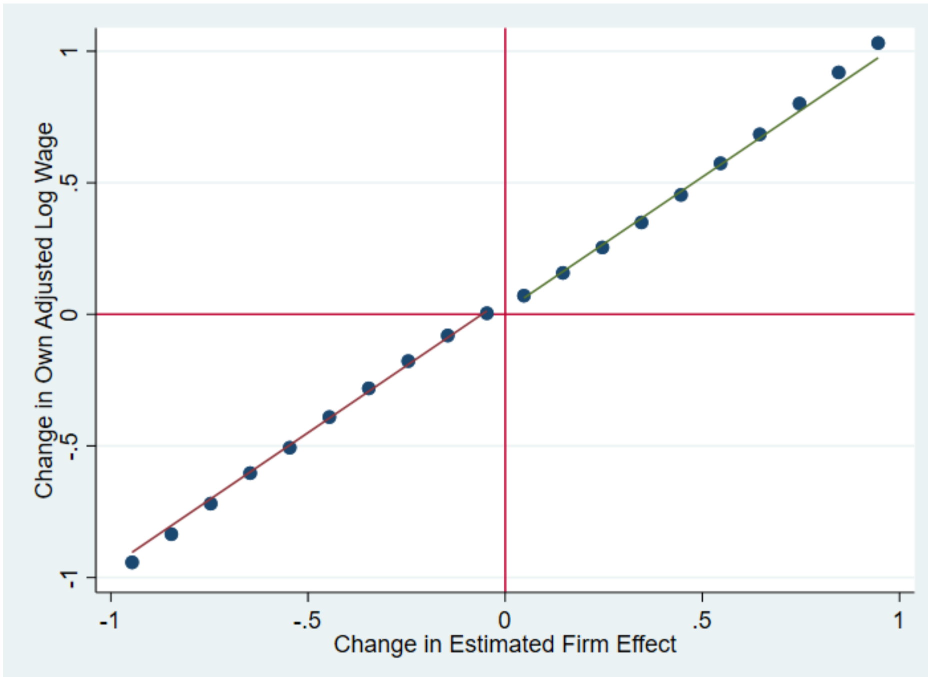
We next conduct a placebo test. We identify a subset of workers who are going to move in a future year but have not yet done so. If earnings trajectories are correlated with changes in firm pay premiums, we would expect some correlate between the change in coworker earnings between the worker's current and future firm, and the change in the worker's pre-move earnings. As shown in Figure B2.

We conduct an analogous exercise where we characterize firms by their estiamted firm effects rather than coworker earnings. Figure B3 plots wage changes for people moving from firm to firm on the vertical axis against the change in coworker mean wages associated with the move on the horizontal axis. Each point corresponds to a decile of the earnings increase (or wage decrease) distribution. Panel A is constructed using all workers that change firms; Panel B is constructed using only workers that are admitted to degree programs. The relationship we observe is linear and approximately symmetric, consistent with the hypothesis that workers do not sort on match effects. The relationship is similar for both STEM and non-STEM majors. An additional prediction of the separable earnings model equation (1) is that earnings should rise (and fall) one-for-one with changes in firm effects. Consistent with this we estimate slopes close to one for each subgroup.

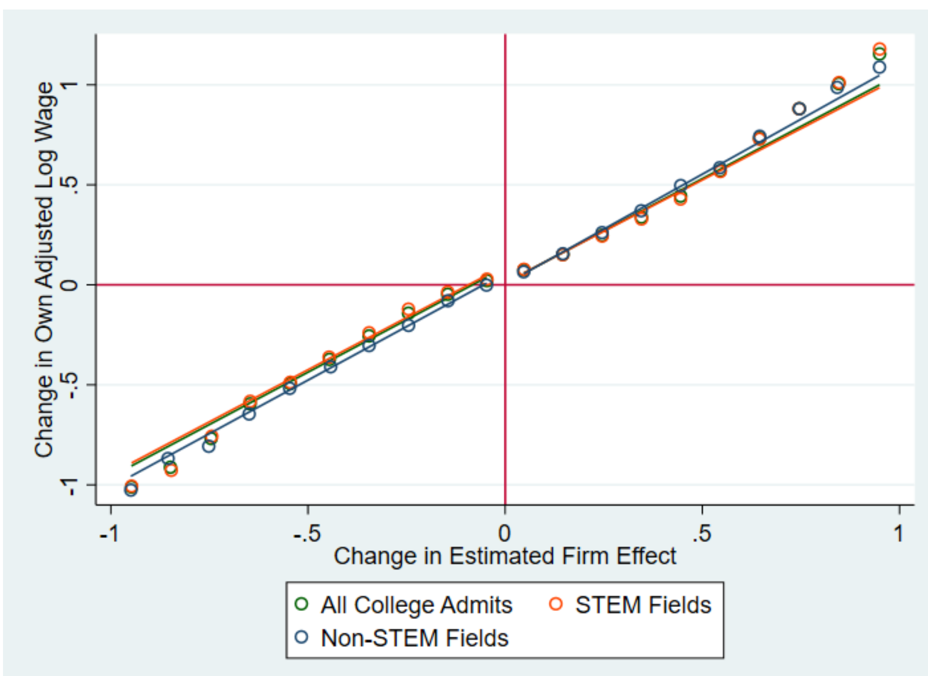
We also conduct a placebo test analogous to Figure B2 in Figure B4. Again, we find no evidence of pre-trends in the earnings of workers that move to higher- or lower-paying firms.

