Firm Sorting, College Major, and the Gender Earnings Gap

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

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

May 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 a common form of pre-labor market skill specialization, college major, mediates access to high-paying firms, and what this means for the gender earnings gap. Using employer-employee tax data from Chile matched to educational records, we show that differences in college major account for more than two-thirds of the firm contribution to the gender earnings gap among college admits. Degrees in Technology, which are numerous, male-dominated, and associated with high firm premiums, drive these effects.

*We greatly appreciate the support of Cristobal Huneeus for this project. We thank Barbara Biasi, Sydnee Caldwell, Pat Kline, Jonathan Leonard, David Levine, and Heather Sarsons for comments.

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.[1] 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; 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 college major 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 and firm tax 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 college and major in which students apply and are admitted to, subject exam scores, as well as earnings, (anonymized) employer identifiers and employer characteristics such as value-added. Our data cover all private sector workers between 2005 and 2019, and a majority of students who take the national college admissions exam between 1977 and 2000. We use these data to estimate earnings models that allow for worker and firm fixed effects (Abowd et al. (1999); henceforth AKM).

We document 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 37% of cross-major variance in earnings. A leading example is Technology majors, for which firm premiums explain 43% of the 19% earnings premium relative to the average major. In contrast, the covariance between worker effects and firm premiums explains only 8% 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 67% and 82% 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, pre-market differences in major explain most of the firm component of the gender gap.

We begin by showing that features of the earnings distribution observed in other countries are present in Chile as well. Men in our data earn 22% more than women; this gap increases to

[1] This literature begins with Card et al. (2016), who study Portugal, and includes Sorkin (2017) (United States), Cruz and Rau (2017) (Chile), Coudin et al. (2018) (France), Bruns (2019) (Germany), Casarico and Lattanzio (2019) (Italy), Morchio and Moser (2020) (Brazil), and Li et al. (2020) (Canada).

34% for students admitted to college. This is comparable to the 23% male-female gap reported in Card et al. (2016; henceforth CCK) using Portuguese data and the 23% gap Sloane et al. (2019) 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 76% 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 10% of the between-major variance in mean earnings across coarse major categories, while the covariance between firm effects and worker effects explains 37%. This effect is driven by strong assortative matching across popular fields like Technology and Education. The mean firm premium for Technology majors is 8.0 log points higher than the premium for the average major, 43% of the 18.5 log point earnings premium for Technology majors. For Education majors, the mean firm premium is 10.5 log points lower than the premium for the average major, 28% of the overall major effect.

Our second main finding is that college major explains a majority of the firm contribution to the gender earnings gap. Among college admits, the male earnings premium is 34.2 log points and the firm effect premium is 7.6 log points, or 22% of the total gap. This is comparable to the firm contribution to gender pay gaps documented in prior studies and other countries. After conditioning on major area, the gender gap in mean firm premium falls by 67%.

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 pay premiums. Overall, the sorting component explains 93-100% of the firm contribution to the gender earnings gap among college admits. Major accounts for 65-80% of the sorting component, but essentially none of the bargaining component.

Finally, we show that the relationship between college major and firm premiums does not simply reflect sorting on vertical measures of skill, as measured either by college admissions exam 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 about 60%. Major retains most of its explanatory power even after conditioning on ability measures. Compared to exam scores, college major is more closely tied to firm effects. 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 three 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 earnings premiums requires an understanding of job matching and firm pay premiums in addition to an understanding of skill formation and skill demand.

Third, our findings have implications for the literature on worker-firm wage decompositions, including both papers that use the two-way fixed effect approach of AKM and papers that use alternative approaches (e.g., Bagger and Lentz, 2019; Bonhomme et al., 2019). This literature generally interprets worker-specific pay premiums as an index of worker quality or skill. Our findings imply that worker-specific pay premiums are not a sufficient statistic for worker skill in the sense that, conditional on worker-specific pay premiums, individuals have different skill bundles, and those bundles predict the firm-specific pay premium they earn.[2]

I Data and Context

We use data from Chile that cover firms, worker earnings, demographics, and college entrance exam scores and admissions outcomes. We summarize the data here; more details on the construction of the data are provided in Online Appendix A.

Data on firms and worker earnings come from administrative and anonymized tax records from the Internal Revenue Service (IRS, or SII for its acronym in Spanish) for the years 2005 to 2019. We calculate firm value-added from balance sheet information on total sales and costs. We use employer-employee data on workers' formal employment contracts to define a worker-firm spell. Earnings within each worker-firm spell include all income from that firm including wages, salaries, bonuses, tips, and other sources of labor income deemed taxable by the IRS.

Data on worker demographics come from the civil registry and includes worker gender and year of birth.

Data on human capital come from college entrance exams and admissions outcomes that were regularly published in the newspaper in Chile and digitized by Hastings, Neilson and Zimmerman (2013), henceforth HNZ. Chilean high school graduates interested in attending college take a national entrance exam, which in the relevant period was called the *Prueba de Aptitude Academica* (PAA). Students submit applications through a centralized assignment system by providing a

[2] An emerging literature studies worker assortative matching allowing for multidimensional skill; this literature has focused on occupations rather than firms (Lindenlaub, 2017; Lise and Postel-Vinay, 2019).

ranked ordered list of college-major pairs, which we will henceforth call 'degree programs'. The system assigns students to degree programs using a deferred acceptance algorithm in which degree programs rank students using a weighted average of standardized exam scores and high school grades. Historically, the results of the exams and admissions process were published in the newspaper each year. We use data on the entrance exam math and verbal scores and the admissions outcomes for the majority of test-takers and accepted applicants from 1977 to 2000.[3]

We attribute students to a major using the last observed degree program to which they applied and were accepted.[4] Degree programs are classified by field of study based on UNESCO Normalized International Classification of Education standards. There are ten categories: Agriculture, Art/Architecture, Business Administration, Education, Health, Humanities, Law, Natural Science, Social Science, and Technology. We call this classification 'major areas'. We label Agriculture, Health, Natural Science, and Technology as STEM fields. We also consider a narrower set of administrative codes, which divide majors into 124 categories. We call this classification 'specific majors'. Examples of such codes include Mechanical Engineering and Pedagogy in Elementary Education.

We merge these data sets using unique tax IDs of workers and firms that are common across sources. To secure the privacy of workers and firms, we cannot observe the merged dataset and the Chilean IRS requires all results that are extracted to be calculated using at least 25 tax IDs.

For the employer-employee data set, we impose the following sample restrictions following the criteria of prior work (Card et al. (2013), henceforth CHK). The earnings records are subject to one important limitation: the data do not include information on hours or weeks worked, so we cannot measure wages. In order to exclude weak labor attachment and reduce the problem of unobserved hours, we limit the analysis to earnings observations above half the 2014 (annualized) full-time minimum wage. In robustness checks described in Section IV, we replicate our analysis using unemployment insurance records on monthly earnings and estimate firm pay premiums using only men's earnings to reduce variability in hours worked. We limit our analysis to workers between the ages of 20 and 60 and consider only the highest paying job for each worker every year. We call this sample the *baseline sample.* We deflate all nominal variables so that they represent 2015 real dollars.

We build a second sample, which restricts to workers that we observe as admitted to a degree program between 1977 to 2000 and their earnings at least six years after the year of admission. We call this second sample the *college sample.* We define the complement of the last sample as the *non-college sample.*

[3]Our data exclude admissions outcomes for universities outside the centralized system, which account for less than 5% of university graduates in the 1980's, and 33% by 2000 (Brunner, 2009; Rodrigo et al., 2010). These newer universities typically serve lower-scoring students.

[4]While we do not directly observe enrollment and matriculation, evidence reported in HNZ indicates that enrollment rates in admitted programs are high.

II 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 firm 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 earnings, 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 annual earnings observations.

We estimate over 8 million worker effects and about 580,000 firm effects. The standard deviation of worker effects is 0.599, about 90% 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 0.243. 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 (CHK):

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

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 earnings. This means that workers that switch firms do not sort based on match effects or worker-specific shocks to earnings. We follow the now-standard specification checks developed in CHK 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 baseline sample, the college sample, 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 earnings decomposition, where the variance of log earnings can be written as

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

The covariance between worker and firm components reflects the degree of assortative matching. If high earnings workers are more likely to work at high earnings firms, then this covariance term will be positive. Below we will express these variance and covariance terms scaled by Var($\log w_{it}$), which can be interpreted as the share of earnings variation explained by some component.

