The Gender Gap: Micro Sources and Macro Consequences*

Iacopo Morchio† Christian Moser‡

June 15th, 2019

Abstract

We investigate the sources of the gender wage gap and its relation to firm heterogeneity. We document a gender wage gap of 20 log points conditional on education interacted with experience, state, industry, and occupation among workers in Brazil. Accounting for unobservable worker and firm heterogeneity, we find that around 46 percent of the residual gender wage gap is between firms, while the remainder is within firms. We highlight lower labor market mobility of women relative to men as an important explanatory factor for pay differences both within and between firms. We develop an equilibrium search model with firm productivity differences, worker ability differences, gender-specific amenities, and employer taste for discrimination. We use the estimated model to show that gender differences in life-cycle mobility across employers are a major contributor to the observed gaps and associated with sizable negative consequences for macroeconomic outcomes such as aggregate productivity, employment, and output.

Keywords: Firm Pay Differences, Worker and Employer Heterogeneity, Linked Employer-Employee Data, Income Inequality, Misallocation

JEL classification:E24, E25, J16, J31, J71

*We thank John Donaldson, David Dorn, Gita Gopinath, Rasmus Lentz, Espen Moen, Ben Moll, Andreas Mueller, Emi Nakamura, Jesse Schreger, Jón Steinsson, Pierre Yared, Steve Zeldes, participants at numerous conferences and workshops, and seminar participants at Columbia University and University of Vienna for useful comments. Moser gratefully acknowledges financial support from the Ewing Marion Kauffman Foundation. All errors are our own.

†Department of Economics, University of Vienna. E-mail: iacopo.morchio@univie.ac.at.

‡Graduate School of Business, Columbia University. E-mail: c.moser@columbia.edu.

1 Introduction

A long literature has studied gaps in employment and pay across genders in a variety of contexts.[1] A significant share of raw pay differences has been attributed to gender differences in contractual work hours, labor market attachment over the life cycle, the field of job training, selection across industries, and occupational choice. Yet substantial gender gaps remain even after controlling for these covariates. Taken at face value, these results suggest that observationally equivalent workers of different gender see significant discrepancies in pay, with women being systematically paid less and having lower propensity to participate in the labor market.

A parallel literature has highlighted the role of employer heterogeneity in explaining pay differences across identical individuals (Abowd et al., 1999a). Yet little progress has been made in linking the gender gap to the role of firms in the labor market. This is partly due to limited data availability combining the most relevant information on human capital and the firm, leaving as an open question what are the microeconomic sources of existing gender gaps. Another shortcoming of the existing literature is that the macroeconomic consequences of existing differences at the micro level are largely unexplored. The reason for this is that aggregating up observed differences at the individual level requires a structural interpretation of the empirical patterns, which to date has been elusive.

In this paper, we aim to fill both of these gaps. On a micro level, we ask: what are determinants of gender gaps in pay and participation? We then ask on a macro level: what are output and welfare gains from re-allocating women? To this end, we study the labor market experience of the universe of formally employed men and women in Brazil. Using administrative linked employer-employee data, we document new facts regarding the role of gender in the workplace, both within and between firms. We then investigate to what extent the observed gender gaps in pay and participation can be explained by observable differences in the labor market experience of men compared to women. Our findings are directly informative for a structural assessment of the macro consequences of the observed gender gaps.

We document a gender wage gap of 20 log points conditional on education interacted with experience, state, industry, and occupation. Accounting for unobservable worker and firm heterogeneity, we find that around 46 percent of the residual gender wage gap is between firms, while

[1] See the literature contained in recent work by Blau and Kahn (2017) and Kleven et al. (2016b).

the remainder is within firms. We highlight lower labor market mobility of women relative to men as an important explanatory factor for pay differences both within and between firms. Finally, we discuss the consequences of gender gaps for macroeconomic variables including aggregate productivity, employment, and output.

To interpret the differences across genders in labor market outcomes within across employers, we develop an on-the-job search model that allows for several different sources of gender gaps. In particular, our model allows for differences in firm-specific amenities and taste-based discrimination against women as potential sources of within-firm gender gaps, and for different probabilities of transitioning between labor market states as additional sources of pay differences across genders. We then calibrate the model to fit the data in order to quantify the role of each source to explain gender gaps.

A common argument against the existence of taste-based discrimination against a population subgroup is that, in a competitive market, such firms would not survive and would be pushed out of the market, making it impossible for discrimination to survive in equilibrium (Becker, 1957). An important implication of our frictional model is that these firms can survive because of the existence of frictions, by operating at a smaller scale than an otherwise identical but non-discriminatory firm. When women look for jobs, they might have to accept jobs at discriminatory firms as stop-gaps, waiting to move to better paying ones.

The second important implication of our model is that the presence of employer preferences that differ across genders has general equilibrium effects. It is not necessary that all firms discriminate for gender gaps to be present in the economy at all jobs. This is because non-discriminating firms know that they are competing for attracting women with discriminating firms, so that they will experience less competition and be able to offer lower wages to women. This also means that there can be far-reaching consequences from changing the behavior of some firms in the economy, as this would increase competition across firms in the labor market.

Although we confirm that large gender gaps exist in employment and pay, our modeling approach suggests that for welfare analysis it is crucial to understand for whether gender pay gaps stem from differences in amenities or discrimination at the firm level. The reason is that, if gender gaps in monetary wages stem from differences in amenity levels with respect to men, removing these gaps might be inefficient as some firms might find themselves unable to pay high monetary wages, while workers value amenities. An immediate consequence of this insight is that output

gains induced by a policy change generally do not directly correspond to welfare gains. Our model with several dimensions of employer heterogeneity allows us to disentangle these sources of observed gender gaps, and therefore to understand the welfare consequences of different policies that we consider in counterfactual simulation exercises.

We find that all three sources of gender gaps—heterogeneity in amenities, discrimination, and mobility—are quantitatively important. In particular, unobserved amenities account for roughly two-thirds of the average gender gap. We also find that, in a world in which women moved from job to job as fast as men, they would in fact be paid more than men, as our baseline calibration implies that women enjoy lower levels of amenities, so that they would have to be compensated by monetary wages instead.

Related literature. There is a long tradition in the microeconomic literature of studying gender gaps (Bertrand et al., 2010; Guvenen et al., 2014; Kleven et al., 2016a; Blau and Kahn, 2017; Adda et al., 2017). A more recent literature has independently highlighted the role of firm heterogeneity as an important determinant of observed pay differences Abowd et al. (1999a); Card et al. (2013); Barth et al. (2016); Song et al. (2016); Alvarez et al. (2018). Little work has bridged the gap between these two strands of the literature, a notable exception being Card et al. (2016).

The macroeconomic literature has stressed the importance of misallocation of capital (Hsieh and Klenow, 2009) and the equilibrium effects of labor market frictions (Burdett and Mortensen, 1998). Recent related work includes Guner et al. (2012); Erosa et al. (2016); Ngai and Petrongolo (2017); Bagger and Lentz (2018) and Erosa et al. (2017), Less attention been paid to the role of the misallocation of human capital embedded in workers of each gender. On this front, the work by Hsieh et al. (2016) is closely related to what we do. An important difference is that, in line with the misallocation literature, we take a firm-level perspective on the experience of workers of both genders in the labor market.

Our contribution is to bridge the microeconomic and macroeconomic perspectives of gender differences in the labor market. Using a large administrative dataset, which contains detailed information on all formal workers in Brazil, we connect differences in pay and employment at the micro level to implied differences in output and efficiency at the macro level.

Outline. Section 2 introduces the administrative linked employer-employee data from Brazil, which lies at the heart of our investigation. Section 3 presents stylized facts on gender gaps at the micro level. Section 4 sets up the model we use for our quantititative analysis. Section 5 discusses the identification of the model parameters. Section 6 outlines the results of our estimation and the implications of the model. Section 7 relates our results to the gender-specific impact of parental leave from the labor market, the racial pay gap, and the gender gap in other countries. Finally, Section 8 concludes.

2 Data

Our main data source is the *Relação Anual de Informações Sociais (RAIS)*, a linked employer-employee register by the Brazilian Ministry of Labor and Employment. Firms' survey response is mandatory, and misreporting is deterred through audits and threat of fines. Collection started in 1986, with coverage becoming near universal from 1994 onward. The data contain detailed information on job characteristics, with 73 million formal sector employment spells recorded in 2012. Although reports are annual, we observe for every job spell the date of accession and separation in addition to average monthly earnings. We keep for each worker the highest-paid among each year's longest employment spells. Our baseline analysis uses average monthly earnings in formal employment ("earnings"), although we also construct a measure of hourly wages as earnings divided by the number of contractual hours ("wages").

In what follows, we restrict attention to male and female workers between age 18 and 64 with valid wage information. We keep for each individual their highest-paid among all last jobs in the calendar year. Table 1 contains a list of summary statistics of labor market outcomes for this subpopulation recorded in the RAIS data for the year 2003. Male and female workers share similar average age of 34.1 compared to 34.3 years, and an urban employment share of 42 compared to 41 percent. Women are significantly more educated than men, with the latter falling short by over one year of schooling on average, with 5.3 compared to 6.4 years. Conditional on participating in the formal sector labor market, men and women spend a similar amount of time at work, 1,688 compared to 1,661 hours per year. This is the case inspite of men being more likely to work full time, that is 40 or 44 hours, compared to women.[2] Earnings for men is on average 15 log points

[2]The last two facts are reconciled by men switching jobs more often, a labor market outcome that we will examine in detail in the following analysis.

higher than that for women, 7.20 compared to 7.05 log Brazilian Reais (BRL). A smaller gender gap of 8 log points exists for wages, with 3.47 compared to 3.39 log BRL. Finally, around 62 percent of all workers in the RAIS data are male in 2003, exposing a large gender disparity in labor force participation.

Table 1. Summary statistics on labor market outcomes for male and female workers

	Males		Females	
	Mean	St.d.	Mean	St.d.
Age (years)	34.11	11.12	34.28	10.83
Education (years)	5.31	2.05	6.36	1.95
Share urban	0.42		0.41	
Hours worked per year	1,688	688	1,661	666
Hours worked per week	42.36	4.60	39.95	7.20
Monthly earnings (log BRL)	7.20	0.79	7.05	0.76
Hourly wage (log BRL)	3.47	0.83	3.39	0.82
# Observations (millions)	25.93		16.03	
# Unique workers (millions)	22.07		13.86	
Labor force share	0.62		0.38	
# Unique employers (millions)	1.09		0.79	

Notes: 2003 Brazilian formal sector census (RAIS) data. "Urban" = RJ, SP, DF.

3 Firms and Gender Gaps in Employment and Pay

3.1 Descriptive evidence

As a first step, we now turn to a descriptive analysis of empirical gender gaps in the raw RAIS data from Brazil. We define the female pay gap as log male pay minus log female pay. We pick as the appropriate pay concept either monthly earnings or hourly wages. Table 2 summarizes our results. We find that the mean female earnings gap is 15 log points. Comparing earnings at different percentiles of the gender-specific earnings distributions, we find that the female earnings gap is strictly increasing toward higher earnings percentiles. For example, workers at the 5th percentile of the male earnings distribution receive 3 log points more than workers at the 5th percentile of the female earnings distribution. This gap grows to over 17 log points at the median of the gender-specific distributions, and reaches 22 log points at the 95th earnings percentile. The

female wage gap is of a smaller magnitude throughout the distribution, with a mean gap of 8 log points.

Table 2. Raw gender pay gaps

	Male	Female	Pay gap
Panel A. Monthly earnings			
Mean	7.205	7.053	0.152
P5	6.251	6.221	0.030
P10	6.349	6.289	0.060
P25	6.640	6.512	0.128
P50	7.035	6.863	0.172
P75	7.605	7.430	0.175
P90	8.286	8.137	0.149
P95	8.786	8.565	0.221
Panel B. Hourly wage			
Mean	3.469	3.391	0.078

Notes: Brazilian formal sector census data (RAIS) for 2003.

One may naturally suspect that some of the observed gender pay gaps are due to women exhibiting differences in pay-relevant dimensions. Confirming this hypothesis, Table 3 shows female labor force participation (FLFP) and pay gaps across 1-digit industries, 1-digit occupations, and the 27 states of Brazil. Panel A shows that, for example, only around 9 percent of all workers in the construction sector are female while women constitute 74 percent of all workers in health services. At the same time, the female earnings gap in construction is -12 log points, meaning that participating women get paid more than men in the same sector, while in health services the gap is 21 log points. Consequently, the fact that women are disproportionately allocated toward higher pay gap industries further exacerbates the overall pay gap. Substantial heterogeneity also exists in the female labor force participation rates across occupations and, to a lesser degree, across states, although pay gaps vary less systematically with the gender ratio along these dimensions.

Table 3. Female labor force participation and pay gaps

Lowest FLPF	FLFP	Earnings gap	Highest FLFP	FLFP	Earnings gap
		Panel A. Industries			
Construction	0.088	-0.122	Health Services	0.740	0.207
Transportation	0.133	0.086	International	0.616	0.297
Nond. manuf.	0.163	0.253	Finance	0.446	0.187
		Panel B. Occupations			
Farmer	0.167	0.208	Public worker	0.665	0.418
Production line	0.206	0.375	Administrator	0.466	0.157
Maintenance	0.210	0.561	Mid-technical	0.466	0.346
		Panel C. States			
Mato Grosso	0.337	0.128	Acre	0.464	-0.013
Alaoas	0.342	-0.002	Piauí	0.457	0.037
Pará	0.357	0.061	Ceará	0.440	0.155

Notes: Brazilian formal sector census data (RAIS) for 2003.