FIGURE B1
EARNINGS CHANGES FOR MOVERS BY COWORKER EARNINGS

(a) All Workers



(b) College Admits

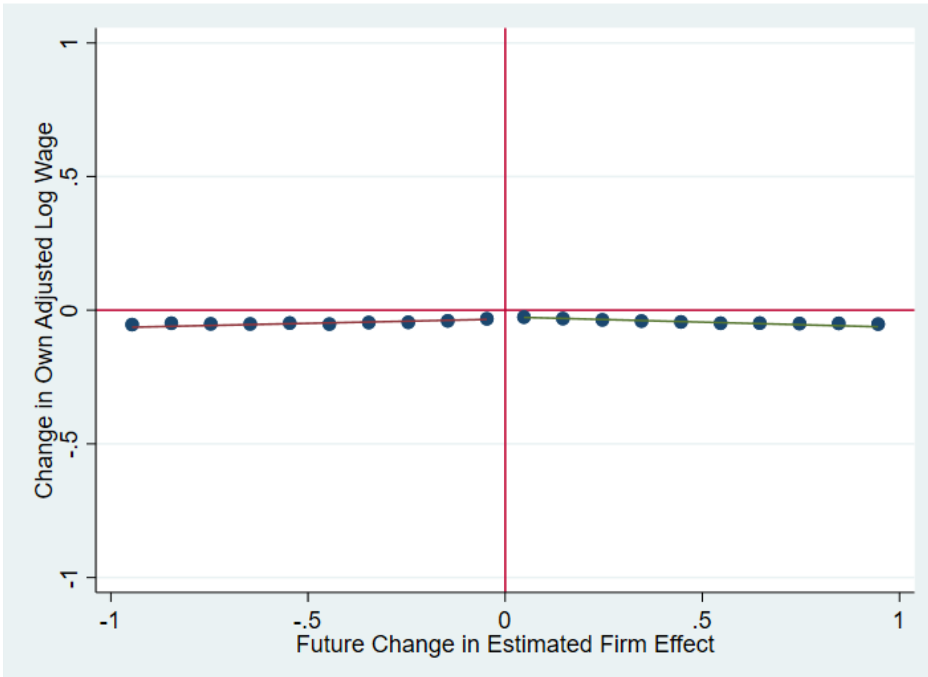


Note: This figure depicts change in own earnings for firm switchers (vertical axis) by difference in coworker earnings at old and new firm (horizontal axis).