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 split-sample approach described in Gerard et al. (2018). Second, we consider an alternative earnings 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 similar results under these alternative approaches.

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 III.C, we consider an augmented model that allows firm pay premiums to vary by gender and college major.

III Results

III.A Earnings Outcomes and Firm Sorting by College Major

In this section we describe earnings outcomes and firm sorting by college major. We compare how firm sorting patterns among workers from the same major compare to sorting patterns between majors.

The top row of Table I decomposes the earnings of college admits using equation (2). Overall, we find that about 63% of the variance in log earnings among college admits can be attributed to the worker component, 16% to the firm component, and 10% to worker-firm sorting, with 10% remaining as a residual. The results of this decomposition are similar to findings in prior work (Card et al., 2018).

We then further decompose variation in earnings into between- and within-major components. We decompose the variance of earnings and the covariance between worker and firm components. We decompose the former using

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

and the latter using

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

where $m(i)$ indexes the major of worker i.

We now come to our first major result. We find that assortative matching is markedly stronger *between* majors than *within* majors. This can be seen in the share of earnings variation explained by the covariance of worker and firm effects. This share is 8% within major areas. By contrast, assortative matching accounts for 37% of variation in earnings across major areas. Major accounts for only 8% of the variance in earnings in the college sample, but accounts for 29% of the sorting component.

Table I also reports means of the following objects for each college major: log earnings, the worker component, the firm component, and log value-added per worker. We demean each outcome, so that reported means for each major are differences from the average earnings, firm, or worker effect premium of the college sample. 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 II 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 Law and the lowest earnings among those studying Education and Humanities. Technology majors have log earnings 18.5 log points higher than average, while Education and Humanities majors have earnings 38.1 and 24.8 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. 46.1% of all men's observations are for those admitted to Technology degree programs, compared to 14.2% of women; for Education, these shares are 10.8% and 31.8% respectively.

The finding that assortative matching is stronger between between majors than within majors is driven in large part by the popular Technology and Education fields. The mean firm premium for Technology majors is 8.0 log points, 43% of the earnings premium for Technology majors relative to the average major. For Education majors, the mean firm premium is -10.5 log points, 28% of the earnings premium for Education majors. All fields except Health follow similar patterns; for Health majors, negative average firm effects partially offset positive average worker effects.

III.B College Major 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 the baseline sample, the overall gender earnings gap is 22.3 log points. The gaps in firm and worker components are 5.0 log points and 19.0 log points. In the college (non-college) sample, the gender earnings gap is 34.2 (22.4) log points, with a 7.6 (4.9) log point gap in the firm component and 26.5 (19.3) log point gap in the worker component. The firm component of the gender gap is thus 22% of the total gap in the baseline sample, non-college sample, and college sample.[5]

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 annual earnings, the worker component, or the firm component and we refer to the remaining terms as the 'residual'.

[Table 2 about here.]

We find that in the college sample the between-major gender earnings gap is 13.7 points and the residual gender earnings gap is 20.5 log points. That is, differences in average earnings by major explain 40% of the gender gap among college admits. The cross-major component of the gender gap is even larger—18.9 log points, or 55%—when we consider narrow major categories.

This brings us to our second main finding: the contribution of majors to the gender earnings gap comes disproportionately through the *firm* component of earnings 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 5.1 log points when we use coarse major categories, and 6.2 log points when we use the finer categories. These values are equal to 67% and 82% of the firm component of the gender gap, respectively. In contrast, major areas and specific majors explain 30% and 46% of the worker component of the gender gap, respectively.

Another way to frame this is to observe that the firm component accounts for 37% of the between-major area gender earnings gap, but only 12% 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 33% of the between-specific-major gender earnings gap, but only 10% of the residual gender earnings gap.

[5]CCK also find that the firm contribution to the gender gap is similar across education levels.

The bottom panel of Table II reports earnings gaps and mean firm and worker contributions to those gaps for individual major areas. The male-female earnings gap is largest in STEM fields of Agriculture, Natural Science, and Technology, where it exceeds 29 log points. It is also large in Business (27.8 log points). The firm contribution to the gap is largest in Natural Science and Technology, 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

$$Y_{it} = \alpha + \beta \text{Male}_i + X_{it}\delta + \mu_{m(i)} + \epsilon_{it}, \tag{4}$$

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 $\mu_{m(i)}$ 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 32.2 log points of which 7.6 log points (24%) 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 32% to 21.9 log points, and the firm contribution to the gap by 18%, to 6.2 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 to 6.1 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 replace controls for exam scores with major area fixed effects. Controls for major reduce the firm sorting component by 59%, to 3.1 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 56%, to 2.7 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 64%, to 2.2 log points. Finally, in column 7 we include specific major fixed effects. These controls decrease the earnings gap to 14.7 log points and the unexplained sorting effect to 1.9 log points, 69% lower 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. We find a similar pattern of estimates when we replace the outcome with log value-added per worker (as in Card et al. (2018)).

[Table 3 about here.]

III.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 earnings model can be written as

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

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.[6] 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:

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

[6] This specification differs from the canonical CCK model, which allows firm effects to vary arbitrarily with gender. It is not feasible to allow firm effects to vary arbitrarily with college major given the limited number of college admits.

where

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

and $\alpha^{M,k}_s = P(m(i) = k|\text{male}; s(J(i,t)) = s)$. $\tilde{\psi}^M_j$ 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 21.0 log points, and the average female firm premium among women is 13.6 log points. The total firm contribution to the gender gap in the augmented model is the difference between these numbers: 7.5 log points. This is similar to the firm contribution we observed in the baseline model (7.6 log points). Most of the firm contribution to the gender gap comes through the sorting channel. Sorting accounts for 7.0 log points (93%) of the firm gender gap when computed using firm effects for men, and 7.5 log points (100%) when computed using firm effects for women.

As in the baseline AKM model, cross-major variation accounts for most of the firm component of the gender gap. Out of the 7.6 log point gap in firm effects, 4.7 log points (63%) comes from variation between major areas and 5.8 log points (76%) comes from variation between specific majors. Essentially all of this is through the sorting channel. Measured using female firm effects, major explains 5.0 log points (79%) 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:

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

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. In contrast to the baseline model, the augmented model suggests substantial firm components to the gender gap in Humanities and Art/Architecture, operating through the bargaining channel. These fields account for a relatively small share of college admits students–4.1% of men and 7.9% 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 mostly captures the sorting component of the firm contribution to gender gaps and its interactions with college major. 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 almost entirely through the sorting channel.

[Table 4 about here.]

IV Additional Analyses and Robustness Tests

Online Appendix C addresses possible concerns about our estimation approach. We evaluate and account for limited mobility bias using approaches developed in Kline et al. (2020), Gerard et al. (2018), and Bonhomme et al. (2019), and earnings data from unemployment insurance records. To address concerns that our annual earnings data miss variation in hours worked, we estimate firm effects using only male workers to reduce variation in hours worked. None of these analyses alter our main findings. We also consider the role of industry in mediating major effects.

V Conclusion

This paper uses administrative data 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 38% of earnings variation across majors is due to the fact that workers admitted to degree programs 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 43% of the mean earnings premium relative to the average over all majors. Overall, differences in major explain the majority—roughly 70-80%, depending on the measure used—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 exam 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 college major 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 (Card et al., 2018), 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.

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.

Bagger, Jesper and Rasmus Lentz, "An Empirical Model of Wage Dispersion with Sorting," *Review of Economic Studies*, January 2019, *86* (1), 153–190.

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.

Bordon, Paola and Chao Fu, "College-major choice to college-then-major choice," *The Review of economic studies*, 2015, *82* (4), 1247–1288.

Brunner, José Joaquín, *Educación superior en Chile: instituciones, mercados y políticas gubernamentales (1967-2007)*, Educaciones Universidad Diego Portales, 2009.

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.

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.

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

Coudin, Elise, Sophie Maillard, and Maxime Tô, "Family, Firms and the Gender Wage Gap in France," Working Papers 2018-09, Center for Research in Economics and Statistics 2018.