3.2 Regression analysis

As a next step, we control for observed differences in pay-relevant worker and job characteristics in the spirit of Mincer (1974). We then compute the "residual female pay gap" controlling for these other worker and job characteristics. Our baseline regression specification for earnings of individul i in year t, denoted y_{it}, is as follows:

$$y_{it} = \alpha^{base} - \mathbf{1}\left[female_i\right]\alpha^{gap} + X_{it}\beta + \varepsilon_{it} \tag{1}$$

where α^{base} is the male pay intercept, $\mathbf{1}\left[female_i\right]$ is an indicator that equals 1 if individual i is female and 0 otherwise, α^{gap} is the conditional female pay gap, X_{it} contains other observable worker and characteristics, and ε_{it} is a residual term. We estimate equation (1) via ordinary least squares (OLS) by imposing the usual strict exogeneity assumption on ε_{it}, namely $\mathbb{E}\left[\varepsilon_{it}|female_i, X_{it}\right] = 0$. We are interested in the estimated gender pay gap, $\hat{\alpha}^{gap}$, which captures the pay discount between a female and a male worker who are otherwise observationally identical.

While we showed above that the raw earnings gap was 15 log points, we find that this gap widens significantly in the regression analysis. If we include in X_{it} the classical Mincerian controls—a linear term in years of education, and linear and quadratic terms in years of potential

experience (constructed as age minus years of education minus six)—we estimate a residual earnings gap of 32 log points. The fact that the residual female earnings gap exceeds the raw earnings gap reflects the fact that females are more likely to possess high-paying characteristics such as education relative to men, yet get paid less in the data.

Controlling for a richer set of worker and job observable characteristics in X_{it}—including years of education dummies interacted with potential experience dummies, state dummies, industry dummies, and occupation dummies—we estimate the female earnings gap to be 22 log points . With the same controls, the corresponding female wage gap amounts to 20 log points, indicating that most differences in female hours worked are correlated with the included covariates, most notably industry and occupation codes.

We conclude that large gender pay gaps remain within narrowly defined subgroups defined by classical pay characteristics, including education and potential experience, state, industry, and occupation.

3.3 Gender segregation across employers

A recent literature highlights the important role of firm heterogeneity in explaining pay dispersion (Abowd et al., 1999b; Card et al., 2013, 2016; Song et al., 2016; Barth et al., 2016; Alvarez et al., 2018) for otherwise identical workers. The fact that large gender pay gaps remain after controlling for rich observable characteristics begs the question: could firm heterogeneity also play a role in explaining the residual gender pay gap described above? While the aforementioned literature argues that the workplace is an important unit of analysis in understanding pay differences across individuals, the firm also seems to be an important actor in the context of productivity and output differences at a macroeconomic level. To the extent that women work at lower-paying and less productive firms, this may explain both the female pay gap and also account for efficiency losses from the misallocation of talent across production units.

Before we turn to a formal analysis of the contribution of firms toward the female pay gap, we investigate how unequally women are distributed across firms. While women make up around 40 percent of Brazil's formal setor labor force in 2003, it is unclear to what extent women are represented in equal proportion across all firms. To assess the extent to which women are distributed (un-) equally across firms, we construct the following firm segregation index for our population

with N_m males and N_f females:

$$S^{firm} = \frac{\sum_{i=1}^{N_m+N_f} \left(\text{female ratio}_{j(i)} - \text{pop. female ratio}\right)^2}{N_m \times (\text{pop. female ratio})^2 + N_f \times (1 - \text{pop. female ratio})^2}$$

where the firm segregation index S^{firm} lies between 0, in the case of each firm featuring a representative share of women, and 1, in the case of all women being concentrated in firms where only women work.

In constructing the firm segregation index, S^{firm}, on the Brazilian RAIS data from 2003, we find the index to take a value of 0.349, which we interpret as economically large. To give a sense of its magnitude, this index corresponds to an average firm-level gender share difference of ± 0.288. As another means of comparison, we find that the same index is signifcantly smaller when computed across industries (0.109), occupations (0.142), or states (0.002). Furthermore, the index value we compute is robust to restricting attention to larger firms for which one can expect the granularity of small firm sizes to be less relevant. We conclude that women are distributed very unequally across firms, with some firms employing far more women than their labor force participation ratio warrants, and others employing far less.

Next, we investigate to what extent women work at lower-paying firms. To this end, we run the following regression to explain average earnings among workers employed by firm j at time t,denoted $\overline{y}_{jt}$:

$$\overline{y}_{jt} = \text{female_employment_share}_{jt} \times \gamma + X_{jt}\beta + \varepsilon_{jt} \tag{2}$$

where $\text{female_employment_share}_{jt}$ average female employment share at that firm in the same year, plus a set of firm-level controls, X_{jt}, which include the firm's industry, occupation structure, and state fixed effects, plus firm-level averages of its workforce's education and epxerience. We estimate equation 2 via OLS, weighted by firm-level employment.

We find that firms with a higher female worker share pay significantly less, as the regression results summarized in Table 4 show. Column 1 of the table shows that the semi-elasticity of mean log earnings with respect to women's employment share is -0.201, significant at the 1 percent level, meaning that an all female firm pays 20 percent less relative to an all male firm. We further decompose this female pay penalty into what we term a "composition effect" and a "net sorting effect." The composition effect captures the share of the mean pay difference explained by women being

paid less on average then men and women are sorted unequally across firms. The net sorting effect then captures the share of the mean pay difference due to the fact that firms employing more women pay less, regardless of who works there. The net sorting effect accounts for a around one quarter of the total pay gap due to the allocation of women across firms in our baseline specification, and up to 45 percent in our alternative specifications controlling for other worker and firm observables. These results suggests that employer identity matters for the female pay gap.

Table 4. Regression of women's average pay on women's employment share

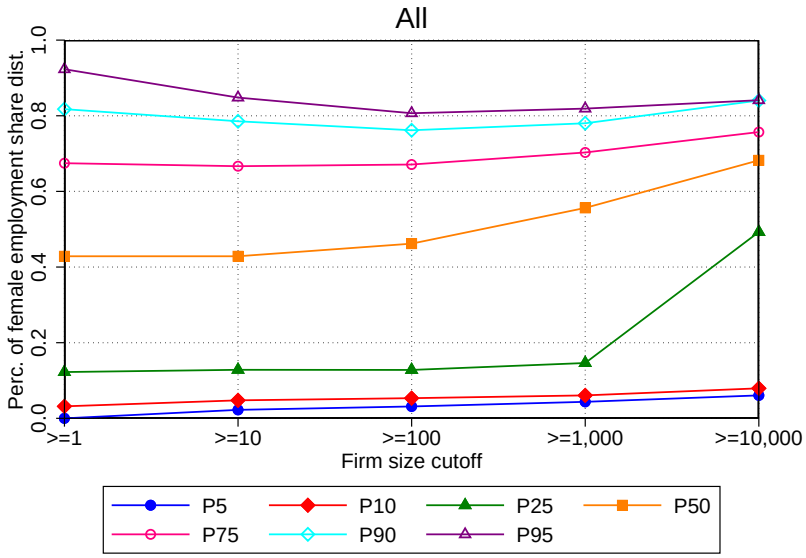
	(1)	(2)	(3)
Women's employment share	-0.201	-0.534	-0.533
Industry, occupation, state FEs	no	yes	yes
Education×experience FEs	no	no	yes
# Observations (millions)	35.9	35.3	35.3
R^2	0.01	0.37	0.44
Composition effect	-0.152	-0.294	-0.300
Net sorting effect	-0.049	-0.240	-0.233

Notes: Brazilian formal sector census data (RAIS) for 2003. Dependent variable is mean employer earnings. Weighted by number of employees. All coefficients significant at 1% level. Composition effect is shifting from all male to all female employees. Net sorting effect is pay difference net of composition effect. Weighted (unweighted) IQR of female emp. share is 0.514 (0.833).

As another way of seeing how unequal the distribution of genders across employers is, Figure 1 shows percentiles of the female employment share distribution across firms for various employment size cutoffs. With female workers making up around 40 percent of all formal workers, a perfectly equal employment allocation would require 38 percent female workers in every firm. Due to indivisibility of bodies, however, this ratio could not be achieved exactly at small firms. At large firms, however, this ratio could be either achieved or closely approximated. Indeed, empirically we see that we are quite far away from the equal-allocation benchmark, even at very large firms with more than 10,000 employees. In the data, the female employment share varies between 9 percent at the 10th percentile and 83 percent at the 90th percentile, even among the largest firm size group.

To emphasize the unequal distribution of women across employers, Figure 2 shows a histogram of female employment shares (unweighted, or at the firm level) in bins spanning intervals

Figure 1. Percentiles of the female employment share distribution across firms for various employment size cutoffs



Notes: Brazilian formal sector census data (RAIS) for 2012.

of 10 percent. The histogram shows two pronounced spikes of mass for the categories 0-10% and 90-100%, indicating that a large fraction of men and women are employed at gender-segregated firms.

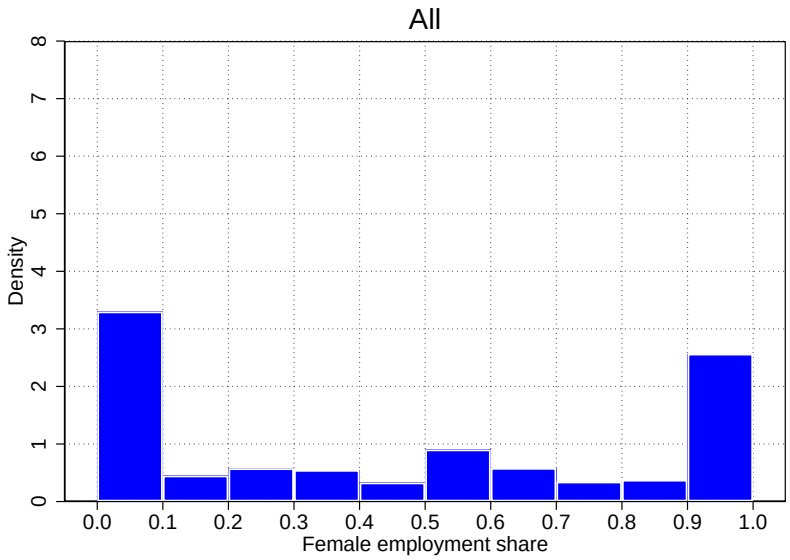
Finally, Figure 3 shows that there is an essentially flat relationship between the gender pay gap in firm fixed effects as a function of the female employment share, a pattern which we confirm to be robust across subsectors and regions of the country.

3.4 Firm heterogeneity in pay

We now formalize our investigation of the role of the firm in explaining the female earnings gap. Following the pioneering work by Abowd et al. (1999b) and a recent paper by Card et al. (2016), we estimate gender-specific firm pay components while at the same time controlling for observed time-invariant worker heterogeneity.[3] This allows for the possibility a given firm has not one but two pay policies, namely one for each gender. Formally, we estimate the following two-way fixed

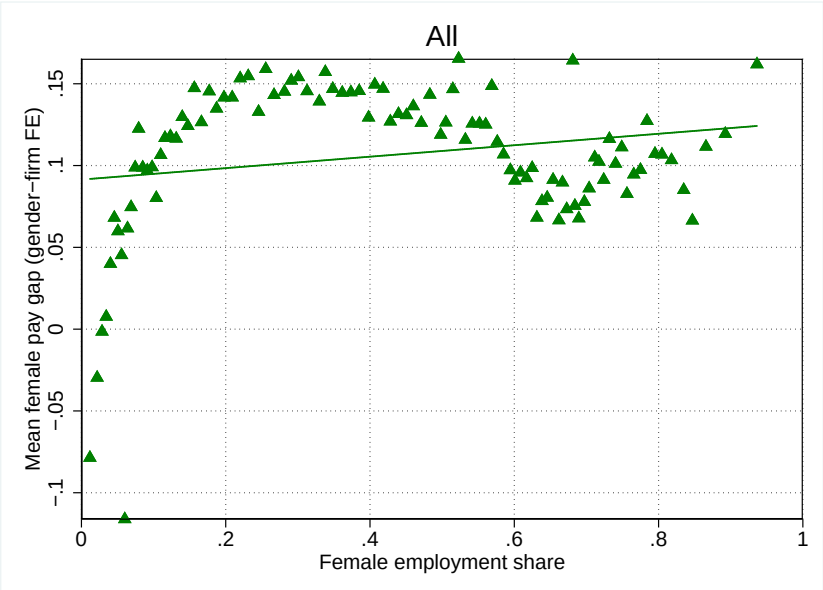
[3]In a separate specification, we also included individual fixed effects in the regression, although this requires an additional normalization of gender-specific firm effects.

Figure 2. Histogram of female employment shares



Notes: Brazilian formal sector census data (RAIS) for 2012.

Figure 3. Female pay gap versus female employment share



Notes: Brazilian formal sector census data (RAIS) for 2012.

effects framework explaining earnings of individual i working at firm j in year t, denoted y_{ijt}:

$$y_{ijt} = X_{it}\beta + \alpha_j^{base} - \mathbf{1}\left[female_i\right]\alpha_j^{gap} + \varepsilon_{ijt} \tag{3}$$

where X_{it} contains observable worker characteristics such as age, education, occupation, and race, plus potentially a person fixed effect controlling for unobservable but fixed worker heterogeneity, α_j^{base} is the male firm fixed effect and α_j^{gap} is the intra-firm pay gap, and ε_{ijt} is a residual term satisfying the usual strict exogeneity condition: $\mathbb{E}\left[\varepsilon_{ijt}|i,j,t\right] = 0$. We estimate equation (3) via ordinary leasts squares (OLS) and focus on estimates of the gender-specific firm fixed effects, α_j^{base} and α_j^{gap}. Based on the estimation results, we can decompose the overall gender gap as follows:

$$\alpha^{gap} = \underbrace{\mathbb{E}\left[\alpha_j^{gap}\middle|\, female\right]}_{\text{within-firm pay gap}} + \underbrace{\left(\mathbb{E}\left[\alpha_j^{base}\middle|\, male\right] - \mathbb{E}\left[\alpha_j^{base}\middle|\, female\right]\right)}_{\text{between-firm pay gap}}$$

where total gender pay differences on the left hand side, α^{gap}, is decomposed into a within-firm pay gap and a between-firm pay gap on the right hand side. The within-firm pay gap is the mean difference of estimated gender gaps within firms, where the average is weighted by the number of women in each firm. Therefore, this component is larger the bigger the pay differences within firms where women work. The between-firm pay gap is the difference in male firm fixed effects due to differences in the allocation of men relative to women across firms. Therefore, this component is larger the more disproportionately men work at high-paying firms compared to women.