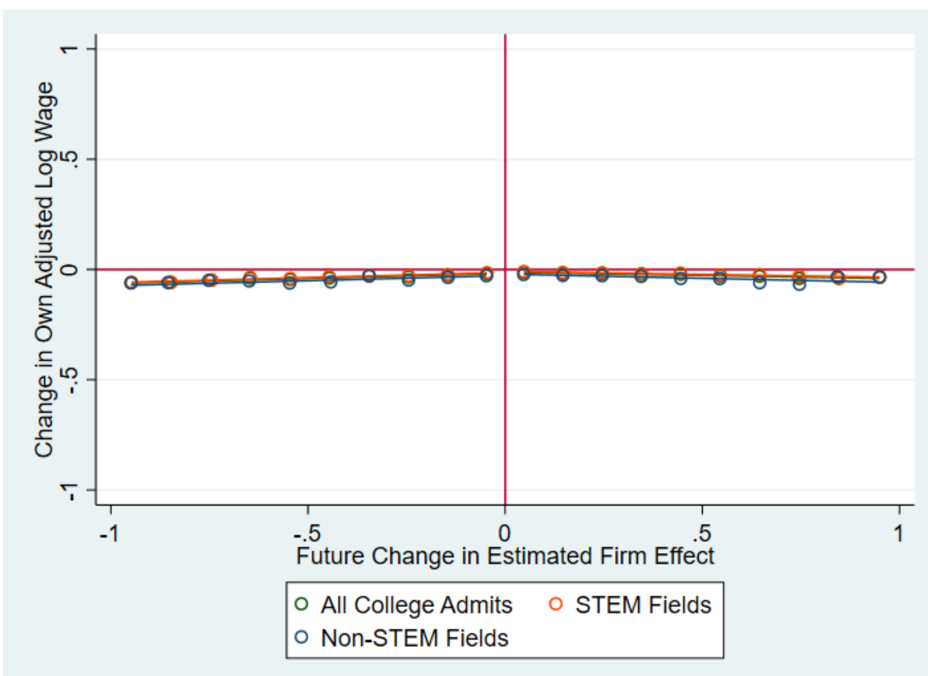
Figure B2

Earnings Changes for Future Movers by Firm Effect

(a) All Workers



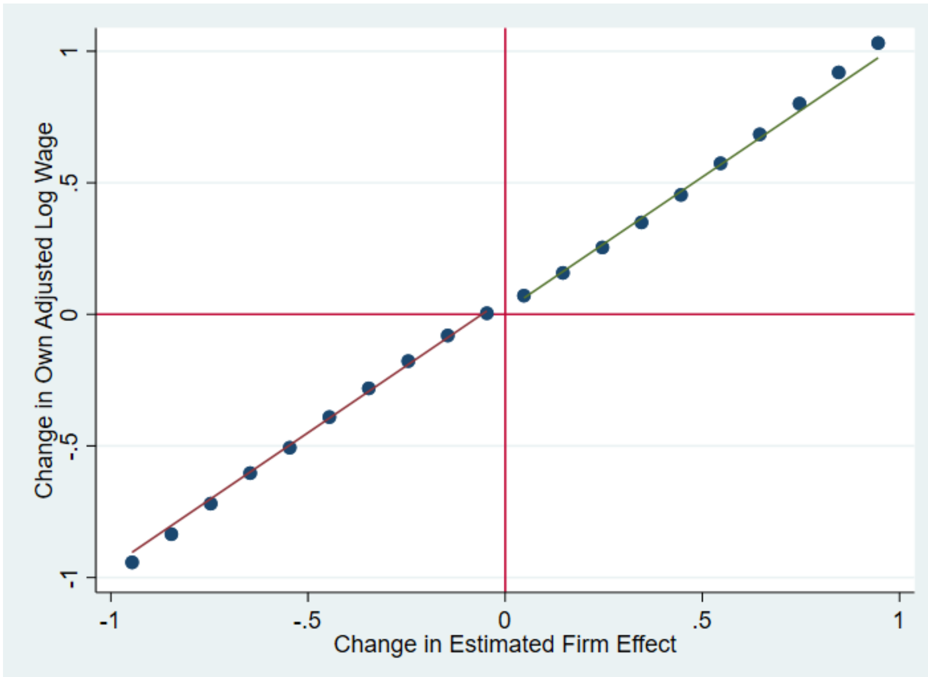
(b) College Admits



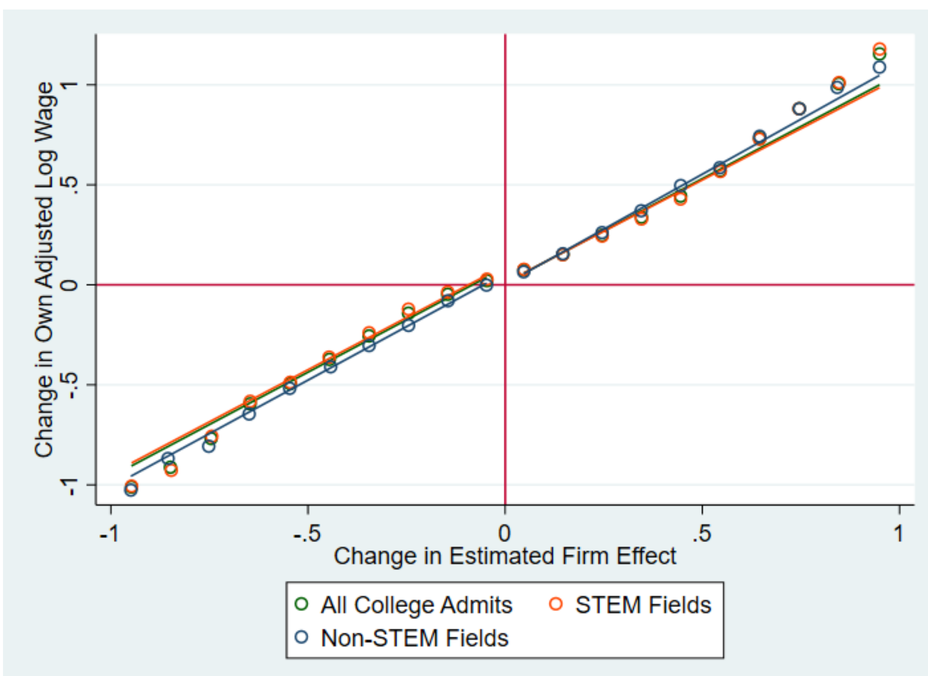
Note: This figure depicts changes in own wages for future firm switchers (vertical axis) by difference in coworker earnings at old and new firm in future switch (horizontal axis).