Cruz, Gabriel and Tomás Rau, "The Effects of Firms' Pay Policies and Equal Pay Laws on the Gender Wage Gap in Chile," May 2017. Unpublished manuscript.

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.

Li, Jiang, Benoit Dostie, and Gaëlle Simard-Duplain, "What Is the Role of Firm-Specific Pay Policies on the Gender Earnings Gap in Canada?," November 2020. IZA Discussion Paper 13907.

Lindenlaub, Ilse, "Sorting Multidimensional Types: Theory and Application," *Review of Economic Studies*, 2017, *84*, 718–789.

Lise, Jeremy and Fabien Postel-Vinay, "Multidimensional Skills, Sorting, and Human Capital Accumulation," July 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.

Rodrigo, Roland M., Juan V. Salamanca, and Marcelo Q. Aliaga, "Evolución Matrícula Educación Superior de Chile: Periodo 1990 – 2009," Technical Report, Servicio de Información de Educación Superior, Chilean Ministry of Education 2010.

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.

TABLE I

EARNINGS OUTCOMES BY MAJOR

Major								Variance Decomposition			
	Male Share (1)	Female Share (2)	Log Earnings (3)	Worker Component (4)	Firm Component (5)	Log VA Per Worker (6)	Firm Ratio (7)	Worker Share (8)	Firm Share (9)	Covariance Share (10)	Residual Share (11)
Overall								0.633	0.162	0.102	0.103
Between Major								0.530	0.104	0.366	-0.000
Within Major								0.645	0.164	0.075	0.117
Agriculture	8.5	4.6	-0.066	-0.008	-0.058	0.425	0.870	0.698	0.152	0.036	0.113
Architecture and Art	3.0	3.6	-0.163	-0.098	-0.067	0.127	0.411	0.658	0.171	0.050	0.121
Business	11.3	11.1	0.163	0.123	0.039	0.445	0.240	0.668	0.154	0.083	0.094
Education	10.8	31.8	-0.381	-0.275	-0.105	-1.377	0.276	0.643	0.156	0.067	0.134
Health	6.8	15.4	0.138	0.179	-0.042	-1.189	-0.304	0.712	0.107	0.009	0.173
Humanities	1.9	4.7	-0.248	-0.210	-0.037	-0.152	0.151	0.636	0.157	0.084	0.123
Law	3.4	2.8	0.317	0.205	0.114	-2.632	0.359	0.612	0.161	0.111	0.116
Natural Science	5.2	5.6	-0.072	-0.071	0.000	0.323	-0.004	0.625	0.171	0.089	0.116
Social Science	3.1	6.2	-0.057	-0.059	0.001	-1.297	-0.019	0.624	0.161	0.082	0.134
Technology	46.1	14.2	0.185	0.106	0.080	0.998	0.430	0.613	0.193	0.102	0.092

This table presents evidence of earnings outcomes by major using the college sample described in Section I. Columns 1 and 2 present the 'Male Share' and 'Female Share', respectively, which refer to the percentage of male and female college admits in each major. Columns 3 through 6 present the average log earnings, worker effect, firm effect and log value-added per worker in each major, relative to average of each variable in the college sample. Column 7 presents the 'Firm Ratio', which is the ratio of the firm component of column 5 and the log earnings of column 3. Columns 8 through 11 present the variance decomposition of log earnings into the worker effect, firm effect, covariance between worker and firm effect, and residual. Firm effects and worker effects are estimates from equation (1), which is described in more detail in Section II. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in Section II. The between-field decomposition weights fields by their number of earnings observations.

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

	Gender Earnings Gap	Firm Component	Worker Component
Overall	0.223	0.050	0.190
		(0.224)	(0.852)
Non-College Admits	0.224	0.049	0.193
		(0.219)	(0.865)
College Admits	0.342	0.076	0.265
		(0.222)	(0.775)
Between Major Area	0.137	0.051	0.080
		(0.372)	(0.584)
Residual	0.205	0.025	0.185
		(0.122)	(0.902)
By Major Area:			
Agriculture	0.292	-0.001	0.300
Architecture and Art	0.174	0.005	0.170
Business	0.278	-0.008	0.298
Education	0.187	0.047	0.147
Health	0.237	0.015	0.221
Humanities	0.125	0.020	0.100
Law	0.035	-0.045	0.083
Natural Science	0.294	0.058	0.241
Social Science	0.143	0.013	0.133
Technology	0.337	0.066	0.280
Between Specific Major	0.189	0.062	0.122
		(0.328)	(0.646)
Residual	0.153	0.015	0.143
		(0.098)	(0.935)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker earnings components as described in Section III.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. The first row, Overall, uses the baseline sample. The second (third) row, Non-College (College) admits, uses the non-college (college) employer-employee sample. The remaining rows, use the college sample.

Table III
Gender Earnings Gap and Firm Sorting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Log Earnings							
Male	0.322	0.219	0.064	0.212	0.154	0.014	0.147
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Outcome: Firm Effect							
Male	0.076	0.062	0.061	0.031	0.027	0.022	0.019
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome: Worker Effect							
Male	0.255	0.158		0.200	0.141		0.141
	(0.002)	(0.002)		(0.002)	(0.002)		(0.002)
Outcome: Log VA Per Worker							
Male	1.225	1.034	1.079	0.553	0.494	0.469	0.396
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
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 using the college sample described in Section I. 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. For the first three outcomes there are 3,816,066 observations for 456,049 workers. For log value-added per worker there are 2,621,730 observations for 398,948 workers. Standard errors are clustered at the worker level. Descriptive statistics for potential work experience and exam scores by gender are reported in Appendix Table B2.

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	0.342	0.210	0.136	0.075	0.070	0.075	0.000	0.005
				(0.219)	(0.205)	(0.219)	(0.000)	(0.015)
Between Major Area	0.137			0.047	0.045	0.050	-0.004	0.002
				(0.343)	(0.328)	(0.365)	(-0.029)	(0.015)
Remainder	0.205			0.028	0.025	0.025	0.004	0.003
				(0.137)	(0.122)	(0.122)	(0.020)	(0.015)
Between Specific Major	0.189			0.058	0.055	0.061	-0.003	0.002
				(0.307)	(0.291)	(0.323)	(-0.016)	(0.011)
Remainder	0.153			0.017	0.015	0.014	0.003	0.003
				(0.111)	(0.098)	(0.092)	(0.020)	(0.020)
By Major Area:								
Agriculture	0.292	0.131	0.123	0.007	0.001	0.004	0.003	0.006
Architecture and Art	0.174	0.123	0.115	0.008	-0.007	0.007	0.001	0.015
Business	0.278	0.202	0.202	0.001	0.003	-0.025	0.025	-0.004
Education	0.187	0.112	0.068	0.044	0.042	0.055	-0.011	0.001
Health	0.237	0.147	0.110	0.037	0.017	-0.005	0.041	0.020
Humanities	0.125	0.165	0.126	0.040	0.008	0.014	0.025	0.032
Law	0.035	0.366	0.380	-0.014	-0.141	0.017	-0.031	0.127
Natural Science	0.294	0.204	0.158	0.046	0.058	0.069	-0.023	-0.011
Social Science	0.143	0.221	0.177	0.044	-0.024	0.016	0.028	0.068
Technology	0.337	0.256	0.201	0.055	0.075	0.064	-0.009	-0.020

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section III.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 III.C. Column 4 reports the total contribution of firm-specific earnings 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 III.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. The first row, Overall, uses the baseline sample. The remaining rows, use the college sample.

A Data Construction

We use data from Chile that cover firms, worker earnings, demographics, and college entrance exam scores and admissions outcomes. The data on firms come from administrative and anonymized tax records from the Internal Revenue Service (IRS, or SII for its acronym in Spanish). The data on worker earnings also come from administrative tax records from the IRS. Data on worker demographics come from Chile's civil registry and includes worker gender and year of birth. Data on college major and college entrance exams come from admissions results that were digitized from historical newspaper publications by Hastings et al. (2013). We merge these data sets using unique tax IDs of workers and firms that are common across sources. To secure the privacy of workers and firms, we cannot observe the merged dataset and the Chilean IRS requires all results that are extracted to be calculated using at least 25 tax IDs.