Our decomposition results shed light on the drivers of the female earnings gap. We estimate a total female earnings gap of 0.252, out of which we attribute 0.137 (or 54%) to the within-firm component and the remaining 0.115 (or 46%) to the between-firm component. In other words, roughly half of the gender gap in Brazil is due to the fact that women systematically work at lower-paying firms relative to men, with the other half explained by women earning systematically less within the same firms where men work.[4]

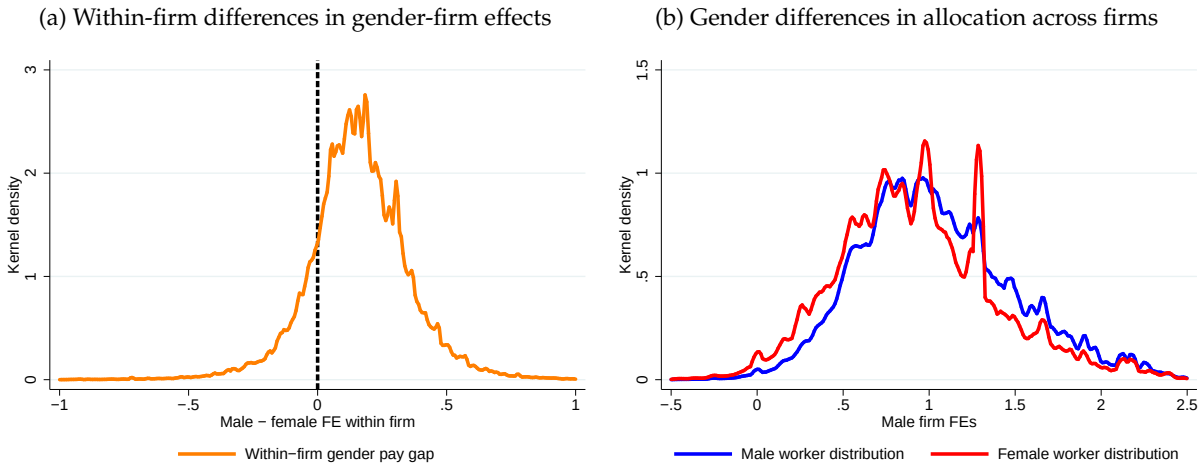
While on average women are paid less both within and across firms, we find significant heterogeneity in the magnitude of these gaps. Figure 4 shows the distributions of the within-firm pay gaps in panel (a). We find that approximately 80 percent of firms have higher estimated firm fixed

[4]We obtain similar results in an alternative decomposition where we take expectiations in equation (3) conditional on the distribution of men.

effects in pay for men than for women. The upper ten percent of the within-firm pay gaps distributions pays at least 50 percent more to men than to observationally equivalent women. At the same time, other firms pay more to women than to men, although this left tail of the distribution is less pronounced.

Panel (b) of Figure 4 shows the distribution of between-firm pay gaps by plotting separately the male firm fixed effect distribution for men and that for women. Evidently, women's distribution is tilted to the left relative to that of men. The largest gender gap exists in the two middle quartiles of the male firm effects distribution, with women disproportionately represented at lower pay firms relative to men.

Figure 4. Distributions of estimated gender-firm effects



Notes: Brazilian formal sector census data (RAIS) for 2003.

3.5 What explains the gender gap between firms?

3.5.1 The role of firm characteristics

In order to investigate the sources of between-firm differences in pay for men relative to women, we now turn to a second stage regression, building on the above first stage regression results. To this end, we regress the estimated gender-specific firm fixed effects in pay from equation (3), $\hat{\alpha}_j^{gender}$, on a host of firm characteristics:

$$\hat{\alpha}_j^{gender} = Z_{jt}\gamma + +\eta_{jt}$$

where Z_{jt} is a vector of characteristics for firm j in year t and η_{jt} is a residual term.

Table 5 shows the results of this second stage regression. We find that larger and older firms pay more in general, but relatively less so for women. For example, comparing columns (1) and (5) in the table, we see that firms that are ten percent larger in terms of their number of employees pay 0.71 percent more to men working at that firm, but only 0.38 percent more to women. The lower firm size-pay premium can therefore explain some of the observed gender pay gap to the extent that either women work at smaller firms on average, or women working at a firm of the same size as men get paid less than their male counterparts. Similarly, comparing columns (3) and (7) shows that older firms pay more to all workers, but disproportionately so for men. A firm that is ten percent older on average pays 0.59 percent more to its male employees, but only 0.30 percent to its female workers. Interestingly, the gender difference in the firm size pay premium is robust to including industry and state fixed effects, while the firm age-pay premium appears to be explained mostly by the differential allocation of women across those dimensions.

Table 5. Second stage regression: determinants of between-firm pay differences

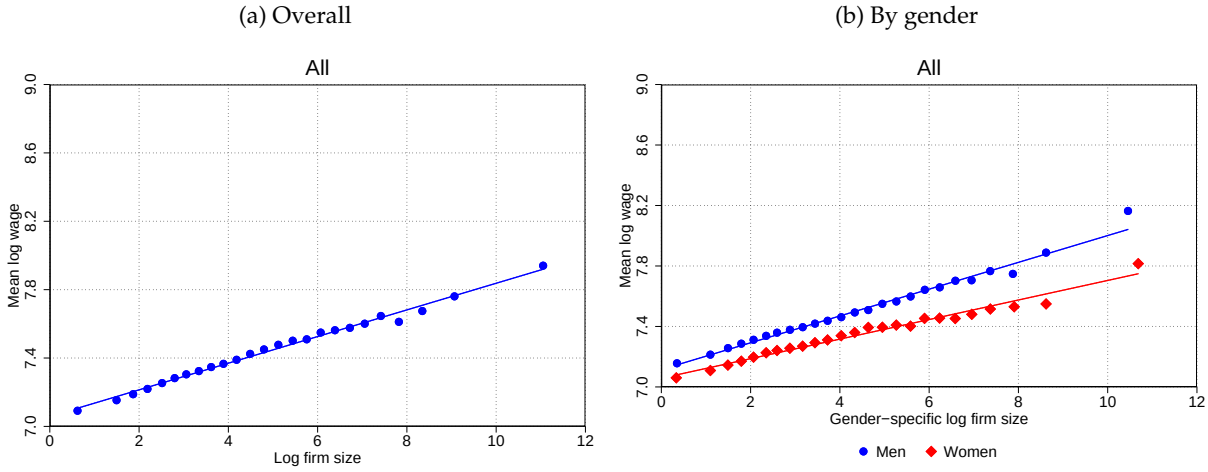
	Male				Female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log firm size	0.071		0.055	0.050	0.038		0.030	0.029
Log gender-firm size		0.076				0.029		
Log firm age			0.059	0.058			0.030	0.057
Industry, state FEs	no	no	no	yes	no	no	no	yes
# Obs. (mm)	17.6	17.6	17.5	17.5	12.0	12.0	11.9	11.9
R^2	0.12	0.13	0.13	0.42	0.06	0.03	0.06	0.38

Notes: Brazilian formal sector census data (RAIS) for 2003. Dependent variable is estimated firm-gender fixed effect from previous regression. Weighted by number of employees. All coefficients significant at 1% level.

A firm characteristic particularly important in predicting higher employer pay that emerged from our previous analysis is firm size. The idea that large firms pay more is consistent with job ladder models à la Burdett and Mortensen (1998), where a firm with relatively higher pay attracts more workers through between-firm competition for workers who engage in on-the-job search. Panel (a) of Figure 5 shows the firm size-pay premium, with an elasticity of earnings with respect to firm size approximately constant around 0.08. Splitting the firm size-pay premium by gender, we see that women get paid less at every firm size bin and that their pay increases at a lower rate

with the size of their employer.

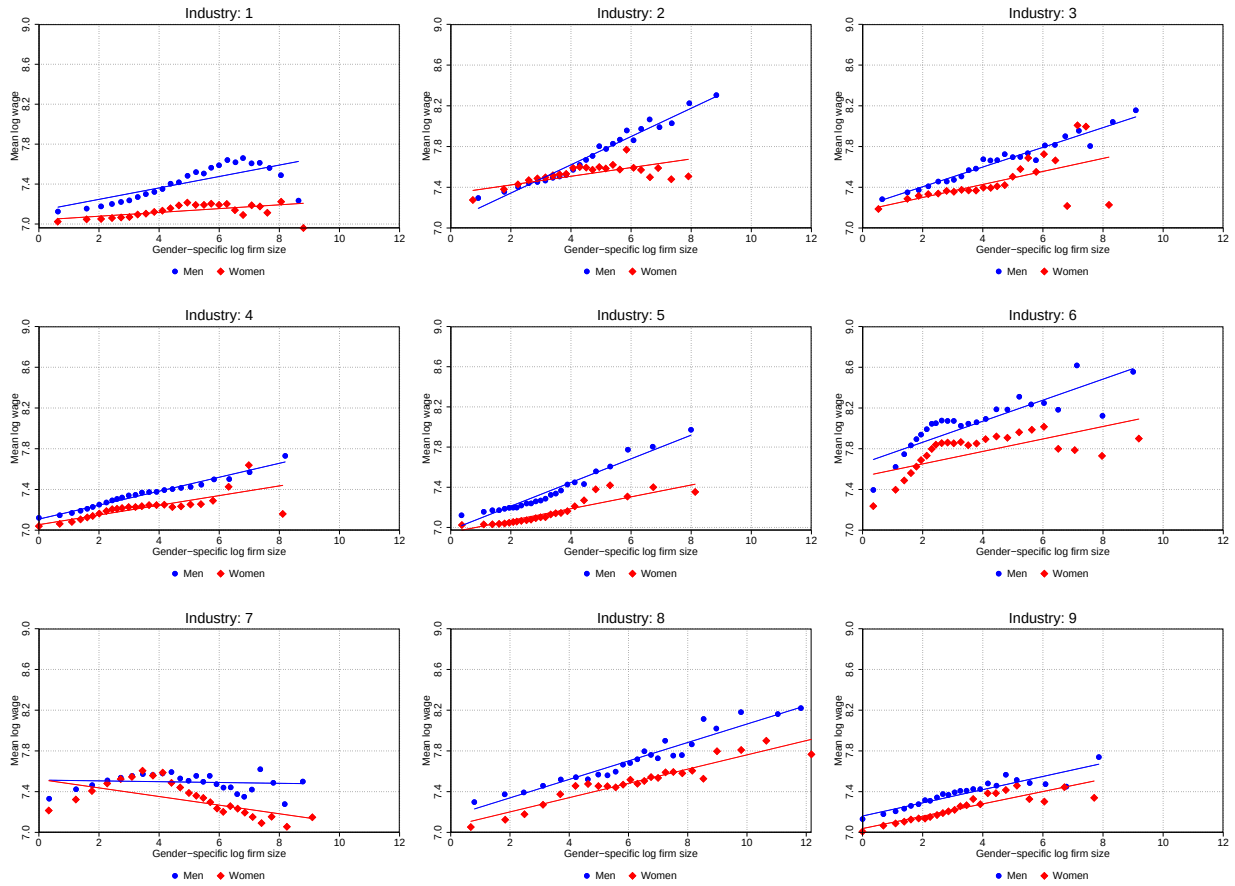
Figure 5. Binscatter plot of mean log wage against log firm size, overall and by gender



Notes: Brazilian formal sector census data (RAIS) for 2012.

Figure 6 shows that this pattern is robust across subgroups defined by 1-digit industries according to Brazil's CNAE 2.0 sector classification system. Appendix A.1 shows that a similar pattern emerges when looking across subgroups by regions, education groups, and occupation groups.

Figure 6. Binscatter plot of gender-specific mean log wage against log firm size, by sector



Notes: Brazilian formal sector census data (RAIS) for 2012.

3.5.2 Gender-specific lifecycle income profiles

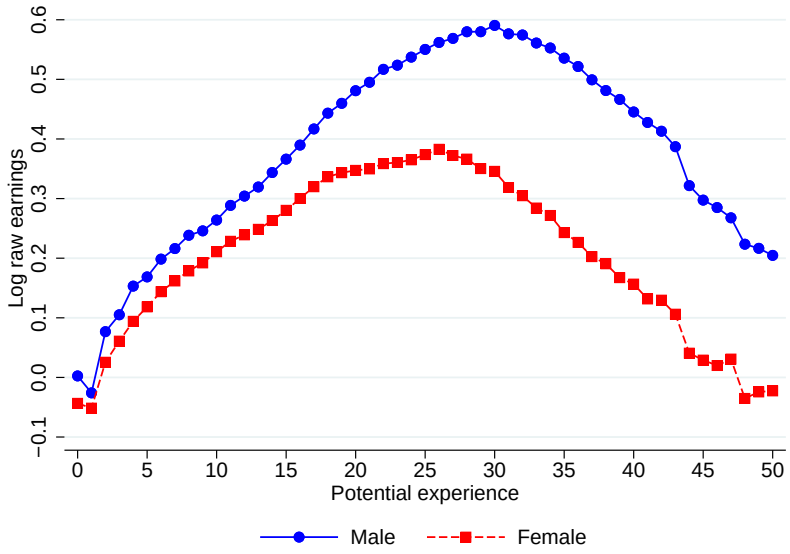
How do women end up working at lower-pay firms? A natural starting point is to look at the life-cycle pattern of earnings for women compared to men. As a first pass, we compute average earning by gender over the first 51 years of potential labor market experience, where we define the latter as age minus years of education minus six (Mincer, 1974). Formally, we write individual i's earnings in year t as follows:

$$y_{it} = \sum_{e=0}^{50} \mathbf{1}\left[exp_{it} = e\right] \left(1[male_i]\alpha_e^{male} + 1[female_i]\alpha_e^{female}\right) + \varepsilon_{it}$$

Figure 7 plots estimated mean earnings across potential experience levels by gender. A few points are worth noting about the graph. First, women and men start out earning very similar amounts in their first job out of school. Over the subsequent 30 years, however, a gap between

mena nd women gradually opens up, partly due to women's earnings growing less fast for the first 25 years and partly due to men's life time earnings profile peaking around five years later than that for women. While the gender earnings gap is close to zero around the time of entering the labor market, at its height the gap reaches a staggering 30 log points, or around 35 percent. Finally, in the later years of the career, the female earnings gap begins to shrink again somewhat.