Figure B3
Earnings Changes for Movers by Firm Effect

(a) All Workers



(b) College Admits

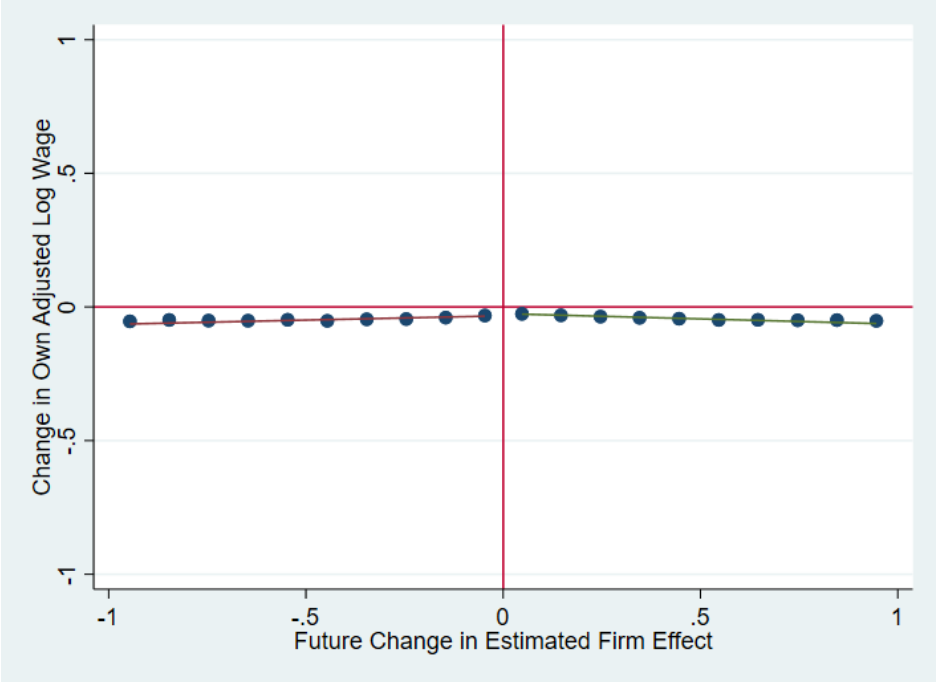


Note: This figure depicts change in own wages for firm switchers (vertical axis) by difference in estimated firm effects at old and new firm (horizontal axis).

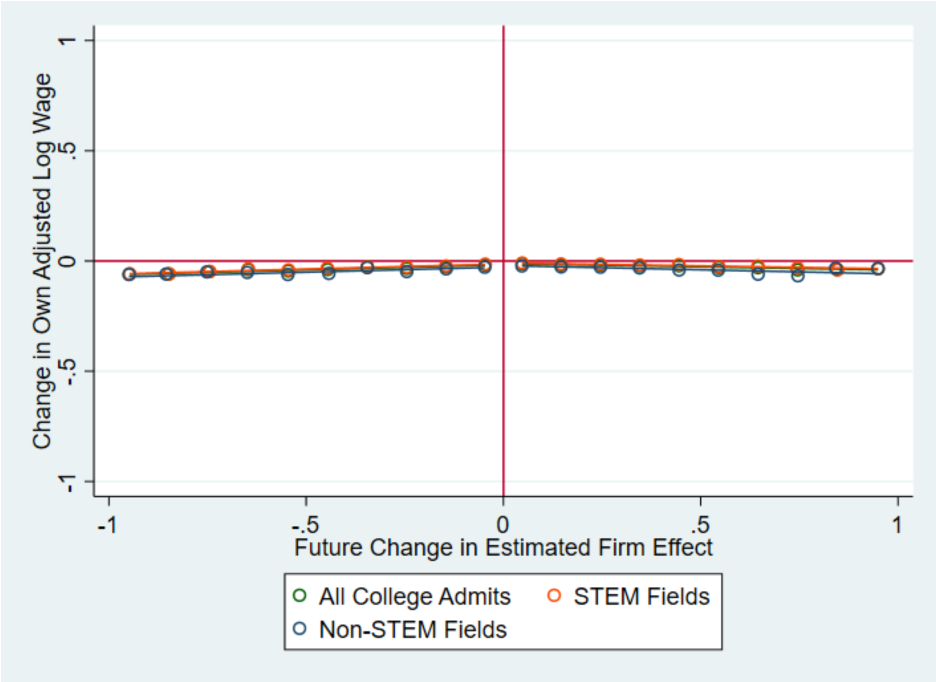
Figure B4

Earnings Changes for Future Movers by Firm Effect

(a) All Workers



(b) College Admits



Note: This figure depicts changes in own wages for future firm switchers (vertical axis) by difference in estimated firm effects at old and new firm in future switch (horizontal axis).

Table B1
Summary of AKM Estimates

	All	College Admits
Worker and firm parameters		
# of worker effects	7,497,531	238,243
# of firm effects	484,185	92,655
Summary of parameter estimates		
Std. dev. of worker effects	0.474	0.508
Std. dev. of firm effects	0.304	0.285
Correlation of worker/firm effects	0.286	0.271
Adjusted R-squared	0.873	
Comparison match model		
Adjusted R^2	0.942	
Std. dev. of match effect	0.162	
Addendum		
Std. dev. log wages		
Sample size	41,753,445	1,615,087

We drop wage observations where monthly wage is half of monthly minimum wage. Limited to largest connected set, which includes 98% of employment.