A.1 Data of Firms and Worker Earnings

Administrative data on firm balance sheet characteristics provides information on total sales and costs (from IRS tax form 29). This data allows for the calculation of value-added for the population of firms for a 15 year period from 2005 to 2019. We also use administrative employer-employee data from the IRS (from the tax affidavit 1887) on workers formal employment contracts to define a worker-firm spell. Earnings within each worker-firm spell include all income from that firm including wages, salaries, bonuses, tips, and other sources of labor income deemed taxable by the IRS. Earnings reported in this data are net of social security benefits so we adjust them to include those benefits. Specialized high skill professions, such as doctors or lawyers, can receive earnings both through wages of labor contracts but also through additional work as an independent contractor for the same firm. Given that we observe all taxable labor income for each worker and its source, we add any income coming from independent contracts (from the tax affidavit 1879) invoiced to the same firm during the period of the formal employment spell as defined above. This allows for a more complete measure of labor earnings for highly skilled workers. We also observe monthly worker earnings from each firm as reported to the social security unemployment insurance program. This data includes wages from formal labor contracts but has the limitation that it is top coded. Relative to previous literature that uses social security data (Card et al., 2013), the tax data has two advantages. First, it has uncensored earnings. Second, we can complement the usual labor contract income with invoices of independent contractors, which can be a relevant source of earnings for high-skilled labor. More details of the unemployment insurance data can be found in Online Appendix A.4.

A.2 Data on Human Capital and College Major

Our data on human capital comes from college entrance exams and admissions outcomes that were regularly published in the newspaper in Chile and digitized by HNZ. Chilean high school graduates interested in attending higher education take a national college entrance exam, which in the relevant period was called the *Prueba de Aptitude Academica* (PAA). The results of these exams were published in the newspaper each year through the 70s, 80s and 90s. Students interested in attending college would then submit applications through a centralized assignment system by providing a ranked ordered list of college-major pairs, which we will henceforth call 'degree programs'. The system assigns students to degree programs using a deferred acceptance algorithm in which degree programs rank students using a weighted average of standardized exam scores and high school grades. Each student is assigned a spot in their most-preferred program willing to accept them. The results of the admissions process we also published in the newspaper each year. We use data on the entrance exam math and language scores and the admissions outcomes that were digitized from archival records of the newspaper publications. This data includes the majority of test-takers and accepted applicants from 1977 to 2000[7].

We attribute students to a major and other degree program characteristics when a student applied and was accepted at that degree program.[8] 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 call this classification 'Major Areas'. We label Agriculture, Health, Natural Science, and Technology as STEM fields. We also consider a narrower set of administrative codes, which divide majors into 124 categories. We call this classification 'Specific Majors' Examples of such codes include Mechanical Engineering and Pedagogy in Elementary Education.

A.3 Sample Construction and Restrictions

For the employer-employee data set, we impose the following sample restrictions following the criteria of previous literature (Card et al., 2013). The earnings records are subject to one important limitation. The data does not include information on hours or weeks worked, so we cannot measure wages. In order to exclude weak labor attachment and reduce the problem of unobserved hours, we limit the analysis to earnings observations above half the 2014 (annualized) full-time minimum wage. In robustness checks described in Section IV, we replicate our analysis using unemployment

[7] One limitation of data digitised from archival newspaper sources is that unsuccessful applications will not be observed unless wait listed. Applications that are ranked below an accepted offer are also not observed in these data (HNZ provides more information on these data). Another limitation is that some universities that begin to appear in the 1990s are not part of the centralized system and are not included in the data. While these off-platform options began to grow in relevance during the 1990s, the centralized system included in our data covered the majority of feasible college options for most of the cohorts in this study between 1977 and 2000. See Brunner (2009) for a review of the higher education system in Chile during this period.

[8] We believe this a good proxy for the college major the worker will eventually have, even though it is measured with error as some students may dropout or switch majors. While we do not directly observe enrollment and matriculation data, evidence reported in HNZ indicates that enrollment rates in the admitted program at these traditional colleges are high. In addition, while we do not observe the major at the time of graduation, we do know the system has significant barriers associated with changing majors. Students are not allowed to switch programs after enrollment without dropping out and re-applying through the centralized system (see Bordon and Fu (2015)). Finally, for most cohorts we observe whether students reapply and we utilize the last program degree observed in the data as the relevant one to assign a worker to a major.

insurance records on monthly earnings and estimate firm pay premiums using only men's earnings to reduce variability in hours worked. Also, we limit our analysis to workers between the ages of 20 and 60 and consider only the highest paying job for each worker every year. We call this sample, the *baseline sample*. We deflate all nominal variables so that they represent 2015 real dollars.

We build a second sample, which conditions on workers that we observe as admitted to a program degree between 1977 to 2000 and their earnings at least six years after the year of admission. We call this second sample the *college sample*. Thus, as a complement of the last sample, we define also the *non-college sample*.

A.4 Unemployment Insurance Data

As a robustness check, we use unemployment insurance (UI) data to measure earnings rather than IRS data. These data cover the year 2009 through 2016. This data set 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 IV, 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. We impute topcoded earnings following Card et al. (2013).[9] Finally, it excludes the public sector. Despite these limitations, the UI data has the potential advantage of identifying better job spells than the IRS data since it is updated every month.

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.

A.4.1 Imputing Topcoded Earnings

A key limitation of the UI 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

[9] Details on the imputation procedure are provided in Appendix A.

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.

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

B AKM Specification Checks

To validate the AKM earnings 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.

Panel A of Figure B1 plots monthly earnings 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 earnings 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.

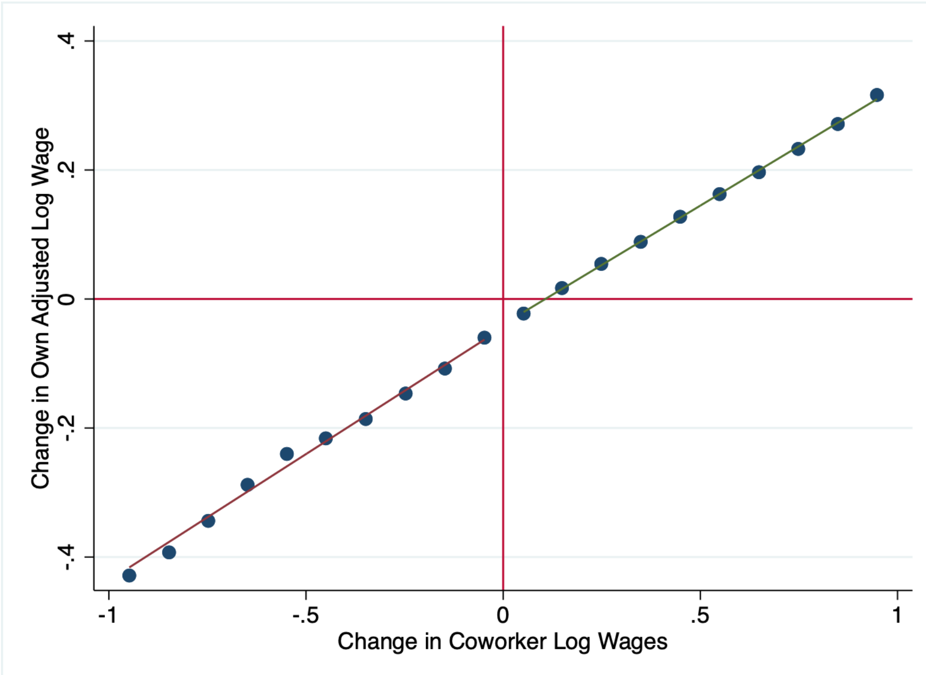
We next conduct a placebo test. We identify a subset of workers who are going to move in a future year but have not yet done so. If earnings trajectories are correlated with changes in firm pay premiums, we would expect some correlate between the change in coworker earnings between the worker's current and future firm, and the change in the worker's pre-move earnings. As shown in Panel B of Figure B1.