Figure 7. Life cycle trajectory of male and female earnings



Notes: Brazilian formal sector census data (RAIS) for 2003.

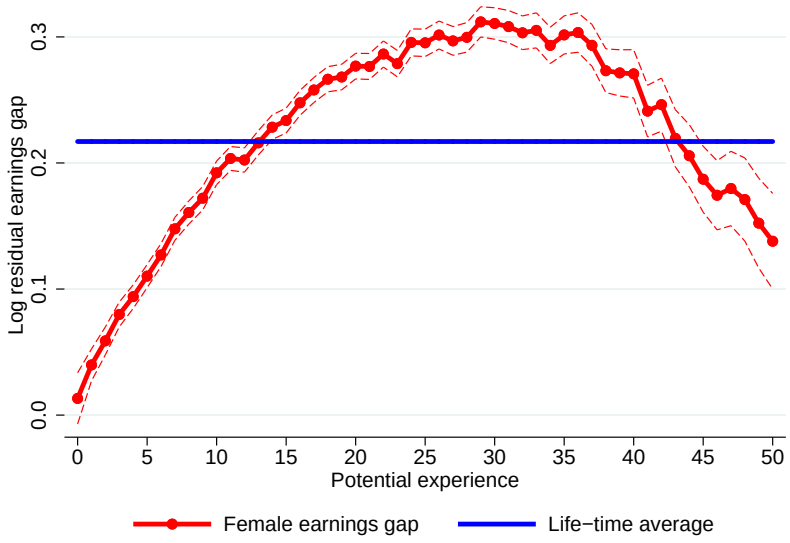
As the previous graph showed merely raw wages, one may expect that some oft he life-cycle gap in earnings may be due to other observable factors such as differences in education, or the industry and occupation of employment. But as our earlier results may already suggest, these observable differences do little to close the gender gap, which we showed was 21.6 log points on average. To investigate the life-cycle pattern of the residual earnings gap, we add observable worker controls to the above specification:

$$y_{it} = \sum_{e=0}^{50} \mathbf{1}\left[exp_{it} = e\right]\left(\alpha_e^{base} - \mathbf{1}\left[female_i\right]\alpha_e^{gap}\right) + X_{it}\beta + \varepsilon_{it}$$

where X_{it} is a vector containing the same set of worker controls mentioned above. In line witht he raw data results, Figure 8 shows that the residual female earnings gap is also inverse-U-shaped

over the life cycle.[5]

Figure 8. Life cycle pattern of conditional female pay gap



Notes: Brazilian formal sector census data (RAIS) for 2003.

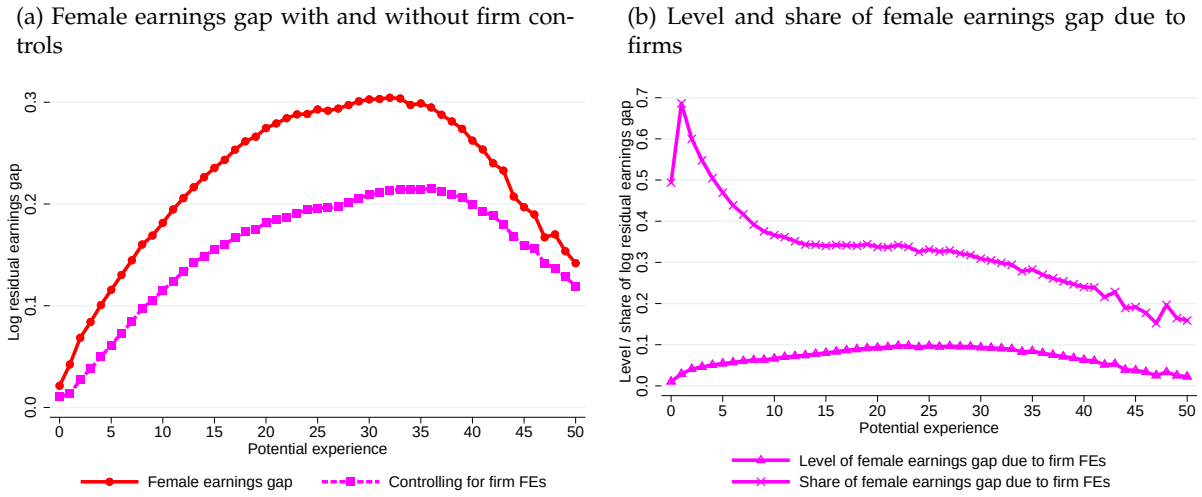
How much of the life cycle earnings gap is explained by firm heterogeneity? To answer this question, we introduce gender firm fixed effects as an additional control into the above specification:

$$y_{it} = \sum_{e=0}^{50} \mathbf{1}\left[exp_{it} = e\right]\left(\alpha_e^{base} - \mathbf{1}\left[female_i\right]\alpha_e^{gap}\right) + X_{it}\beta + \alpha_j^{base} - \mathbf{1}\left[female_i\right]\alpha_j^{gap} + \varepsilon_{it}$$

Panel (a) of Figure 10a shows the estimated residual earnings gap with and without such gender-specific firm controls. Introducing the firm fixed effects into the specification reduces the residual earnings gap by up to 30 percent, meaning that firm heterogeneity explains a substantial share of the overall earnings gap. But the contribution of firm heterogeneity towards gender pay differences itself follows an inverse-U shape over the life cycle. Contribution of between-firm pay differences to female earnings gap follows inverse U-shape over life-cycle, as shown in panel (b) of the figure.

[5]See also Erosa et al. (2016).

Figure 9. The role of firm heterogeneity in explaining life cycle pattern of the gender pay gap



Notes: Brazilian formal sector census data (RAIS) for 2003.

3.5.3 Gender-specific lifecycle job ladders

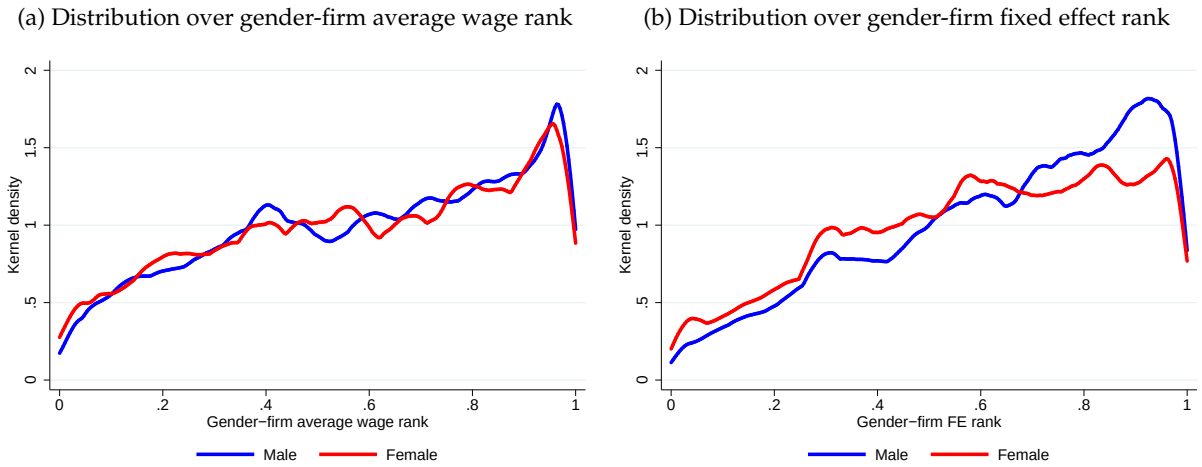
The pronounced life cycle pattern in the gender earnings gap and in the contribution of firm heterogeneity in particular points towards differences in labor markets as an important source of gender inequality in Brazil. A large literature building on Burdett and Mortensen (1998) has theorized that the workings of a "job ladder," by which workers continuously reallocate towards better employment opportunities over their life cyle, can explain a host of empirical labor market regularities. We focus here on a related but new aspect of this job ladder hypothesis, namely that "the" job ladder may look quite different for men relative to women.

One potential gender difference in the allocation of workers across firms in the job ladder could stem from the fact that men immediately when starting their career (or restarting it out of unemployment) are matched with better firms compared to their female peers. To test this sub-hypothesis, we compare starting rungs of the job ladder for labor market entrants from unemployment across genders. To this end, we first need to take stance on how to define the job ladder. We do so in two complementary ways. First, we rank firms in ascending order by the average wage they pay to their workforce. Second, we rank firms according to the estimated gender-specific firm fixed effects from equation (3) above.

Figure 10 shows the distributions of entry-level firm ranks from a kernel density estimate using an Epanechnikov kernel. There is little difference in the average wage rank at which men and

women start out of unemployment, shown in panel (a) of the figure. Panel (b) shows that women start at lower gender-firm fixed effect ranks compared to men. Quantitatively, we find this channel to explain a small share of the overall gender earnings gap, both on average and over the life cycle.

Figure 10. Distribution of first job from unemployment across firms



Notes: Brazilian formal sector census data (RAIS) for 2003.

The fact that it is not primarily women's starting firm rank relative to men leads us to conclude that something about the speed of climbing the job ladder must differ across genders. To investigate this second sub-hypothesis, wse track workers over time across firms and compute labor market transitions, both into and out of formal sector employment as well as between jobs in the formal sector. Table 6 summarizes the results of our investigation. There is little indication that women in Brazil are subject to longer unemployment spells than men. While it is certainly true that less women than men participate in the formal sector labor market, we find that conditional on a women having entered the formal labor force she is as likely as a man in the same position to return to formal sector employment following a nonemployment spell. A second dimension of labor mobiltiy is that women face a "paper floor" in the labor market, meaning that they are more likely to drop out of employment than men (Guvenen et al., 2014). But again, we find that female workers are no more likely to lose their job relative to men. A third dimension of labor fluidity is the degree of upward mobility, measured by job-to-job transition rates. In this context, we find that women exhibit substantially different patterns than men, which are consistent with the existence of a "glass ceiling": female workers are 28% less likely to find a job while employed relative to their male counterparts.

Table 6. Labor market transition rates by gender

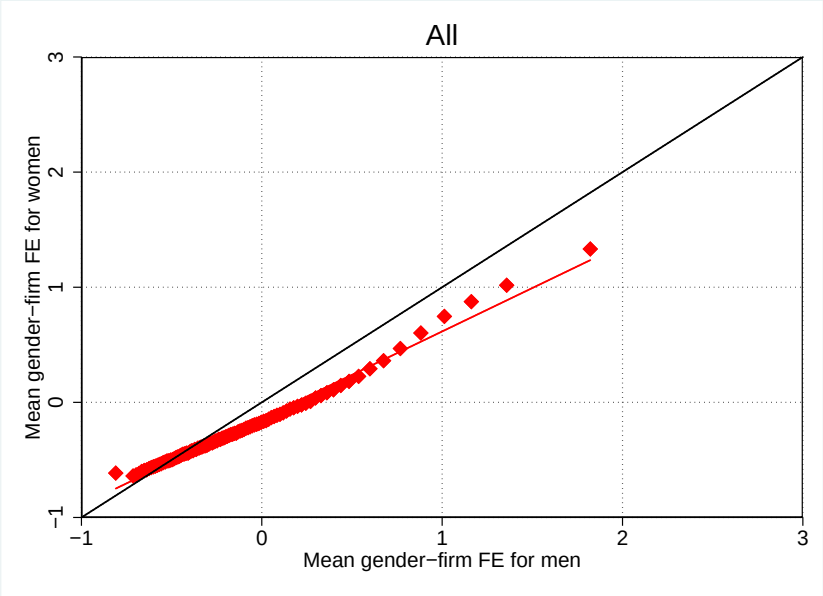
	Male	Female
Prob. of unemployment-to-employment (UE) transition	0.165	0.157
Prob. of employment-to-unemployment (EU) transition	0.182	0.172
Prob. of employment-to-employment (EE) transition	0.016	0.012

Notes: 2001-2005 Brazilian formal sector census data (RAIS). Statistics are computed at monthly frequency for workers of age 25-45 and weighted by number of workers.

3.6 What explains the gender gap within firms?

We now turn to examining the determinants of within-firm pay differences. To get a sense of the joint distribution of gender-firm effects in the population, based on the results of estimating equation equation (3), we first compute the average female firm fixed effect as a function of the male fixed effect in the same firm. Figure 11 shows the results of this exercise. Higher-paying firms for men on average are higher-paying firms for women. But the gradient of firm fixed effects for women in those for men is less than one, implying that a gender-firm gap opens up toward higher male firm pay ranks. This gender-firm gap is approximately 0% at the bottom of the firm pay distribution but as high as 50 percent at the top, pointing to sizable within-firm pay differences.