C Alternative Modeling Approaches and Extensions

C.1 Limited Mobility Bias and Split Sample Estimation

A concern about worker-firm analyses raised in previous research is that estimated correlations between worker and firm effects (as we report in e.g. Table I) are subject to *limited mobility bias*: they are biased downward in finite samples and the size of the bias is inversely related to the degree of worker mobility among firms (Abowd et al., 2004; Andrews et al., 2008). However, we find little evidence of limited mobility bias in our sample. First, we apply the methodology and code of Kline et al. (2020) and find that leave-out estimates of the sorting correlation are essentially identical to the unadjusted correlation.

Second, we estimate a corrected correlation using a split-sample IV design similar to Gerard et al. (2018). We randomly split our sample of workers into two subsamples and estimate equation (1) separately for each subsample. We take our two sets of estimates for ψ_j and β to construct two worker effect estimates for each individual. For each subsample $s \in \{1, 2\}$, we then estimate the regression model

$$\hat{\alpha}_i^s = \lambda_0 + \lambda_1 \hat{\psi}^s_{J(i,t)} + \xi_{it}, \tag{10}$$

instrumenting for $\hat{\psi}_j^s$ using $\hat{\psi}_j^{-s}$, the estimated firm effect from the other subsample.

When we estimate equation (10) using the full data, we get an estimate for λ_1 of 0.457. Using our two-sample approach, which corrects for limited mobility bias, we get estimates of 0.467 and 0.469. Limited mobility bias appears to be quite limited in our data, perhaps because worker mobility is high in Chile. We also find that split sample analogs for the covariance shares reported in Table I are not materially different from the baseline estimates.

C.2 Alternate Approaches to Firm Effect Estimates

Bonhomme et al. (2019) propose an alternative approach to limited mobility bias based on firm clustering. Limited mobility bias emerges when there are many firms with few movers. Bonhomme et al. (2019) address the problem by reducing the dimensionality of the problem and grouping firms into K distinct classes based on the similarity of their earnings distribution. We follow Lamadon et al. (2019) and adapt the AKM specification but replace firm effects with firm *class* effects.

Following Bonhomme et al. (2019) we group firms into K distinct classes based on the similarity of their earnings distribution using k-means clustering. Mathematically, we partition the J firms in the sample into classes solving the following k-means problem (Bonhomme et al., 2019):

$$\min_{k(1),\dots,k(J),H_1,\dots,H_k} \sum_{j=1}^{J} n_j \int \left(\hat{F}_j(y) - H_{k(j)}(y)\right)^2 d\mu(y), \tag{11}$$

where $\hat{F}_j$ denotes the empirical CDF of log earnings in firm j, n_j is the number of workers in firm j, μ is a discrete measure, $k(1), ..., k(J)$ denotes a partition of firms into K classes, and $H_1, ..., H_k$ are CDFs. We first divide firms into 10 industries, and then set the number of classes within each industry at $K = 10$.[3] We then estimate equation (1), but we replace firm fixed effects with firm class fixed effects, $\widetilde{\psi}_{C(i,t)}$.

Analogs of each of the main tables using this approach are provided here. As in our main analysis, we find that (a) covariance between firm and worker effects is much stronger between than within majors (Table C1), (b) the firm component of the gender earnings gap is mostly a between-major phenomenon, with major accounting for between 66% and 77% of the firm component of the gender gap (Table C2), and (c) that major retains substantial explanatory power for the firm component of the gender gap even conditional on measures of worker ability (Table C3).

We also incorporate firm clustering in an alternative to our augmented AKM model equation (5). In this specification we use an alternative classification of gender-major specific firm effects, $\psi_{J(i,t)}^{m(i)g(i)}$:

$$\psi_{J(i,t)}^{m(i)g(i)} = \psi_{J(i,t)}^{0} + \sigma_{\tilde{s}(J(i,t))m(i)}^{m} + \sigma_{\tilde{s}(J(i,t))g(i))}^{g}. \tag{12}$$

There are two differences to note. First, rather than classify firms by sectors ($s(J)$) we further subdivide each sector into three firm classes using k-means clustering, where $\tilde{s}(J)$ denotes the sector-class of firm J. Second, to increase power we constrain gender-sector-class effects and major-sector-class effects to be additively separable.

Table C4 decomposes the gender wage gap using this alternative specification. Overall, this specification yields results similar to that of the augmented model examined in the main text. Major operates through the sorting channel, not the bargaining channel.

C.3 Imputing Topcoded Earnings

A key limitation of the earnings data is that monthly earnings are topcoded. Overall, 3% of observations are topcoded, and 17% of observations are topcoded in our sample of college admits. To account for topcoding, we follow Card et al. (2013) and impute topcoded earnings using a Tobit model. We use a series of Tobit models (in practice, the STATA function `mi impute intreg`) fit separately by year, gender, exam score decile (where those without exam scores are assigned to a distinct category), and age range (four 10-year ranges). These Tobit models for a given year include the worker's average earnings and topcoding rate in all other years, the average earnings and topcoding rate of his or her coworkers in that year, and firm size.