We conduct an analogous exercise where we characterize firms by their estimated firm effects rather than coworker earnings. Figure B2 plots monthly earnings changes for people moving from firm to firm on the vertical axis against the change in coworker mean earnings associated with the move on the horizontal axis. Each point corresponds to a decile of the earnings increase (or earnings 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.

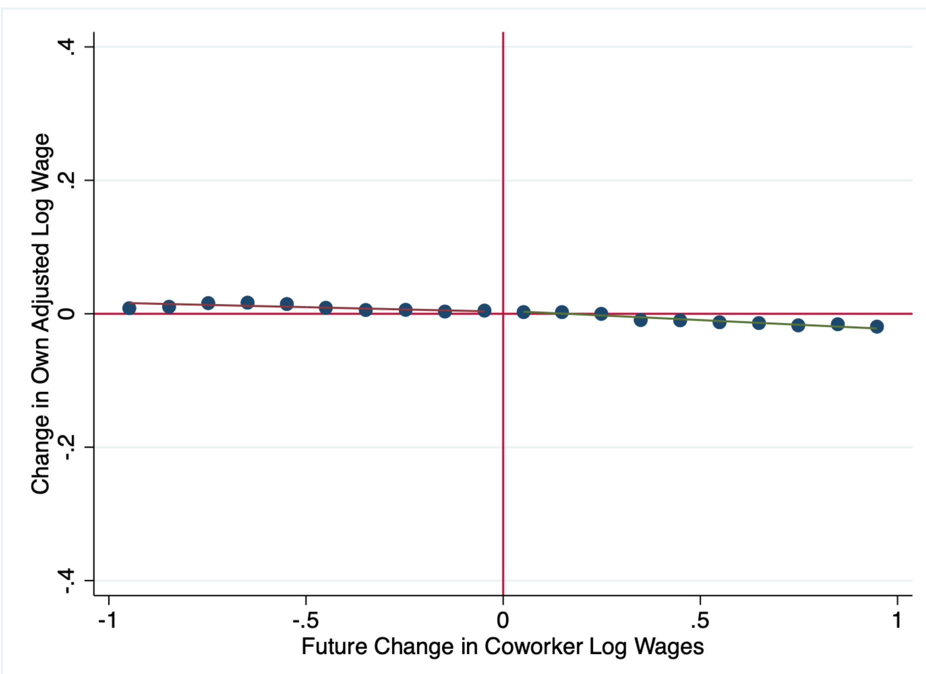
We also conduct a placebo test analogous to Panel B of Figure B1 in Figure B3. Again, we find no evidence of pre-trends in the earnings of workers that move to higher- or lower-paying firms.

Figure B1
Earnings Changes for Movers by Coworker Earnings

(a) Contemporaneous Moves



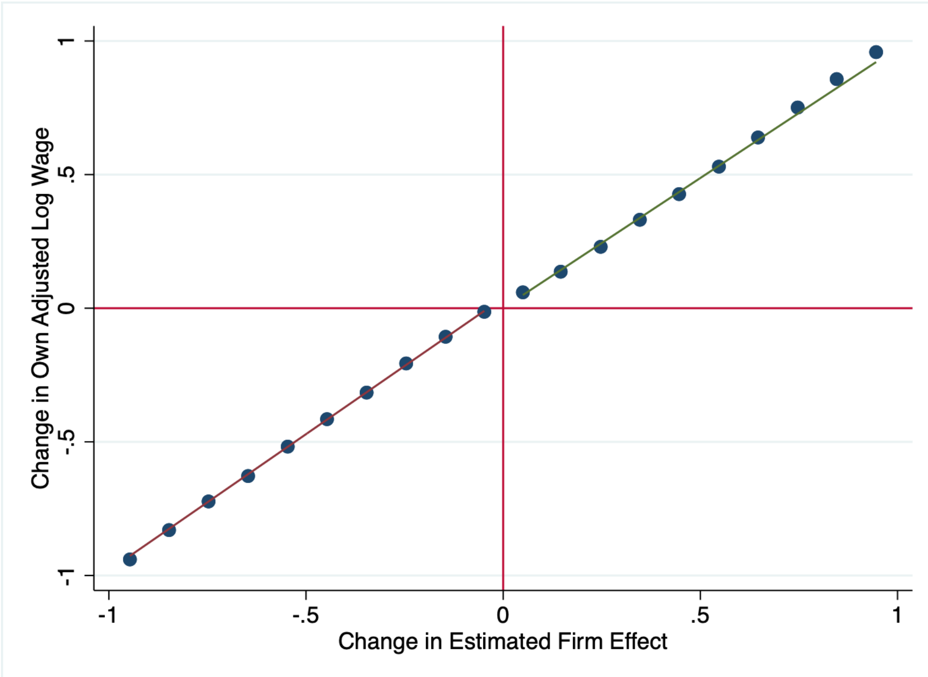
(b) Future Moves



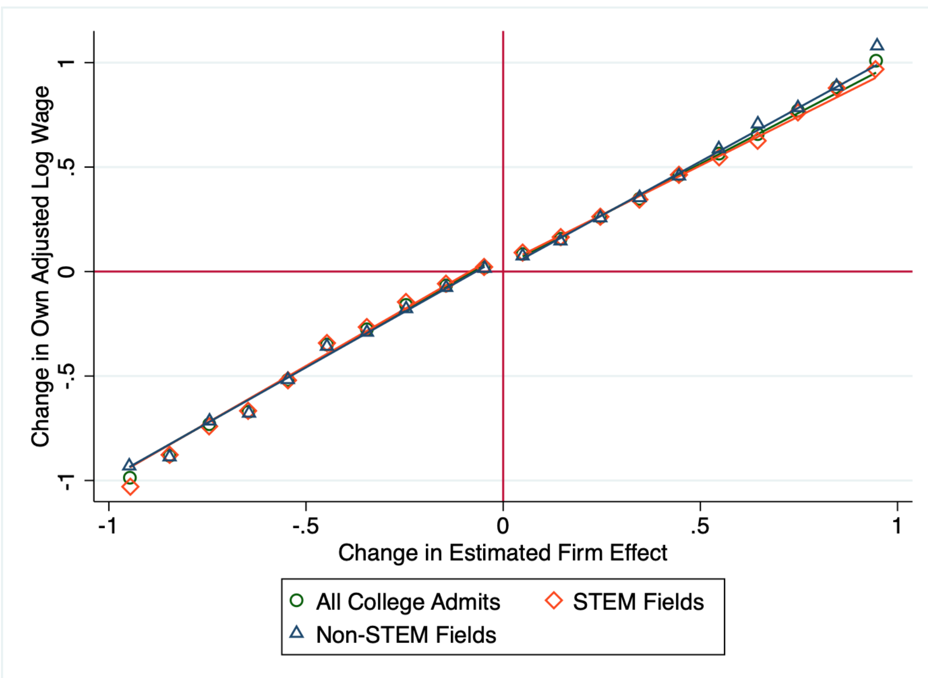
Note: This figure presents specification tests for the AKM model, motivated by Card et al. (2013). Panel A depicts change in own earnings for firm switchers (vertical axis) by the difference in coworker earnings at their old and new firms (horizontal axis). Panel A depicts contemporaneous changes in own earnings for *future* firm switchers (vertical axis) by the difference in coworker earnings at their current and future firms (horizontal axis).