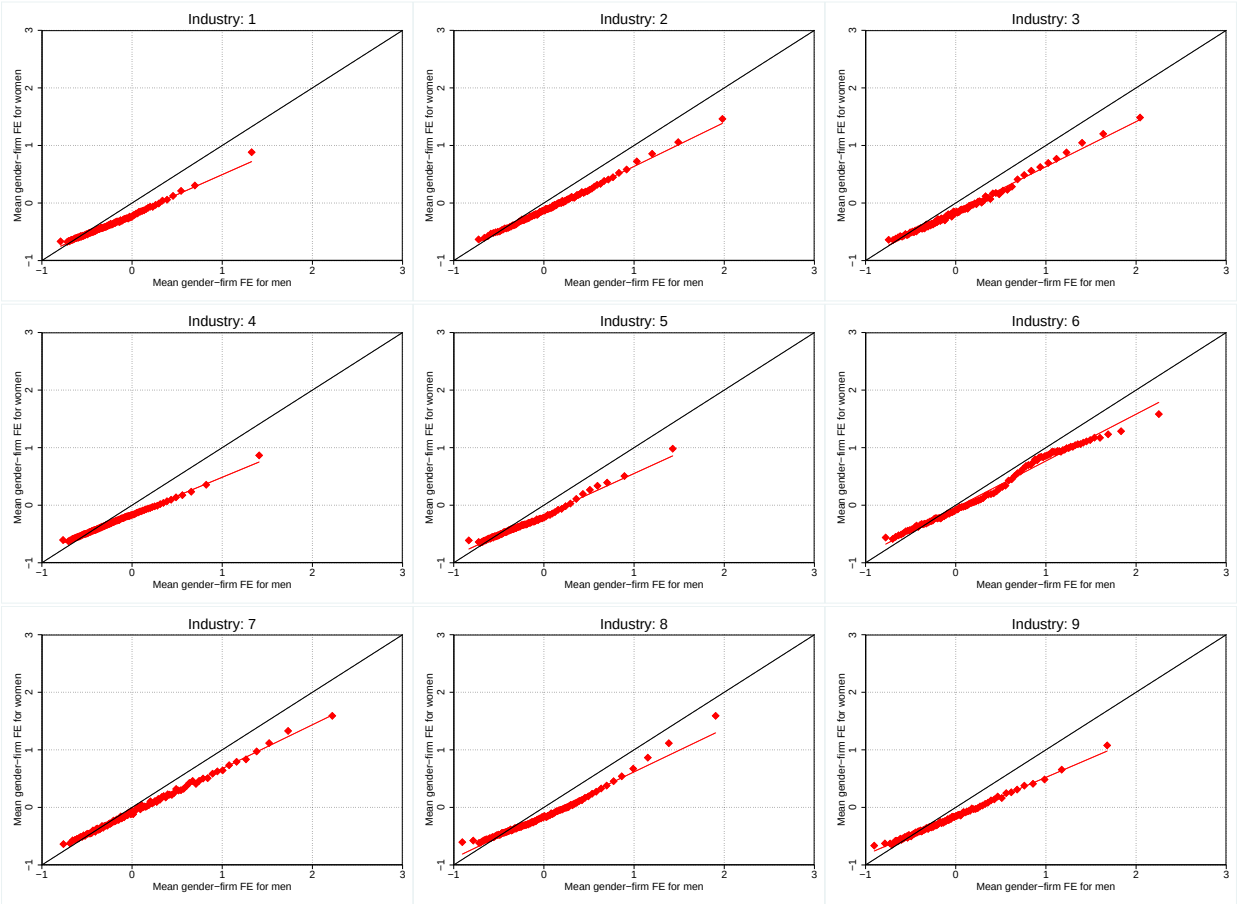
Figure 11. Average female firm fixed effect across male firm fixed effects distribution



Notes: Brazilian formal sector census data (RAIS) for 2012.

Figure 12 shows the same relationship between gender-firm fixed effects for women against that for men for nine 1-digit industries according to Brazil's CNAE 2.0 sector classification system. The stable relationship that emerges from these figures is that women get paid less than men at most employers, that higher-paying employers for men are also higher-paying employers for women, and that the difference between female employer fixed effects and that for men is increasing in the level of men's employer fixed effect. Appendix A.2 shows that this pattern is robust across regions, education groups, and occupation groups, suggesting that it is not drive by sorting on those observables.

Figure 12. Average female firm fixed effect across male firm fixed effects distribution, by sector



Notes: Brazilian formal sector census data (RAIS) for 2012.

What determines the these within-firm pay differences? To answer this question, we proceed analogously to the between-firm section and regress in a second stage the estimated firm-pay gap to observable firm characteristics:

$$\hat{\alpha}_j^{gap} \equiv \hat{\alpha}_j^{male} - \hat{\alpha}_j^{female} = Z_{jt}\delta + \xi_{jt}$$

where Z_{jt} is again a vector containing the observable characteristics of firm j in year t, and ξ_{jt} is an error term. Table 7 shows the results from this regression analysis. We find that factors associated with smaller pay gaps within firms include the share of female employees and whether or not a female worker is in the highest-paid position at the firm, with elasticities of -0.055 for each of the two when introduced separately into the regression. In a horserace between the two factors, it turns out that having a female worker in a leadership position is more than three times

as important as having a more female general workforce. This conclusion is robust to controlling for log earnings of the highest-paid female employee, suggesting that our previous estiamtes are not purely mechnical.

Table 7. Regression of female pay gap on its determinants

	(1)	(2)	(3)	(4)	(5)
Share of female employees	-0.055		-0.014	-0.031	-0.006
Ind: female highest-paid worker		-0.055	-0.050	-0.055	-0.046
Log highest female earnings				0.009	0.005
Industry and state FEs	no	no	no	no	yes
# Observations (millions)	21.1	21.1	21.1	18.8	18.8
R^2	0.01	0.02	0.02	0.03	0.09

Notes: 5% sample of 2001–2005 Brazilian formal sector census data (RAIS). Dependent variable is within-firm difference in gender-firm FEs. Weighted by number of employees. All coefficients significant at 1% level.

Next, we turn to the role of occupations, tasks, and skill contents of workers of different genders within firms. We first convert 5-digit Brazilian CBO occupation codes to US Census occupation codes, which we then link to skill task contents from the Dictionary of Occupational Titles (Autor et al. (2003), Acemoglu and Autor (2011)). A first look at the distribution of skill and task contents is presented in Table 8. We find that men tend to be employed in more "brawny" occupations that require routine manual and non-routine manual physical skills, while females are employed in more "brainy" and interpersonal occupations that require non-routine manual personal and non-routine cognitive analytical skills.

Table 8. Skill and task requirements of occupations held by each gender

	Male	Female	Difference (male - female)
Routine Manual	0.368	-0.023	0.391
Routine Cognitive	-0.254	-0.056	-0.198
Non-routine Manual Physical	0.446	-0.136	0.582
Non-routine Manual Personal	-0.324	0.269	-0.593
Non-routine Cogn. Personal	-0.230	0.102	-0.332
Non-routine Cogn. Analytical	-0.130	0.210	-0.340
# Observations (millions)	20.1	13.6	

Notes: 2003 Brazilian formal sector census data (RAIS), combined with US Census Occupation Codes and Dictionary of Occupational Titles. Variables are standard normal z-scores with population mean 0 and variance 1.

Given that men and women are distributed unequally across job and skill positions within firms, how does their remuneration depend on their occupation? In other words, conditional on task skill content, do women earn the same returns to their skills as men do? To this end, we introduce task and skill contents directly as a regressor into our main regression specification:

$$y_{ijt} = X_{it}\beta + \alpha_j^{gender} + taskskill_{ijt} \times \gamma^{gender} + \varepsilon_{ijt}$$

where $taskskill_{ijt}$ is the occupation-specific measure derived from the Dictionary of Occupational Titles, and γ^{gender} is a coefficient that we allow to differ across genders. We alreadt described differences in the allocation of women compared to men across task and skill groups, $taskskill_{ijt}$, so we turn next to gender differences in the returns to these categories.

Table 9 presents our finding of almost uniformly lower returns for women compared to men for the same task and skill contents associated with a given occupation. Comparing columns (1) and (4), for example, shows that women earn less than men within most task-skill categories. This pattern is robust, although the magnitude changes somewhat, when indroducing other controls into the regression, including interacted education and experience dummies as well as state and industry fixed effects, see columns (2) and (6). Most importantly, when we introduce firm controls into the specification, substantial share of the gap in gender-specific returns to skill and task contents vanishes, implying that firm heterogeneity mediates the gender differences in returns to occupational contents.

Table 9. Regression of pay on gender-specific skills and task requirements across occupations

	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Routine Manual	-0.078	-0.059	-0.052	-0.203	-0.058	-0.058
Routine Cognitive	0.141	0.077	0.063	0.095	0.062	0.045
Non-routine Manual Physical	0.014	0.056	0.077	-0.080	-0.031	-0.032
Non-routine Manual Personal	-0.093	-0.047	-0.007	-0.176	-0.055	-0.029
Non-routine Cogn. Personal	-0.083	-0.022	0.032	0.023	-0.050	0.038
Non-routine Cogn. Analytical	0.323	0.143	0.105	0.211	0.127	0.073
Edu×exp FEs	no	yes	yes	no	yes	yes
State, industry FEs	no	yes	no	no	yes	no
Firm FEs	no	no	yes	no	no	yes
# Observations (millions)	21.0	21.0	21.0	13.6	13.6	13.6

Notes: 2003 Brazilian formal sector census data (RAIS), combined with US Census Occupation Codes and Dictionary of Occupational Titles. Variables are standard normal z-scores with population mean 0 and variance 1.

Finally. we turn to investigating on-the-job earnings growth differences across women and men. To the extent that men get promoted more frequently than women, this may be an important driver of gender earnings differences within firms. To quantify the contribution of this channel, we estimate the following specification, which introduces a gender-specific tenure profile into our main specification:

$$y_{ijt} = X_{it}\beta + tenure_{ijt} \times \gamma^{gender} + \alpha_j^{gender} + \alpha_t + \varepsilon_{ijt}$$

where $tenure_{ijt}$ is tenure of individual i at firm j in year t, and γ^{gender} is a gender-specific coefficient. Table 10 shows the estimation results from this exercise. We find that females' earnings growth with tenure is 18 percent lower than that of males, indicating a there exists a gender gap across tenure levels, not just labor market experience in general. For example, comparing columns (1) and (4) we see that men have a 0.34 log points return to an additional month of tenure at their firm, implying a roughly 4.2 percentage points annualized earnings growth on-the-job. This stands in comparison to a 0.28 log points return to tenure for women, implying approximately 3.4 percentage points annualized earnings growth on-the-job. These results are robust to controlling for a rich set of covariates, including education and experience fixed effects, introduced separately or flexibly interacted. When introducing occupation fixed effects, however, the gender difference shrinks by over 70 percent, indicating that most of the gender gap in the returns to tenure is due to men

moving up through occupational ranks more quickly than women do.

Table 10. Regression of pay on gender-specific tenure variable

	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Tenure (months)	0.0034	0.0027	0.0028	0.0028	0.0020	0.0026
Year, gender-firm FEs	yes	yes	yes	yes	yes	yes
Education FEs	yes	no	yes	yes	no	yes
Edu×exp FEs	no	yes	no	no	yes	no
Occupation FEs	no	no	yes	no	yes	yes
# Obs. (millions)	25.4	25.4	25.4	25.4	25.4	25.4
R^2	0.66	0.69	0.72	0.66	0.69	0.72

Notes: 5% sample of 2001-2005 Brazilian formal sector census data (RAIS). Dependent variable is within-firm difference in gender-firm FEs. Weighted by number of employees. All coefficients significant at 1% level.

4 Model

Time is continuous. The economy is populated by a measure male and female workers, indexed by their productivity a and their gender $g \in \{M, F\}$. Their measure is indexed by $\mu_{a,g}$, where $\mu_{a,m} + \mu_{a,f} = 1$. We refer to the duple $\{a, g\}$ as the *type* of the worker. There exists a large outside measure of firms $M > 1$, indexed by j.

4.1 Workers

Workers can be either employed or unemployed. Each worker is assigned a fixed level of productivity $a \in [\underline{a}, \bar{a}]$, drawn from a group-specific distribution H^g. While unemployed, workers get flow utility $b_{a,g}$. Search is segmented in the sense that workers with different ability a and different gender g search in different submarkets. Also, search is random in the sense that workers cannot direct their search to specific firms. Workers receive job offers at random both during employment and unemployment, with arrival rates $\lambda^e_{a,g}$ and $\lambda^u_{a,g}$ respectively. Additionally, to allow workers to move down the ladder when employed, workers receive "offers they can't refuse", commonly referred to as godfather shocks, at rate $\lambda^G_{a,g}$, as in Moscarini and Postel-Vinay (2018), both during

employment and unemployment. Therefore, we define $\lambda^e_{a,g} = s^e_{a,g}\lambda^u_{a,g}$ and $\lambda^G_{a,g} = s^G_{a,g}\lambda^u_{a,g}$, where $s^e_{a,g}$ and $s^G_{a,g}$ are the search effort exercised by employed workers due to on-the-job search and by all workers due to godfather shocks, respectively. Finally, workers are infinitely lived and care for the value of the stream of expected consumption, discounted at rate ρ.

A job offer is an opportunity to work at firm j for the wage w and a level of firm-specific amenity π_j, drawn from the distribution $F(w+\pi)$, an endogenous object that will be determined in equilibrium. A job is terminated either endogenously, because the worker moves to a better contract, or exogenously with type- and gender-specific rates $\delta_{a,g}$.