We replicate the main tables using this approach: Table C5 summarizes earnings outcomes by major; Table C6 decomposes the gender earnings gap into firm and worker components; and Table C7 decomposes the gender earnings gap using a regression approach. When we impute topcoded earnings, the gender pay gap increases from about 25 log points to 32 log points, with most of that increase coming from the worker component. However, imputing topcoded earnings does not alter our main findings.

Results are also similar if we also exclude workers that were admitted to the two most selective universities—Universidad de Chile and Pontificia Universidad Católica de Chile—where admits

[3] We implement this clustering using Bonhomme et al. (2019) companion R package, `rblm`.

have the highest rates of topcoded earnings. Excluding these workers reduces the topcoding rate to 14%.

C.4 Variation in Hours

A second key limitation of the earnings data is that they do not contain information on hours worked. As a result, variation in worker and firm effects may in part be driven by systematic variation in hours worked across workers and firms. This feature is particularly relevant in our context because the gender earnings gap may be driven in part by differences in hours of work by men and women.

We re-examine differences in earnings and worker and firm components across majors after limiting our AKM estimation sample to men to reduce the role of variation in hours worked. Table C8 summarizes earnings outcomes by major using data for men only. As in the main text, the covariance between firm and worker effects is much stronger between than within majors.

C.5 The Role of Industry

A significant proportion of the differences in earnings across majors may be explained by the fact that majors are associated with particular industries, and industries vary substantially in their average firm effects. We can decompose the sorting covariance term further into

$$\begin{aligned}\mathrm{Cov}(\alpha_i + X_{it}'\beta, \psi_{J(i,t)}) =& \mathrm{Cov}(\alpha_i + X_{it}'\beta, \bar{\psi}_{s(J(i,t))} + \eta_{J(i,t)}) \\ =& \underbrace{\mathrm{Cov}(\alpha_i + X_{it}'\beta, \bar{\psi}_{s(J(i,t))})}_{\text{between industries}} + \underbrace{\mathrm{Cov}(\alpha_i + X_{it}'\beta, \eta_{J(i,t)})}_{\text{within industries}}, \end{aligned} \tag{13}$$

where $s(J(i,t))$ is the *sector* of firm J, $\bar{\psi}_s$ is the (employee-weighted) average firm effect across firms in sector s, and $\eta_{J(i,t)}$ is the residual firm effect for J defined such that $\psi_{J(i,t)} = \bar{\psi}_{s(J(i,t))} + \eta_{J(i,t)}$.

57% percent of the between-major covariance is explained by the fact that majors associated with high worker effects have workers that sort to high-paying *industries*. By contrast, industry explains only 19% of the within-major covariance.[4]

[4]In Appendix Table C1 we replicate the analysis after grouping firms into classes as in Bonhomme et al. (2019). The results are qualitatively similar. In particular, assortative matching is about twice as strong between fields of study than within fields of study.

TABLE C1
EARNINGS OUTCOMES BY MAJOR, BLM

							Variance Decomposition			
Major	Male Share	Female Share	Log Earnings	Worker Component	Firm Component	Firm Ratio	Worker Share	Firm Share	Covariance Share	Residual Share
Overall							0.597	0.081	0.180	0.142
Between Major							0.563	0.067	0.349	0.021
Within Major							0.601	0.082	0.164	0.153
Business	11.4	14.6	0.109	0.092	0.015	0.131	0.621	0.079	0.161	0.139
Agriculture	9.3	6.7	-0.115	-0.055	-0.055	0.480	0.636	0.078	0.135	0.151
Architecture and Art	2.2	3.2	-0.242	-0.197	-0.039	0.161	0.589	0.085	0.163	0.163
Natural Science	5.4	7.2	-0.079	-0.077	0.000	-0.001	0.597	0.077	0.178	0.148
Social Science	2.3	5.5	-0.186	-0.146	-0.038	0.201	0.606	0.073	0.161	0.160
Law	2.3	2.4	-0.027	-0.013	-0.058	0.216	0.592	0.069	0.202	0.137
Education	7.4	24.8	-0.392	-0.293	-0.097	0.248	0.620	0.075	0.147	0.158
Humanities	1.9	4.7	-0.300	-0.261	-0.034	0.115	0.592	0.076	0.167	0.165
Health	3.3	9.3	0.074	0.104	-0.032	-0.434	0.673	0.056	0.109	0.162
Technology	54.5	21.7	0.178	0.124	0.052	0.293	0.573	0.093	0.180	0.154

Firm effects and worker effects are estimates from equation (1), which is described in more detail in Section III. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in III. The between-field decomposition weights fields by their number of earnings observations.