Figure B2
Earnings Changes for Movers by Firm Effect

(a) All Workers



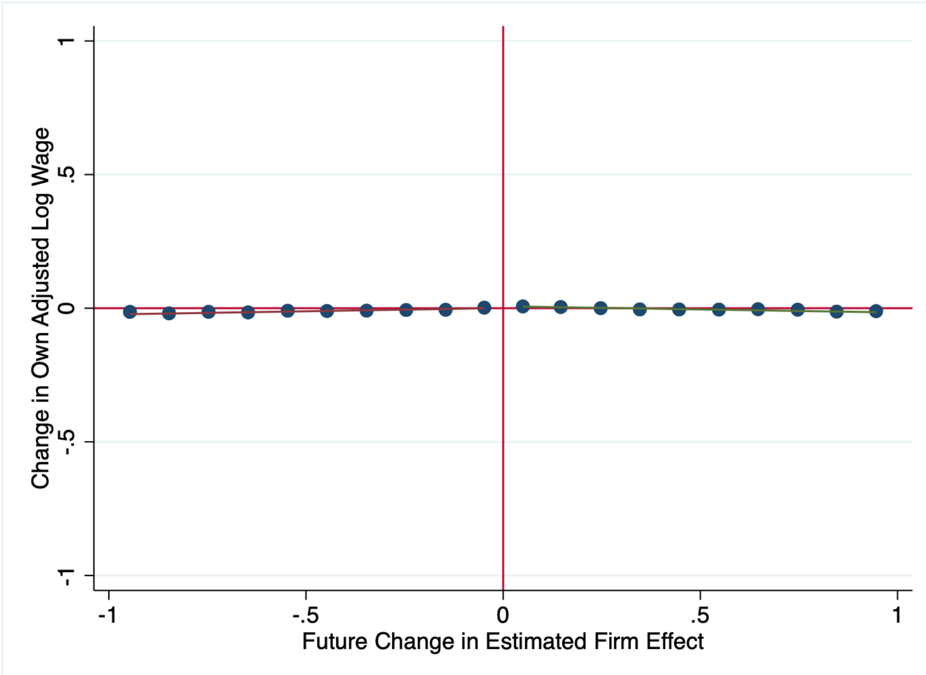
(b) College Admits



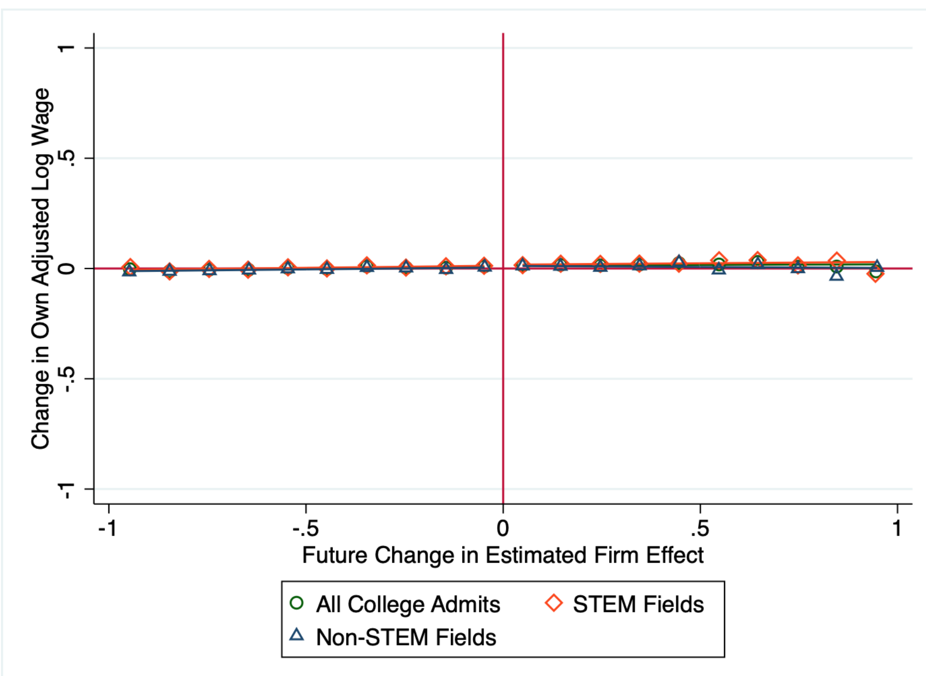
Note: This figure presents specification tests for the AKM model, motivated by Card et al. (2013). The figure depicts the change in own monthly earnings for firm switchers (vertical axis) by the difference in estimated firm effects at their old and new firms (horizontal axis). Panel A plots this relationship for all firm switchers. Panel B plots this relationship for firm switchers that are college admits, separately for all college admits, those admitted to STEM fields, and those admitted to non-STEM fields.

Figure B3
Earnings Changes for Future Movers by Firm Effect

(a) All Workers



(b) College Admits



Note: This figure presents specification tests for the AKM model, motivated by Card et al. (2013). The figure depicts changes in own monthly earnings for future firm switchers (vertical axis) by difference in estimated firm effects at old and new firm in future switch (horizontal axis). Panel A plots this relationship for all firm switchers. Panel B plots this relationship for firm switchers that are college admits, separately for all college admits, those admitted to STEM fields, and those admitted to non-STEM fields.

Table B1
Summary of AKM Estimates

	All	College Admits
Worker and firm parameters		
# of worker effects	8,022,664	457,143
# of firm effects	578,758	129,345
Summary of parameter estimates		
Std. dev. of worker effects	0.599	0.600
Std. dev. of firm effects	0.319	0.323
Correlation of worker/firm effects	0.243	0.166
Adjusted R^2	0.904	
Comparison match model		
Adjusted R^2	0.932	
Std. dev. of match effect	0.135	
Addendum		
Std. dev. log wages	0.806	0.802
Sample size	42,734,493	3,819,785

We drop wage observations where monthly wage is half of monthly minimum wage. Limited to largest connected set, which includes 98% of employment. Sample restrictions are described in further detail in Section I and Appendix Section A.

Table B2
Descriptive Statistics for
Explanatory Variables in Table III

	Men	Women
Experience	21.5	20.9
	(7.5)	(7.6)
Math Exam Score	664.8	622.4
	(89.4)	(92.9)
Verbal Exam Score	609.7	604.6
	(83.9)	(78.4)
Observations	2,081,719	1,734,347

This table presents the descriptive statistics for explanatory variables in Table III, potential work experience (years since last admissions exam) and math and verbal exam scores, separately by gender. Means are reported with standard deviations in parentheses.

C Alternative Modeling Approaches and Extensions

C.1 Using UI Data

We replicate the main tables using UI data: Table C2 summarizes earnings outcomes by major; Table C3 decomposes the gender earnings gap into firm and worker components; Table C4 decomposes the gender earnings gap using a regression approach; and Table C5 decomposes the gender earnings gap using the augmented model. Our main findings are unchanged.

C.2 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).

We assess whether limited mobility bias is a problem in our context using three approaches: the leave-out approach of Kline et al. (2020); the split-sample approach of Gerard et al. (2018); and the firm clustering approach of Bonhomme et al. (2019). We apply the first and third approaches in UI data rather than IRS data because, at the time of writing, these approaches have additional software requirements that we were unable to meet in the IRS computing environment. We apply the second approach in both the UI and IRS data. Across all three approaches, we find little evidence of limited mobility bias in our sample, perhaps because worker mobility is high in Chile.[10]

C.2.1 Kline et al. (2020) Leave-Out Approach

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. The baseline correlation between worker and firm effects is 0.3234; the leave-out correlation is 0.3244.

C.2.2 Gerard et al. (2018) Split-Sample Approach

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. We then compute split sample analogs for the covariance shares reported in Table I are not materially different from the baseline estimates. In the UI data, the full sample covariance share is 0.172; the split-sample covariance share is 0.177. In the IRS data, the full sample covariance share is 0.103; the split-sample covariance share is 0.124.

[10]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.

C.2.3 Bonhomme et al. (2019) Firm Clustering Approach

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):

$$\min_{k(1),...,k(J),H_1,...,H_k} \sum_{j=1}^{J} n_j \int \left(\hat{F}_j(y) - H_{k(j)}(y)\right)^2 d\mu(y), \tag{10}$$

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 $K = 10$.[11] 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 C6), (b) the firm component of the gender earnings gap is mostly a between-major phenomenon, with major accounting for between 63% and 81% of the firm component of the gender gap (Table C7), and (c) that major retains substantial explanatory power for the firm component of the gender gap even conditional on measures of worker ability (Table C8).

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, $\psi^{m(i)g(i)}_{J(i,t)}$:

$$\psi^{m(i)g(i)}_{J(i,t)} = \psi^{0}_{J(i,t)} + \sigma^{m}_{\tilde{s}(J(i,t))m(i)} + \sigma^{g}_{\tilde{s}(J(i,t))g(i))}. \tag{11}$$

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 C9 decomposes the gender earnings 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.

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

C.3 Variation in Hours

A key limitation of the IRS 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 C10 summarizes earnings outcomes by major using IRS data for men only. As in the main text, the covariance between firm and worker effects is much stronger between than within majors.