Thus, the problem of the employed worker can be written as follows:

$$\begin{aligned}\rho S_{a,g}(w+\pi) =& w+\pi+\lambda^e_{a,g}\int_{w+\pi}^{\bar{w}+\bar{\pi}} S_{a,g}(w'+\pi') - S_{a,g}(w+\pi)\, dF_{a,g}(w'+\pi') \\ &+\lambda^G_{a,g}\int_{\underline{a}+\underline{\pi}}^{\bar{w}+\bar{\pi}} S_{a,g}(w'+\pi') - S_{a,g}(w+\pi)\, dF_{a,g}(w'+\pi') + \delta_{a,g}[W_{a,g} - S_{a,g}(w+\pi)]\end{aligned} \tag{4}$$

The problem of the unemployed worker reads

$$\rho W_{a,g} = b_{a,g} + (\lambda^u_{a,g} + \lambda a, g^G)\int_{\underline{a}+\underline{\pi}}^{\bar{w}+\bar{\pi}} \max\{S_{a,g}(w+p) - W_{a,g}, 0\}\, dF_{a,g}(w+\pi) \tag{5}$$

Strict monotonicity of the value function $S_{a,g}$ in $w+\pi$ implies that the optimal strategy of an unemployed worker will be characterized by a reservation "combination" of wage and amenity level $\phi_{a,g}$. Thus, an unemployed worker will accept any offer for which $w+\pi > \phi_{a,g}$ and reject all others. The reservation combination can be written as the sum of the flow value of unemployment plus the forgone option value of being unemployed, as in Burdett and Mortensen (1998):

$$\phi_{a,g} = b_{a,g} + (\lambda^u_{a,g} - \lambda^e_{a,g})\int_{\phi_{a,g}}^{\bar{w}+\bar{\pi}} \frac{1-F_{a,g}(w+\pi)}{\rho+\delta_{a,g}+\lambda^G_{a,g}+\lambda^e_{a,g}(1-F_{a,g}(w+\pi))}\, d(w+\pi)\,. \tag{6}$$

Since in equilibrium no firm will post a contract worth less than $\phi_{a,g}$ in that submarket, the unemployment rate for each type and gender is

$$u_{a,g} = \frac{\delta_{a,g}}{\delta_{a,g}+\lambda^u_{a,g}+\lambda^G_{a,g}}\,. \tag{7}$$

In steady state, the offer distribution in terms of $w + \pi$ is

$$G_{a,g}(w+\pi) = \frac{F_{a,g}(w+\pi)}{1+\kappa_{a,g}(1-F_{a,g}(w+\pi))} \tag{8}$$

where $\kappa_{a,g} = \lambda^e_{a,g}/(\delta_{a,g} + \lambda^G_{a,g})$ governs the relative speed of climbing the job ladder.

4.2 Firms

There is a large outside measure of firms $M > 1$, indexed by j. Firms are ex-ante heterogeneous in the sense that each firm j is characterized by a fixed productivity p_j, a gender-specific firm-specific amenity level $\pi_{j,g}$ and an employer taste for women z_j, drawn at birth from the joint distribution $D(p, \pi, z)$. Because of the interaction between heterogeneity and search frictions, firms will also be ex-post heterogeneous in terms of posted wages and sizes for each gender.

Workers care for the amenity level they receive while staying at their current firm and for the monetary wage they earn at that firm. Firms aim to maximize their steady state profits.

Thus, firms are fully characterized by the quadruple $\{p, \pi_m, \pi_f, z\}$. They produce

$$y(p, \{l_{a,m}\}_{a\in A_m}, \{l_{a,f}\}_{a\in A_f}) = p\int_{a\in A_m} a\, l_{a,m}\, dA_m + p\int_{a\in A^f} a\, l_{a,f}\, dA_f\,. \tag{9}$$

Output is additively separable across worker types, and markets are segmented by types. We now discuss these two assumptions in some detail. First, the assumption that output is additively separable in genders allows the model to be flexible in terms of the gender structure of every single firm. Assuming complementarities between genders would imply that male and female firm sizes are more correlated than in the data. Also, it would make our model considerably harder to solve.

Second, we assume that vacancy posting is separate across genders because we find that the main modeling alternatives have counterfactual implications. We consider two natural alternative assumptions: that the vacancy cost is a convex function of the *sum* of male and female vacancies; and that vacancy posting is joint so that firms cannot direct vacancies to a specific genders. The first of these alternative assumptions has the counterfactual implication that, since the marginal benefit of a vacancy is different between genders, every firm would hire only men, only women, or be indifferent between any male-female composition.[6] In the data, instead, we observe substantial variation in female shares, which are most often than not different from zero and one. The second

[6] We discuss in detail the proof of this argument in Appendix B.1.

alternative assumption, that vacancies cannot target genders, implies too little variation in female shares across firms with respect to what we observe in the data.[7]

Under the assumption of separate vacancy posting, the firm's problem reads:

$$\begin{aligned}\max_{w_{a,m},w_{a,f},v_{a,m},v_{a,f}} \quad & \Big\{(p\,a - w_{a,m})l_{a,m}(w_{a,m}+\pi_m, v_{a,m}) \\ & + (p\,a - z - w_{a,f})l_{a,f}(w_{a,f}+\pi_f, v_{a,f}) - c_{a,m}(v_{a,m}) - c_{a,f}(v_{a,f})\Big\}\end{aligned} \tag{10}$$

where $c_{a,m}(v_{a,m})$ and $c_{a,f}(v_{a,f})$ are increasing, convex functions that satisfy $c(0) = 0$. Thus, a firm understands that a higher amenity level means that, at a fixed wage level, workers will be more likely to accept its offer compared to the offer of another firm with the same wage and a lower amenity level. Define the effective wage $\tilde{w}_{a,g} = w_{a,g} + \pi_{a,g}$. Then we can write the firm's problem as

$$\begin{aligned}\max_{\tilde{w}_{a,m},\tilde{w}_{a,f},v_{a,m},v_{a,f}} \quad & \Big\{(p\,a + \pi_m - \tilde{w}_{a,m})l_{a,m}(\tilde{w}_{a,m}, v_{a,m}) \\ & + (p\,a + \pi_f - z - \tilde{w}_{a,f})l_{a,f}(\tilde{w}_{a,f}, v_{a,f}) - c_{a,m}(v_{a,m}) - c_{a,f}(v_{a,f})\Big\}\end{aligned} \tag{11}$$

Therefore, firms place themselves on a ladder according to $\tilde{w}$.

A firm makes profits if and only if

$$\phi_{a,g} < \tilde{w} < p\,a + \pi_g - z\mathbb{I}[g = \text{female}]\,. \tag{12}$$

Thus, only firms for which $p\,a + \pi_g - z\mathbb{I}[g = \text{female}] > \phi_{a,g}$ will operate. In general, our model allows firms to hire any combination of both genders, including hiring only one gender.

The matching function follows the standard Cobb-Douglas functional form:

$$m_{a,g} = A^g[\mu_{a,g}(u_{a,g} + \lambda^e_{a,g}(1 - u_{a,g}) + s^G_{a,g})]^\alpha V^{1-\alpha}\,. \tag{13}$$

Thus, the total number of vacancies posted in submarket $\{a, g\}$ is equal to

$$V_{a,g} = \int v_{a,g}(p, \pi, z)\, dD_{a,g}(p, \pi, z) \tag{14}$$

[7] We discuss in detail the joint vacancy posting model, its implications on the female share and a solution algorithm for that model in Appendix B.2.

4.3 Equilibrium

Define $\lambda^e_{a,g} = s_{a,g}\lambda^u_{a,g}$. The following equation represents the law of motion of firm sizes:

$$\begin{aligned}\dot{l}_{a,g}(\tilde{w},v) =& -\delta_{a,g} l_{a,g}(\tilde{w},v) - \lambda^e_{a,g}(1-F_{a,g}(\tilde{w}))l_{a,g}(\tilde{w},v) - \lambda^G_{a,g} l_{a,g} + \\ & v q_{a,g}\left[\frac{u_{a,g}+s^G_{a,g}}{u_{a,g}+(1-u_{a,g})s^e_{a,g}+s^G_{a,g}} + \frac{(1-u_{a,g})s^e_{a,g}}{u_{a,g}+(1-u_{a,g})s_{a,g}+s^G_{a,g})}G_{a,g}(\tilde{w})\right]\end{aligned} \tag{15}$$

Solving for the stationary solution:

$$l_{a,g}(\tilde{w}_{a,g},v_a) = \left(\frac{1}{\delta_{a,g}+\lambda^G_{a,g}+\lambda^e_{a,g}(1-F_{a,g}(\tilde{w}_{a,g}))}\right)^2 \frac{v_{a,g}}{V_{a,g}}\mu_{a,g}(u_{a,g}+s^G_{a,g})\lambda^u_{a,g}(\delta_{a,g}+\lambda^G_{a,g}+\lambda^e_{a,g}) \tag{16}$$

4.4 Equilibrium Characterization

To find the firm policy functions, define $T_{a,g} = \mu_{a,g}[(u_{a,g}+s^G_{a,g})\lambda^u_{a,g}(\delta_{a,g}+\lambda^G_{a,g}+\lambda^e_{a,g})]/V_{a,g}$. we rewrite the firm's problem as a function of the steady state mass of employed workers as follows:

$$\begin{aligned}\max_{\tilde{w}_{a,m},\tilde{w}_{a,f},v_{a,m},v_{a,f}} \quad & \left\{T_{a,m}v_{a,m}(a\,p+\pi_m-\tilde{w}_{a,m})\left(\frac{1}{\delta_{a,m}+\lambda^G_{a,m}+\lambda^e_{a,m}(1-F_{a,m}(\tilde{w}_{a,m}))}\right)^2\right. \\ & \left. + \left\{T_{a,f}v_{a,f}(a\,p+\pi_f-z-\tilde{w}_{a,f})\left(\frac{1}{\delta_{a,f}+\lambda^G_{a,f}+\lambda^e_{a,g}(1-F_{a,f}(\tilde{w}_{a,f}))}\right)^2 - c_{a,m}(v_{a,m}) - c_{a,f}(v_{a,f})\right\}\right.\end{aligned} \tag{17}$$

The associated FOC read:

$$\begin{aligned} c^{m\prime}(v_{a,m}) &= T_{a,m}(ap+\pi_m-\tilde{w}_{a,m})\left(\frac{1}{\delta_{a,m}+\lambda^G_{a,m}+\lambda^e_{a,m}(1-F_{a,m}(\tilde{w}_{a,m}))}\right)^2 \\ c^{f\prime}(v_{a,f}) &= T_{a,f}(a\,p+\pi_f-z-\tilde{w}_{a,f})\left(\frac{1}{\delta_{a,f}+\lambda^G_{a,f}+\lambda^e_{a,f}(1-F_{a,f}(\tilde{w}_{a,f}))}\right)^2 \\ 1 &= (a\,p+\pi_m-\tilde{w}_{a,m})\frac{2\lambda^e_{a,m}f_{a,m}(\tilde{w}_{a,m})}{\delta_{a,m}+\lambda^G_{a,m}+\lambda^e_{a,m}(1-F_{a,m}(\tilde{w}_{a,m}))} \\ 1 &= (a\,p+\pi_f-z-\tilde{w}_{a,f})\frac{2\lambda^e_{a,f}f_{a,f}(\tilde{w}_{a,f})}{\delta_{a,f}+\lambda^G_{a,f}+\lambda^e_{a,f}(1-F_{a,f}(\tilde{w}_{a,f}))}\end{aligned} \tag{18}$$

Our model justifies our focus on the AKM decomposition of log wages. Define $p_{a,g} = ap + \pi_{a,g} - z_{a,g}$ as the composite of productivity, amenities and employer taste for a gender, with the understanding that $z_{a,m} = 0$ for all ability types and firms, and that amenities and tastes can also be dependent on worker ability. Abstracting from vacancy posting, we can write the wage offered

by a firm as

$$\tilde{w}(p, \pi_{a,g}, z_{a,g}, a) = \quad a\,p + \pi_{a,g} - z_{a,g} \\ - \int_{\phi_{a,g}}^{\bar{p}_{a,g}} \left[\frac{1 - \Gamma_0(\phi_{a,g}) + \kappa_{a,g}(1 - \Gamma_0(p_{a,g}))}{1 - \Gamma_0(\phi_{a,g}) + \kappa_{a,g}(1 - \Gamma_0(x))} \right]^2 d\,x$$

As a consequence, under assumptions on the proportionality of amenities and employer tastes to a, we are able to derive an exact decomposition of a worker's wage:

Proposition 1. *Suppose that the outside option of all workers, the amenity levels of both genders at all firms and the level of employer taste at all firms are proportional to a worker's ability. Then, if workers of the same gender all share the same mobility parameter κ_g, wages can be written as*

$$w(p, \pi_{a,g}, z_{a,g}, a) = \underbrace{a}_{\textit{Worker Effect}} \underbrace{\left[(p - z_g) - \int_{\phi_g}^{\bar{p}_g} \left[\frac{1 - \Gamma_0(\phi_g) + \kappa_g(1 - \Gamma_{0,g}(p_g))}{1 - \Gamma_{0,g}(\phi_g) + \kappa(1 - \Gamma_{0,g}(x))} \right]^2 d\,x \right]}_{\textit{Gender-specific firm effect}}$$

where $p_g = p + \pi_g - z_g$, $z_m = 0$ for all firms and the firm effect is independent of a.

Therefore, we can write log wages as $\log w = \log a + \log F$ for each gender, where the first part is the worker effect and the second part is the gender-specific firm effect, independent of ability type.[8]

To solve the model, we extend the solution algorithm in Engbom and Moser (2018) to an environment with employer heterogeneity in productivity, amenities and discrimination, and two genders.

5 Identification

We adopt a two-step identification strategy: first, we identify the type- and gender-specific labor market parameters by constructing an ordering of firms that identifies the ladder, and by identifying worker flows across labor market states and ladder rungs. Then, we use information on wage differences across ladder rungs to identify firm-specific productivity and amenities. In this preliminary version, we assume away heterogeneity in ability and therefore we drop the subscript a for readability.

[8] In our current draft, we assume that labor market parameters are the same across ability types, but this need not be the case. When labor market parameters are different across types, the decomposition suggested in Proposition 1 will not be exact. However, Engbom and Moser (2018) show that the AKM decomposition manages to predict more than 99% of earnings variation in model-generated data.

In our model, higher-ranked firms hire relatively more workers from employment than unemployment, because there are more workers at a lower rung in the ladder who find their offers more attractive than their present one. Therefore, we start by ranking firms by their poaching index, as in Bagger and Lentz (2018):

$$r_j = \frac{N_j^{EE}}{N_j^{UE} + N_j^{EE}} \tag{19}$$

where j is the firm index, N_j^{EE} is the number of hires from employment of firm j and $N_j^{U}E$ is the number of hires from nonemployment of firm j. We use this information and the share of hires from nonemployment of firm j out of total hires from nonemployment to construct $f_{g,r}$ and $F_{g,r}$, respectively the gender-specific density function of the offer distribution at rank r and the gender-specific CDF of the offer distribution at rank r. The latter is equivalent to $F_g(\tilde{w}_{g,r})$, where $\tilde{w}_{g,r}$ is the effective wage $w + \pi$ paid by the firm to gender g at rank r. Consider a case in which $r_2 > r_1$: we consider movements from firm 1 to 2 to be upward movements, and movements from firm 2 to 1 to be downward movements. With this information, we can compute labor market transition rates.