Table C2
Contribution of Firm-Specific Pay Premiums and
Major to the Gender Earnings Gap, BLM

	Gender Earnings Gap	Firm Component	Worker Component
Overall	0.199	0.057	0.157
		(0.286)	(0.789)
Non-College Admits	0.202	0.058	0.159
		(0.287)	(0.787)
College Admits	0.252	0.056	0.195
		(0.222)	(0.774)
Between Major Area	0.134	0.037	0.095
		(0.276)	(0.709)
Residual	0.118	0.019	0.101
		(0.161)	(0.856)
By Major Area:			
Business	0.146	0.002	0.147
Agriculture	0.221	0.015	0.202
Architecture and Art	0.065	0.005	0.061
Natural Science	0.190	0.036	0.154
Social Science	0.049	0.006	0.039
Law	0.048	-0.001	0.038
Education	0.029	0.021	0.019
Humanities	-0.007	0.009	0.006
Health	0.041	0.019	0.020
Technology	0.210	0.040	0.170
Between Specific Major	0.163	0.046	0.114
		(0.282)	(0.699)
Residual	0.089	0.009	0.081
		(0.101)	(0.910)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section IV.B using firm and worker component estimates described in Section C.2. Entries in parentheses represent the percent of the overall male-female earnings gap (in column 1) that is explained by the source described in column heading.

Table C3
Gender Earnings Gap and Firm Sorting, BLM

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Log Earnings							
Male	0.228	0.143	0.060	0.114	0.074	0.009	0.069
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Outcome: Firm Effect							
Male	0.055	0.039	0.032	0.022	0.016	0.013	0.008
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome: Worker Effect							
Male	0.178	0.110		0.098	0.065		0.066
	(0.002)	(0.002)		(0.002)	(0.002)		(0.002)
Exam Scores		✓			✓		✓
Worker Effects			✓			✓	
Major Area				✓	✓	✓	
Specific Major							✓

This table presents estimates of equation (4), a regression model for the male-female differences in earnings outcomes. All specifications include a full set of indicators for potential work experience (years since last admissions exam). There are 2,791,187 observations for 420,435 workers. Standard errors are clustered at the worker level.

TABLE C4

AUGMENTED MODEL DECOMPOSITION, ALTERNATIVE SPECIFICATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Means of Firm Premiums			Decompositions of Contribution of Firm Component			
					Sorting		Bargaining	
	Gender Wage Gap	Male Premium among Men	Female Premium among Women	Total Contribution of Firm Components	Using Male Effects	Using Female Effects	Using Male Distribution	Using Female Distribution
Overall	0.252	0.323	0.245	0.078	0.066	0.062	0.013	0.016
				(0.310)	(0.262)	(0.246)	(0.052)	(0.063)
Between Major Area	0.134			0.040	0.039	0.036	0.001	0.004
				(0.299)	(0.291)	(0.269)	(0.007)	(0.030)
Remainder	0.118			0.039	0.027	0.026	0.012	0.012
				(0.331)	(0.229)	(0.220)	(0.102)	(0.102)
By Major Area:								
Business	0.146	0.270	0.248	0.022	0.005	0.004	0.017	0.018
Agriculture	0.221	0.309	0.260	0.049	0.035	0.033	0.015	0.016
Architecture and Art	0.065	0.287	0.259	0.028	0.012	0.011	0.015	0.016
Natural Science	0.190	0.354	0.291	0.064	0.050	0.049	0.013	0.015
Social Science	0.049	0.277	0.258	0.019	0.008	0.006	0.011	0.013
Law	0.048	0.309	0.290	0.019	0.003	0.003	0.016	0.016
Education	0.029	0.203	0.170	0.033	0.026	0.022	0.007	0.011
Humanities	-0.007	0.240	0.237	0.003	-0.008	-0.009	0.011	0.012
Health	0.041	0.303	0.279	0.024	0.016	0.015	0.008	0.009
Technology	0.210	0.358	0.287	0.072	0.055	0.054	0.017	0.018
Between Specific Major	0.163			0.051	0.050	0.047	0.002	0.004
				(0.313)	(0.307)	(0.288)	(0.012)	(0.025)
Remainder	0.089			0.027	0.016	0.015	0.011	0.012
				(0.303)	(0.180)	(0.169)	(0.124)	(0.135)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section C.2. Entries in parentheses represent the percent of the overall male-female earnings gap (in column 1) that is explained by the source described in column heading.