C.4 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

$$\begin{aligned}\mathrm{Cov}(\alpha_i + X_{it}'\beta, \psi_{J(i,t)}) =& \mathrm{Cov}(\alpha_i + X_{it}'\beta, \bar{\psi}_{s(J(i,t))} + \eta_{J(i,t)}) \\ =& \underbrace{\mathrm{Cov}(\alpha_i + X_{it}'\beta, \bar{\psi}_{s(J(i,t))})}_{\text{between industries}} + \underbrace{\mathrm{Cov}(\alpha_i + X_{it}'\beta, \eta_{J(i,t)})}_{\text{within industries}}, \end{aligned} \quad (12)$$

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.

Table C1
Summary of AKM Estimates, UI Data

	All	College Admits
Worker and firm parameters		
# of worker effects	7,487,592	413,563
# of firm effects	475,233	129,133
Summary of parameter estimates		
Std. dev. of worker effects	0.499	0.638
Std. dev. of firm effects	0.306	0.301
Correlation of worker/firm effects	0.287	0.307
Adjusted R^2	0.880	
Comparison match model		
Adjusted R^2	0.945	
Std. dev. of match effect	0.165	
Addendum		
Std. dev. log wages	0.703	0.845
Sample size	41,706,005	2,766,161

We drop wage observations where monthly wage is half of monthly minimum wage. Limited to largest connected set, which includes 99% of employment. Sample restrictions are described in further detail in Section A.4.

TABLE C2
EARNINGS OUTCOMES BY MAJOR, UI DATA

							Variance Decomposition			
Major	Male Share	Female Share	Log Earnings	Worker Component	Firm Component	Firm Ratio	Worker Share	Firm Share	Covariance Share	Residual Share
Overall							0.615	0.127	0.172	0.086
Between Major							0.513	0.086	0.387	0.014
Within Major							0.626	0.131	0.148	0.095
Agriculture	9.3	6.5	-0.121	-0.060	-0.061	0.503	0.670	0.123	0.112	0.095
Architecture and Art	2.1	3.1	-0.277	-0.195	-0.081	0.293	0.620	0.151	0.123	0.106
Business	11.7	14.8	0.144	0.133	0.010	0.070	0.669	0.111	0.137	0.083
Education	7.5	25.4	-0.505	-0.368	-0.134	0.265	0.630	0.142	0.118	0.110
Health	3.4	9.7	-0.003	0.029	-0.032	11.240	0.663	0.111	0.114	0.112
Humanities	1.7	4.6	-0.363	-0.290	-0.070	0.193	0.638	0.133	0.123	0.106
Law	2.2	2.2	0.036	0.080	-0.043	-1.182	0.639	0.108	0.174	0.079
Natural Science	5.3	9.7	-0.091	-0.091	0.001	-0.012	0.607	0.131	0.169	0.093
Social Science	2.2	5.0	-0.234	-0.156	-0.075	0.319	0.637	0.124	0.139	0.100
Technology	54.8	21.8	0.218	0.142	0.075	0.344	0.599	0.139	0.172	0.090

This table presents evidence of earnings outcomes by major using the college sample of the UI data described in Section A.4. Column 1 and 2 presents the 'Male Share' and 'Female Share', respectively, which refer to the percentage of male and female college admits in each major. Columns 3 through 5 present the average log earnings, worker effect and firm effect in each major, relative to average of each variable in the college sample. Column 6 presents the 'Firm Ratio', which is the ratio of the firm component of column 5 and the log earnings of column 3. Column 7 through 10 present the variance decomposition of log earnings into the worker effect, firm effect, covariance between worker and firm effect, and residual. Firm effects and worker effects are estimates from Equation (1), which is described in more detail in Section II. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in Section II. The between-field decomposition weights fields by their number of earnings observations.

Table C3
Contribution of Firm-Specific Pay Premiums and Major to the Gender Earnings Gap, UI Data

	Gender Earnings Gap	Firm Component	Worker Component
Overall	0.207	0.058	0.163
		(0.280)	(0.787)
Non-College Admits	0.206	0.057	0.163
		(0.277)	(0.791)
College Admits	0.346	0.086	0.264
		(0.249)	(0.763)
Between Major Area	0.176	0.054	0.120
		(0.307)	(0.682)
Residual	0.170	0.032	0.144
		(0.188)	(0.847)
By Major Area:			
Agriculture	0.273	0.031	0.241
Architecture and Art	0.101	0.016	0.093
Business	0.221	0.004	0.229
Education	0.074	0.040	0.050
Health	0.081	0.017	0.066
Humanities	0.021	-0.003	0.031
Law	0.076	0.007	0.071
Natural Science	0.243	0.055	0.196
Social Science	0.083	0.011	0.074
Technology	0.293	0.065	0.236
Between Specific Major	0.217	0.067	0.149
		(0.309)	(0.687)
Residual	0.128	0.019	0.115
		(0.148)	(0.898)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker earnings using firm and worker component estimates described in Section C.2.3 and the UI data. 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. The first row, Overall, uses the baseline employer-employee UI sample. The second (third) row, Non-College (College) admits, uses the non-college (college) employer-employee UI sample. The remaining rows, use the college employer-employee UI sample.

Table C4
Gender Earnings Gap and Firm Sorting, UI Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Log Earnings							
Male	0.325	0.215	0.045	0.176	0.124	0.012	0.116
	(0.003)	(0.003)	(0.001)	(0.003)	(0.003)	(0.001)	(0.003)
Outcome: Firm Effect							
Male	0.087	0.069	0.053	0.040	0.035	0.023	0.025
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome: Worker Effect							
Male	0.248	0.154		0.148	0.100		0.101
	(0.002)	(0.002)		(0.003)	(0.002)		(0.003)
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 using the college employer-employee UI sample described in Section A.4. 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 2,080,717 observations for 308,445 workers. Standard errors are clustered at the worker level. Descriptive statistics for potential work experience and exam scores by gender are reported in Appendix Table B2.

Table C5

Augmented Model Decomposition of the Gender Earnings Gap, UI Data

	(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	0.346	0.336	0.244	0.092	0.078	0.072	0.019	0.014
				(0.325)	(0.278)	(0.262)	(0.063)	(0.048)
Between Major Area	0.176			0.047	0.047	0.043	0.003	-0.001
				(0.313)	(0.328)	(0.291)	(0.022)	(-0.007)
Remainder	0.170			0.045	0.028	0.028	0.016	0.016
				(0.339)	(0.229)	(0.229)	(0.110)	(0.110)
Between Specific Major	0.201			0.059	0.059	0.055	0.004	-0.000
				(0.331)	(0.331)	(0.307)	(0.018)	(-0.000)
Remainder	0.118			0.032	0.016	0.016	0.015	0.015
				(0.314)	(0.180)	(0.180)	(0.146)	(0.135)
By Major Area:								
Agriculture	0.273	0.318	0.255	0.063	-0.017	0.075	-0.012	0.080
Architecture and Art	0.101	0.324	0.217	0.107	-0.082	0.025	0.081	0.189
Business	0.221	0.279	0.263	0.016	0.043	-0.032	0.048	-0.027
Education	0.074	0.191	0.167	0.024	0.038	0.037	-0.012	-0.014
Health	0.081	0.272	0.289	-0.017	0.035	0.034	-0.051	-0.052
Humanities	0.021	0.284	0.192	0.092	-0.053	-0.044	0.135	0.145
Law	0.076	0.321	0.276	0.045	-0.052	0.036	0.010	0.098
Natural Science	0.243	0.380	0.270	0.110	0.004	0.055	0.055	0.107
Social Science	0.083	0.260	0.275	-0.014	-0.017	0.060	-0.075	0.003
Technology	0.293	0.376	0.295	0.081	0.076	0.048	0.033	0.005

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section III.C. Estimates are derived using UI data. 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 III.C . Column 4 reports the total contribution of firm-specific earnings 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 III.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. The first row, Overall, uses the baseline employer-employee UI sample. The remaining rows, use the college employer-employee UI sample.