To begin with, we identify δ_g and λ_g^u simply as the transition rates from employment to unemployment and viceversa, respectively. We must now identify λ_g^e and λ_g^G. The added difficulty is that some transitions that are induced by the "godfather shocks" are upward transitions and we must tell these apart from the upward movements induced by on-the-job offers that can be refused. To do this, first notice that all outgoing transitions from a firm at rank r can be written as

$$J2J(\tilde{w}_r) = n(\tilde{w}_r)[\lambda^e(1 - F(\tilde{w}_r)) + \lambda^G]$$

where $n(\tilde{w}_r)$ is the number of workers who are working at that effective wage or that, in other words, are working at rank r. Intuitively, outgoing transitions can happen because a worker has received a better offer while employed, or because she was hit by a godfather shock. Rearranging the expression we obtain

$$\lambda^G = \frac{J2J(\tilde{w}_r) - \underbrace{n(\tilde{w}_r)(\lambda^e + \lambda^G)(1 - F(\tilde{w}_r))}_{\textbf{\# of upward transitions}}}{n(\tilde{w}_r)F(\tilde{w}_r)}. \tag{20}$$

Therefore, we can estimate the godfather shock probability by counting only downward transi-

tions and adjusting by the probability that such a transition would occur, given that a worker is at that specific rung of the ladder. Once we know λ^G, we can use it to back out λ^e:

$$\lambda^e = \frac{\frac{J2J(\tilde{w})}{n(\tilde{w})} - \lambda^g}{1 - F(\tilde{w})} \tag{21}$$

We estimate transition rates separately for each gender. Our results are summarized in Table 11. Consistently with our previous results, we find that women exhibit lower transition rates in general, both across labor market states and job-to-job.

Table 11. Transition rates: estimated on 2005-2017 RAIS Data.

Parameter	Value
Across labor market states	
λ^u_m	0.0915
λ^u_f	0.0793
δ^m	0.0342
δ^f	0.0293
Job-to-job	
λ^e_m	0.0416
λ^e_f	0.0342
λ^G_m	0.0066
λ^G_f	0.0053

5.1 Identifying Amenities, Productivity and Employer Tastes

In this subsection we discuss in detail how we identify firm-specific amenities, productivity and employer tastes. We identify amenities by exploiting an intuition similar to Sorkin (2018): in our environment, the relationship between pay and rank is informative of how large amenities are at a certain firm. We augment this intuition by exploiting the first order conditions of our model, which imply a relationship between wages, amenities and productivity at the firm level.

To begin with, notice that in our model, if firm 2 is ranked higher than firm 1 by workers, it must be that $w_2 + \pi_2 > w_1 + \pi_1$. Therefore, by looking at wage differentials across rungs of the ladder, we can infer something about the amenity differentials across rungs of the ladder. In practice, our problem is to find a sequence $\{\pi_1, ..., \pi_R\}$ of amenities such that $w_1 + \pi_1 \leq w_2 + \pi_2 \leq ... \leq w_R + \pi_R$. We find that the most accurate solution to the problem can be obtained, in most cases, by solving a least squares problem in which we search for the sequence of amenities that

minimizes the distance between rungs in the ladder.

However, by using the first order conditions of our model, we can achieve better identification and also simultaneously identify firm-specific productivity p and employer taste z. The intuition is that, in our model, effective wages $w + \pi$ must be increasing in $p + \pi_m$ for men and $p + \pi_f - z$ for women. Therefore, we introduce the additional set of constraints that $p_{g,1} + \pi_{g,1} \leq p_{g,2} + \pi_{g,2} \leq ... \leq p_{g,R} + \pi_{g,R}$, where $p_{f,r} = p_r - z_r$ is a modified productivity term that takes into account employer taste for women. This seems to complicate our identification as we now also need to identify productivity p and employer taste z. However, by using the first-order conditions in equation 18, we can substitute for productivity:

$$p = w_m + \frac{\delta_m + \lambda_m^G + \lambda_m^e(1 - F_m(\tilde{w}_m))}{2\lambda_m^e f_m(\tilde{w}_m)}.$$

For women, our first order conditions imply that

$$p - z = w_f + \frac{\delta_f + \lambda_f^G + \lambda_f^e(1 - F_f(\tilde{w}_f))}{2\lambda_f^e f_f(\tilde{w}_f)}.$$

Thus, after recovering p and π_m by executing our algorithm on male data, we can recover π_f and z by executing our algorithm on female data given p.

Because our estimation strategy only relies on the first order conditions with respect to wages, it allows for an arbitrary distribution of vacancy posting costs between firms and across genders, which are accounted for implicitly in our data estimates of f_g and F_g.

Using our estimates of the offer distribution above, we set $F_g(\tilde{w}_g) = F_{g,r}$ and $f_g = f_{g,r}\partial_r/(\partial\tilde{w}_g)$, and rewrite the constraints as functions only of known parameters, w_g and π_g.

$$\min_{\{\pi_{g,1},...\pi_{g,N}\}} \sum (w_{g,r+1} + \pi_{g,r+1} - w_{g,r} + \pi_{g,r})^2$$

$$\text{s.t.}$$

$$w_{g,r} + \pi_{g,r} \leq w_{g,r+1} + \pi_{g,r+1} \quad \forall r \in \{1, ..., R-1\}$$

$$p_{g,r} + \pi_{g,r} \leq p_{g,r+1} + \pi_{g,r+1} \quad \forall r \in \{1, ..., R-1\}$$

The final solution are firm-specific estimates of p, π_m, π_f and z. In Appendix B.7 we present evidence that our algorithm is effective at recovering the firm-specific parameters. We perform

Monte Carlo simulations of model-generated data and subsequently run our algorithm on the data using only the information that is available to us in the RAIS data. Our algorithm recovers measures of productivity and amenities that are strongly correlated to those in the simulated data, and manages to replicate quite closely the features of the distributions.

Notice that in no case are we able to identify the mean of amenities for either gender. The reason is that, if we increase all amenities by x and the outside option ϕ_g by the same quantity x, our model produces the same wages, ranks and firm sizes for every firm. For this reason, we normalize the mean of amenities to zero, and we will not run counterfactuals on equalizing outside options between men and women, or on changing the means of amenities for one gender.

Finally, the outside options for each gender ϕ_g are automatically identified, by definition, as the lower bound of $w_g + \pi_g$ that is returned by our algorithm.

6 Results

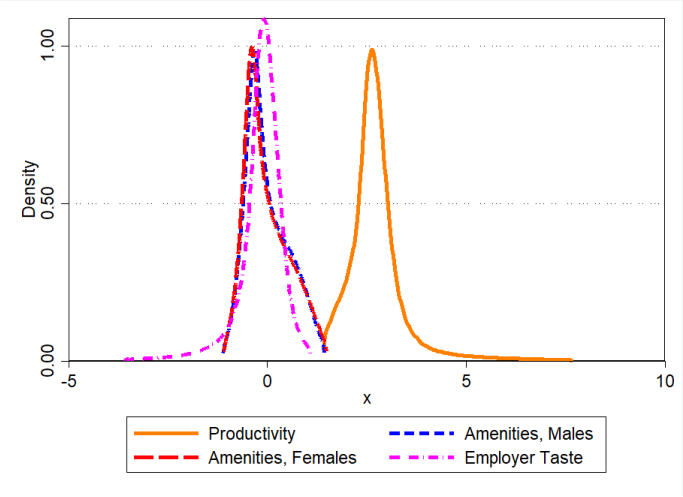
In this section we discuss our estimation results in terms of their implications for what the distributions of productivity, amenities and discrimination look like, and the relationship between wages, poaching ranks and gender gaps with productivity, amenities and employer tastes.

6.1 Estimation Results

We find that dispersion in productivity across firms is substantially larger than dispersion in amenities for either gender or in employer tastes. The dispersion in male and female amenities is very similar; as is common in models based on Burdett and Mortensen (1998), productivity exhibits a long right tail. Instead, we find that employer taste for women exhibits a long left tail.

Table 12 reports the pairwise correlation table between wages, ranks and the firm-specific parameters we estimate. We find that, as expected, productivity is strongly correlated with both male and female wages (correlations of 0.68 and 0.65 respectively). Amenities are weakly positively correlated with amenities, and strongly positively correlated across genders, with a correlation around 0.75 (see also figure 14). Also, while log wages are typically positively related to poaching ranks, amenities are stronger predictors of ranks than wages are, for both genders. Interestingly, amenities are substantially uncorrelated to wages. Another interesting result we obtain is that negative employer taste for women is negatively related to female amenities: firms that have a

Figure 13. Estimation results. Marginal distribution of p (productivity), π_m (amenities for men), π_f (amenities for women) and z (employer taste).



stronger distaste for women tend to provide less amenities to them (correlation of -0.35).

Table 12. Estimation results. Correlation table of gender-specific AKM fixed effects, p (productivity), π_m (amenities for men), π_f (amenities for women) and z (employer taste).

	(1)							
	$\log(w_m)$	$\log(w_f)$	Rank, M	Rank, F	p	π_m	π_f	z
Male AKM FE	1							
Female AKM FE	0.863	1						
Rank, M	0.434	0.422	1					
Rank, F	0.450	0.407	0.813	1				
p	0.680	0.649	0.640	0.568	1			
π_m	-0.164	-0.138	0.776	0.578	0.248	1		
π_f	-0.0375	-0.139	0.613	0.797	0.256	0.749	1	
z	-0.221	-0.341	-0.300	-0.476	-0.392	-0.211	-0.354	1

Poaching ranks for each gender are mostly increasing in wages: however, we find that some firms that are at the bottom of the pay scale are ranked on average higher than most other firms, suggesting that they compensate low pay with higher amenities (figure 15). Similarly, we find that amenities are typically increasing in poaching ranks, but that top-ranked firms offer on average lower amenities than many other firms (figure 16).

As expected, the gender gap at the firm level is increasing in employer taste z, with substantial dispersion induced also by differences in amenities, labor market parameters and in the relationship to other ladder rungs (figure 17).

Figure 14. Estimation results. Female amenities vs male amenities, with linear fit.

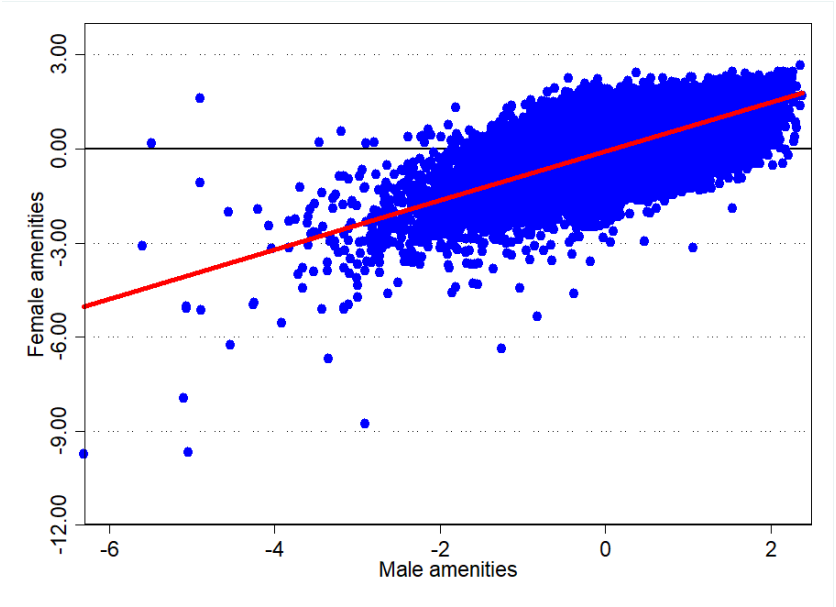


Figure 15. Estimation results. Poaching rank vs male AKM firm fixed effects (left pane) and vs female AKM firm fixed effects (right pane).

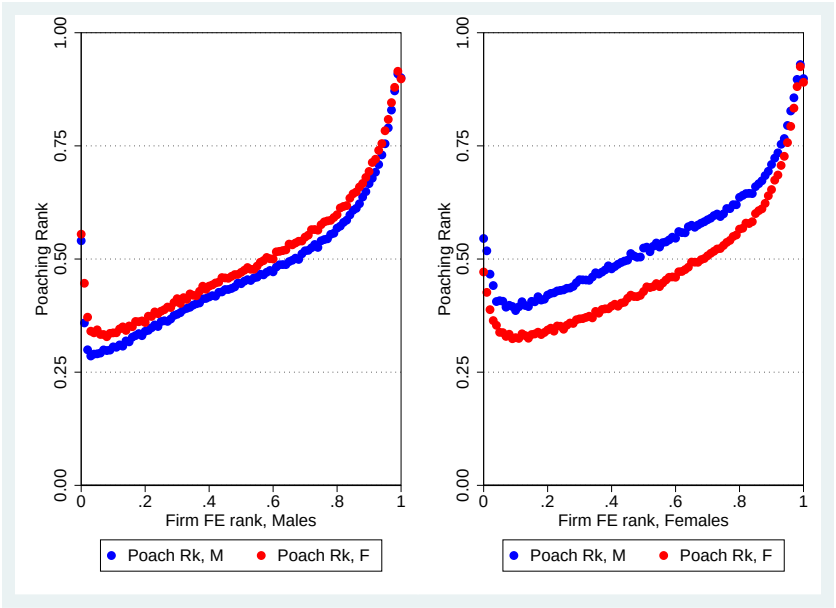


Figure 16. Estimation results. Amenities vs male poaching rank (left pane) and vs female poaching rank (right pane).