Table C5

Earnings Outcomes by Major, Imputed Earnings

							Variance Decomposition			
Major	Male Share	Female Share	Log Earnings	Worker Component	Firm Component	Firm Ratio	Worker Share	Firm Share	Covariance Share	Residual Share
Overall							0.591	0.126	0.167	0.116
Between Major							0.488	0.097	0.401	0.014
Within Major							0.601	0.129	0.142	0.128
Business	11.4	14.6	0.139	0.124	0.014	0.101	0.644	0.110	0.133	0.113
Agriculture	9.3	6.7	-0.139	-0.075	-0.060	0.431	0.641	0.121	0.109	0.129
Architecture and Art	2.2	3.2	-0.288	-0.198	-0.082	0.285	0.595	0.148	0.119	0.138
Natural Science	5.4	7.2	-0.095	-0.094	-0.000	0.004	0.586	0.128	0.164	0.122
Social Science	2.3	5.5	-0.228	-0.152	-0.074	0.324	0.605	0.125	0.135	0.135
Law	2.3	2.4	-0.002	0.059	-0.045	21.152	0.617	0.104	0.165	0.114
Education	7.4	24.8	-0.458	-0.324	-0.135	0.294	0.609	0.142	0.108	0.141
Humanities	1.9	4.7	-0.354	-0.282	-0.069	0.195	0.607	0.133	0.120	0.140
Health	3.3	9.3	0.037	0.066	-0.032	-0.871	0.634	0.108	0.117	0.141
Technology	54.5	21.7	0.213	0.134	0.076	0.359	0.574	0.136	0.165	0.125

Firm effects and worker effects are estimates from equation (1), which is described in more detail in Section III. We impute topcoded earnings as described in Section C.3. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in III. The between-field decomposition weights fields by their number of earnings observations.

Table C6
Contribution of Firm-Specific Pay Premiums and Major to the Gender Earnings Gap, Imputed Earnings

	Gender Earnings Gap	Firm Component	Worker Component
Overall	0.207	0.058	0.156
		(0.280)	(0.754)
Non-College Admits	0.205	0.057	0.156
		(0.278)	(0.761)
College Admits	0.319	0.085	0.230
		(0.266)	(0.721)
Between Major Area	0.161	0.054	0.106
		(0.335)	(0.658)
Residual	0.158	0.031	0.123
		(0.196)	(0.778)
By Major Area:			
Business	0.213	0.007	0.203
Agriculture	0.261	0.032	0.222
Architecture and Art	0.082	0.018	0.057
Natural Science	0.238	0.057	0.177
Social Science	0.077	0.016	0.055
Law	0.088	0.006	0.064
Education	0.043	0.037	0.008
Humanities	0.006	-0.006	0.013
Health	0.078	0.018	0.056
Technology	0.275	0.062	0.208
Between Specific Major	0.200	0.067	0.132
		(0.335)	(0.660)
Residual	0.119	0.019	0.098
		(0.160)	(0.824)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section IV.B using firm and worker component estimates described in Section C.2. We impute topcoded earnings as described in Section C.3. Entries in parentheses represent the percent of the overall male-female earnings gap (in column 1) that is explained by the source described in column heading.

Table C7

Gender Earnings Gap and Firm Sorting, Imputed Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Log Earnings							
Male	0.287	0.186	0.056	0.153	0.101	0.021	0.095
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)
Outcome: Firm Effect							
Male	0.085	0.067	0.057	0.038	0.033	0.024	0.022
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome: Worker Effect							
Male	0.206	0.123		0.120	0.074		0.076
	(0.002)	(0.002)		(0.002)	(0.002)		(0.002)
Exam Scores		✓			✓		✓
Worker Effects			✓			✓	
Major Area				✓	✓	✓	
Specific Major							✓

This table presents estimates of equation (4), a regression model for the male-female differences in earnings outcomes. All specifications include a full set of indicators for potential work experience (years since last admissions exam). We impute topcoded earnings as described in Section C.3. There are 2,791,187 observations for 420,435 workers. Standard errors are clustered at the worker level.

Table C8
Earnings Outcomes by Major, Men Only

						Variance Decomposition			
Major	Male Share	Log Earnings	Worker Component	Firm Component	Firm Ratio	Worker Share	Firm Share	Covariance Share	Residual Share
Overall						0.547	0.178	0.134	0.141
Between Major						0.421	0.136	0.402	0.041
Within Major						0.554	0.181	0.117	0.148
Business	11.4	0.077	0.093	-0.017	-0.223	0.596	0.161	0.103	0.140
Agriculture	9.3	-0.143	-0.055	-0.084	0.589	0.610	0.155	0.083	0.152
Architecture and Art	2.2	-0.310	-0.197	-0.103	0.331	0.558	0.194	0.094	0.154
Natural Science	5.4	-0.090	-0.080	-0.009	0.103	0.553	0.174	0.135	0.138
Social Science	2.3	-0.255	-0.159	-0.091	0.357	0.594	0.158	0.108	0.140
Law	2.3	-0.106	-0.026	-0.067	0.626	0.577	0.157	0.137	0.129
Education	7.4	-0.470	-0.321	-0.144	0.306	0.573	0.175	0.110	0.142
Humanities	1.9	-0.403	-0.290	-0.106	0.263	0.582	0.153	0.115	0.150
Health	3.3	-0.002	0.051	-0.048	-22.745	0.619	0.143	0.082	0.156
Technology	54.5	0.123	0.063	0.055	0.450	0.526	0.196	0.127	0.151

Firm effects and worker effects are estimates from equation (1), which is described in more detail in Section III, using men only. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in III. The between-field decomposition weights fields by their number of earnings observations.