Table C6
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							0.669	0.069	0.174	0.088
Between Major							0.638	0.046	0.304	0.012
Within Major							0.672	0.071	0.158	0.099
Agriculture	10.1	6.7	-0.087	-0.031	-0.055	0.629	0.712	0.065	0.124	0.099
Architecture and Art	1.9	2.8	-0.303	-0.251	-0.050	0.165	0.666	0.075	0.149	0.110
Business	12.0	14.6	0.175	0.162	0.011	0.062	0.712	0.058	0.146	0.084
Education	8.1	27.6	-0.488	-0.392	-0.093	0.191	0.661	0.070	0.151	0.118
Health	3.8	10.6	0.019	0.050	-0.031	-1.626	0.722	0.051	0.109	0.118
Humanities	1.7	4.5	-0.352	-0.316	-0.033	0.191	0.658	0.066	0.161	0.115
Law	1.8	1.6	-0.014	0.006	-0.017	1.248	0.680	0.054	0.186	0.080
Natural Science	5.3	6.8	-0.086	-0.091	0.007	-0.082	0.651	0.071	0.182	0.096
Social Science	2.1	4.8	-0.239	-0.193	-0.043	0.179	0.673	0.062	0.157	0.108
Technology	53.3	20.0	0.221	0.162	0.057	0.259	0.651	0.081	0.176	0.092

This table presents evidence of earnings outcomes by major using the college employer-employee UI sample described in Section I and the BLM estimation method described in Section C.2.3. Column 1 and 2 presents the 'Male Share' and 'Female Share', respectively, which refer to the percentage of male and female college admits in each major. Columns 3 through 5 present the average log earnings, worker effect, and firm effect in each major, relative to the average of each variable in the college sample. Column 6 presents the 'Firm Ratio', which is the ratio of the firm component of column 5 and the log earnings of column 3. Columns 7 through 10 present the variance decomposition of log earnings into the worker effect, firm effect, covariance between worker and firm effect, and residual. Firm effects and worker effects are estimates from Equation (1), which is described in more detail in Section II. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in Section II. The between-field decomposition weights fields by their number of earnings observations.

Table C7
Contribution of Firm-Specific Pay Premiums and Major to the Gender Earnings Gap, BLM

	Gender Earnings Gap	Firm Component	Worker Component
Overall	0.198	0.057	0.155
		(0.288)	(0.783)
Non-College Admits	0.200	0.059	0.157
		(0.295)	(0.785)
College Admits	0.350	0.063	0.292
		(0.180)	(0.834)
Between Major Area	0.180	0.040	0.138
		(0.222)	(0.767)
Residual	0.169	0.023	0.154
		(0.136)	(0.911)
By Major Area:			
Agriculture	0.282	0.016	0.267
Architecture and Art	0.089	0.001	0.098
Business	0.222	0.002	0.233
Education	0.071	0.023	0.065
Health	0.087	0.022	0.067
Humanities	0.020	-0.003	0.032
Law	0.091	0.002	0.097
Natural Science	0.245	0.044	0.209
Social Science	0.086	0.009	0.083
Technology	0.305	0.053	0.259
Between Specific Major	0.221	0.051	0.169
		(0.231)	(0.765)
Residual	0.129	0.012	0.123
		(0.093)	(0.953)

This table decomposes the gender earnings gap for subgroups of workers into firm and worker earnings components as described in Section III.B using firm and worker component estimates described in Section C.2.3. 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. The first row, Overall, uses the baseline employer-employee UI sample. The second (third) row, Non-College (College) admits, uses the non-college (college) employer-employee UI sample. The remaining rows, use the college employer-employee UI sample.

Table C8
Gender Earnings Gap and Firm Sorting, BLM

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome: Log Earnings							
Male	0.329	0.219	0.022	0.179	0.126	0.001	0.119
	(0.003)	(0.003)	(0.001)	(0.003)	(0.003)	(0.001)	(0.003)
Outcome: Firm Effect							
Male	0.063	0.048	0.031	0.028	0.023	0.012	0.014
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome: Worker Effect							
Male	0.276	0.180		0.163	0.115		0.115
	(0.003)	(0.003)		(0.003)	(0.003)		(0.003)
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 using the college employer-employee UI sample described in Section I. 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 C9
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	0.346	0.336	0.244	0.092	0.078	0.072	0.014	0.019
				(0.285)	(0.238)	(0.223)	(0.047)	(0.060)
Between Major Area	0.176			0.047	0.047	0.043	-0.001	0.003
				(0.286)	(0.292)	(0.267)	(-0.006)	(0.019)
Remainder	0.170			0.044	0.029	0.028	0.016	0.016
				(0.287)	(0.185)	(0.178)	(0.102)	(0.102)
Between Specific Major	0.201			0.059	0.059	0.055	-0.000	0.004
				(0.293)	(0.293)	(0.273)	(-0.000)	(0.020)
Remainder	0.118			0.032	0.017	0.016	0.015	0.015
				(0.271)	(0.144)	(0.136)	(0.127)	(0.137)
By Major Area:								
Agriculture	0.273	0.318	0.255	0.063	-0.017	0.075	0.080	-0.012
Architecture and Art	0.102	0.324	0.217	0.107	-0.082	0.025	0.189	0.081
Business	0.222	0.279	0.263	0.016	0.043	-0.032	-0.027	0.048
Education	0.075	0.191	0.167	0.024	0.038	0.037	-0.014	-0.012
Health	0.082	0.272	0.289	-0.017	0.035	0.034	-0.052	-0.051
Humanities	0.021	0.284	0.192	0.092	-0.053	-0.043	0.145	0.135
Law	0.076	0.321	0.276	0.045	-0.053	0.036	0.098	0.010
Natural Science	0.243	0.380	0.270	0.110	0.004	0.055	0.107	0.055
Social Science	0.083	0.260	0.275	-0.014	-0.017	0.064	0.003	-0.075
Technology	0.293	0.376	0.295	0.081	0.076	0.048	0.005	0.033

This table decomposes the gender earnings gap for subgroups of workers into firm and worker components as described in Section C.2.3. Estimates are derived using UI data. 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 C.2.3 . Column 4 reports the total contribution of firm-specific earnings 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 III.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. The first row, Overall, uses the baseline employer-employee UI sample. The remaining rows, use the college employer-employee UI sample.

Table C10
Earnings Outcomes by Major, Men Only

					Variance Decomposition			
Major	Log Earnings	Worker Component	Firm Component	Firm Ratio	Worker Share	Firm Share	Covariance Share	Residual Share
Overall					0.639	0.196	0.075	0.091
Between Major					0.544	0.152	0.319	-0.015
Within Major					0.645	0.196	0.061	0.098
Agriculture	-0.130	-0.032	-0.095	0.731	0.712	0.166	0.015	0.106
Architecture and Art	-0.227	-0.128	-0.102	0.447	0.669	0.199	0.025	0.107
Business	0.133	0.131	0.000	0.003	0.682	0.176	0.056	0.085
Education	-0.405	-0.293	-0.118	0.291	0.648	0.186	0.065	0.102
Health	0.138	0.205	-0.068	-0.493	0.725	0.131	-0.018	0.162
Humanities	-0.320	-0.258	-0.061	0.192	0.641	0.168	0.083	0.108
Law	0.183	0.127	0.061	0.333	0.629	0.180	0.080	0.112
Natural Science	-0.087	-0.079	-0.008	0.094	0.625	0.208	0.070	0.098
Social Science	-0.124	-0.092	-0.031	0.248	0.646	0.171	0.066	0.117
Technology	0.098	0.037	0.062	0.637	0.612	0.221	0.080	0.087

This table presents evidence of earnings outcomes by major using the college sample described in Section I. Column 1 presents the 'Male Share', which refer to the percentage of male college admits in each major. Columns 2 through 4 present the average log earnings, worker effect, and firm effect in each major, relative to average of each variable in the college sample. Column 5 presents the 'Firm Ratio', which is the ratio of the firm component of column 4 and the log earnings of column 2. Columns 6 through 9 present the variance decomposition of log earnings into the worker effect, firm effect, covariance between worker and firm effect, and residual. Firm effects and worker effects are estimates from Equation (1), which is described in more detail in Section II, using men only. Earnings outcomes are demeaned. We describe the decomposition of the variance in earnings in more detail in Section II. The between-field decomposition weights fields by their number of earnings observations.