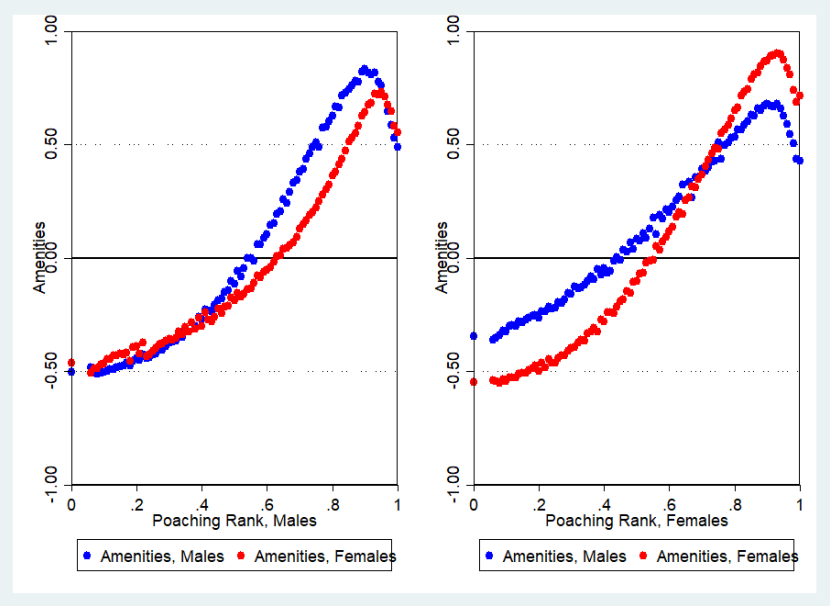


Figure 17. Estimation results. Gender gap in firm AKM fixed effects vs employer taste z.

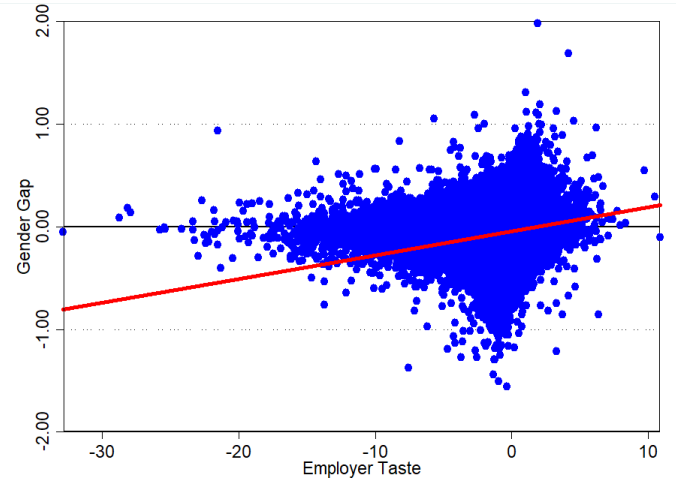


Table 13. Regression of estimated amenities π_m (left column) and π_f (right column) on firm-specific measures of amenities.

	Men	Women
Employer provides food stamps	0.820*** (0.020)	0.614*** (0.010)
Share of own gender among employees	-0.026 (0.041)	0.234*** (0.022)
Share part-time (<40h) workers	0.042 (0.049)	-0.182*** (0.025)
Share of workers fired for unjust cause	-3.093*** (0.071)	-2.065*** (0.038)
Share with ≥10% wage growth	1.882*** (0.051)	0.430*** (0.026)
Income risk (actual/contract. earnings)	-0.039** (0.018)	-0.361*** (0.010)
Share with paternity leave	-21.455*** (4.765)	1.795*** (0.208)
Share with accident	3.209*** (0.413)	-1.944*** (0.277)
Share with work-related illness	-2.754*** (0.879)	-3.606*** (0.388)
Share with unpaid leave	-1.111* (0.627)	1.160*** (0.319)
Other controls	yes	yes
R^2	0.105	0.119
Number of observations (employers)	442,727	442,093

Note: ***, **, and * denote significance at the 1%, 5%, and 10% level. Units are std. dev.s of estimated AKM estab. FEs.

6.2 Amenities and observables

Our estimates of amenities have been identified as residuals that rationalize the discrepancy between wage ranks and poaching ranks. In this subsection we investigate the relationship of our estimates of amenities with measures of nonpay characteristics that can be interpreted as measures of amenities. If our estimates are a good measure of amenities, these should positively covary with nonpay characteristics that are desirable to workers, and negatively to nonpay characteristics that are typically perceived as undesirable. Our results are summarize in Table 13

We find that our estimates of amenities align well with measures of nonpay characteristics. For instance, the fact that employers provide food stamps to their employees is positively associated to our measures of amenities. Similarly, workers of all genders value negatively the fact that, at a certain employer, many workers fall sick to work-related illnesses, or that many workers are fired for unjust cause.

Some attributes exhibit opposite signs in their association to male and female amenities. For instance, having many colleagues of the same sex on paternity leave is seen as positively by females but negatively by males, possibly because male workers see paternity leave by itself less positively, and are unhappy to be forced to "pick up the slack". Analogously, men seem to value dangerous jobs (those with a higher share of accidents) while women value them negatively, suggesting the

importance of gender-specific culture in explaining job-related amenities. Finally, women seem to value working at a firm with a higher female share, while the male share has a statistically insignificant association to male amenities.

6.3 Model Implications

In this subsection we investigate the quantitative importance of each source of gender gaps to explain the data. To do so, we first solve and simulate our model in the baseline scenario, feeding to the model the parameters we have previously estimated and the joint distribution of productivity, amenities for each gender and employer tastes that we observe in the data. Then, we shut down differences across genders one by one, keeping constant the cost of posting vacancies and solving the general equilibrium of the model under the changed scenario.[9] Our results are summarized in Table 14.

Table 14. Counterfactual simulations. Model results when sources of differences between men and women are shut down. Baseline results (column 1) against differences in amenities shut down (2), employer tastes shut down (3), different rates of J2J transition shut down (4) and all combinations (5-8).

		Differences Across Genders						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Amenities	✓	-	✓	✓	-	-	✓	-
Discrimination	✓	✓	-	✓	-	✓	-	-
J2J arrival rate	✓	✓	✓	-	✓	-	-	-
Var(Gender Gap) (Δ%)	0.10	(-9.17)	(-60.35)	(4.18)	(-64.53)	(-4.77)	(-67.16)	(-70.15)
Var(w^f) (Δ%)	0.47	(-8.54)	(-29.96)	(5.40)	(-35.76)	(-3.46)	(-27.98)	(-33.60)
β(GWG, rank w^m)	0.13	0.18	0.24	0.10	0.27	0.14	0.21	0.24
β(GWG, female share)	-1.27	-1.21	2.27	-1.40	2.12	-1.33	2.21	2.05
Std dev.(female share)	0.15	0.15	0.07	0.16	0.07	0.16	0.07	0.07
JF Rate, Females	0.08	0.08	0.08	0.07	0.08	0.07	0.07	0.07
Gender Wage Gap (Δ%)	-	(-1.70)	(19.35)	(-9.70)	(17.04)	(-11.30)	(11.12)	(8.83)

We investigate the impact of changes in model parameters on the average gender gap, the variation in gender gaps at the firm level, the variance in women's wages and the relationship between gender gaps and female shares, as well as the job-finding rate of women. We perform a total of seven experiments: in the first, we set amenities of women equal to those of men at the firm level. In the second, we set all firm-specific employer tastes $z = 0$. In the third, we set $\lambda_f^e = \lambda_m^e$,

[9] In principle this also involves recalibrating the outside option of male and female workers across experiments, which is allowed to change in response to changes in parameters. In this preliminary version we keep the outside option constant for simplicity but we plan to allow for this change in future versions.

thus making women as fast as men in climbing the job ladder. Because of our estimation strategy based on AKM fixed effects, we need to impose a normalization of the means of fixed effects. We initially normalize the difference between male and female AKM firm fixed effects to zero, so that the gender wage gap implied by the model will not be equal to the gender wage gap seen in the data. For this reason, our main interest in this exercise lies in the proportion of the variance of gender gaps that we will be able to explain.

We find that employer tastes and differences in amenities are particularly important determinants of variation in gender wage gaps at the firm level. Heterogeneity in employer tastes alone accounts for around 60% of the variation in gender wage gaps across firms, and for 30% of the variance in female wages.[10] Differences in amenities across genders account for 9% of the variation in gender gaps and for roughly 9% of the variance of female wages.

Shutting down differences between men and women in climbing the ladder reduces the average gender gap by approximately 10 percentage points, meaning that differences in the speed of climbing account for almost half of the total gender gap. The impact of different speeds on wage variation is however more limited, with differences in λ_g^e accounting for 4% (5%) of the variance of gender gaps (female wages).

All these differences together explain around 70% of the variation in gender gaps across firms, with the remainder being accounted for by differences in separation rates, outside options and godfather shock arrival rates.

Most fundamentals have a negligible impact on the job-finding rate of women and on the dispersion of female shares, except for employer tastes: in particular, shutting down dispersion in employer tastes more than halves the standard deviation of female shares implied by the model.

7 Discussion

7.1 Motherhood and maternity leave

First, we investigate to what extent the event of women giving birth to children may explain some of the previously documented facts. The unequal incidence of parenthood across mothers and

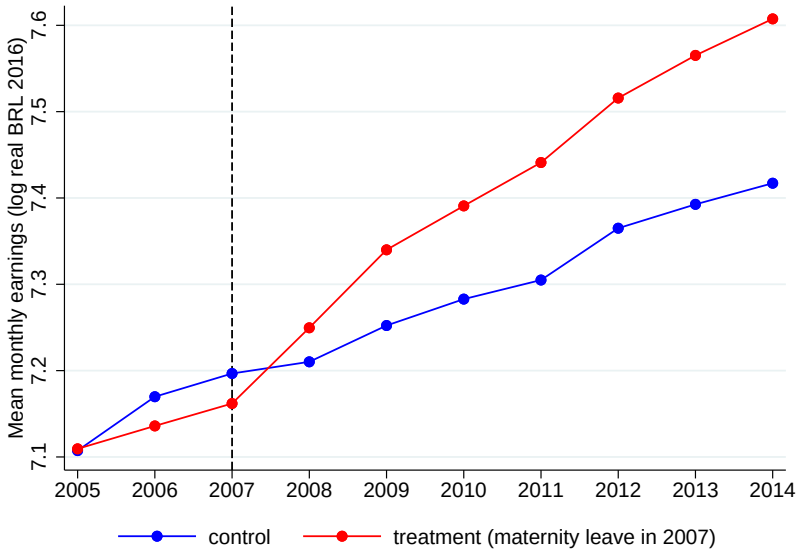
[10] We find that, on average, firms seem to have a taste for women rather than men, so that shutting down employer tastes *increases* the gender gap by 19 percentage points. This somewhat surprising result is due to the initial normalization of AKM firm fixed effects. We plan to run robustness checks of our results under different normalizations of AKM fixed effects.

fathers is an often-cited factor associated with labor market inequalities (Kleven et al., 2016b). To explore the role of motherhood in our context, we exploit information in our administrative data on maternity and paternity leaves from work taken during the period 2007–2014. To this end, we construct an event study of women in the years immediately before and after taking maternity leave in 2007.

Figure 18 shows the result of this study. In blue is the mean earnings trajectory for the "control" group that consists of women who did not have a child in 2007, while the red line shows the same trajectory for the "treatment" group of women who took maternity leave in 2007. We find no evidence of a pay gap arising aroun the time of maternity leave in this event study approach. If anything the wage growth of mothers, conditional on remaining emloyed in the subsequent years, accelerates relative to that of women who did not take a leave.

Of course this result should be taken with a grain of salt. An expected caveat is that labor force participation drops significantly for women around the time of motherhood, suggesting that there are earnings losses occurring at the extensive rather than at the intensive margin. A second caveat is that in the presence of dynamic selection, those women remaining in the labor force are a (positively) selected subgroup of all the women who gave birth to a child in 2007. A third caveat is that statistical discrimination of motherhood may be already priced into wages at the time of hiring and consequently remain undetected in this simple event study approach. For these reasons, we consider the finding that motherhood is not associated with wage losses as tentative and hopefully the subject of future investigations.

Figure 18. Event study of female pay gap around maternity leave



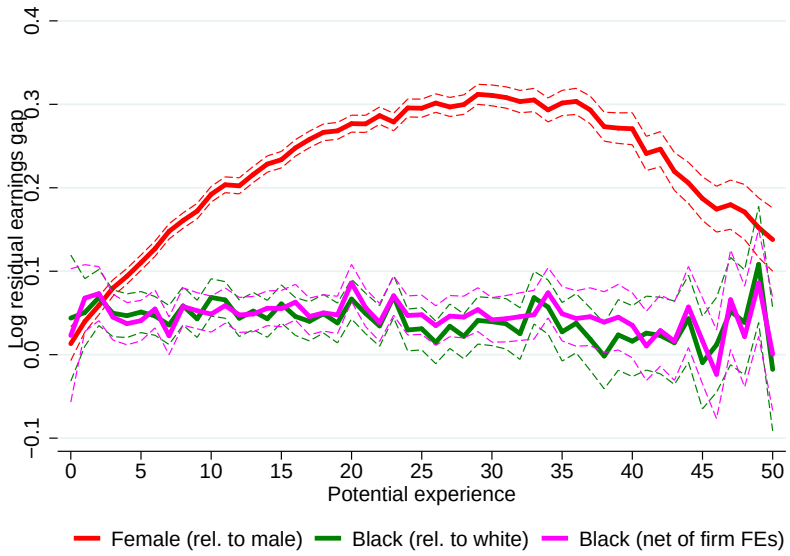
Notes: 2005-2014 Brazilian formal sector census data (RAIS).

7.2 Comparison with racial earnings gaps

In previous sections, we highlighted the pronounced life cycle pattern of the female earnings gap and its relation to gender differences in firm pay and labor market flows. An interesting comparison group in this context is the population of non-white minorities in Brazil, which are comprised of black, "brown" (*pardo*, in Portuguese), Asian, and native indigenous individuals. A natural question to ask is then: Is there anything special about women's labor market experience?

The answer to this question, it turns out, is 'yes': the female earnings gap is fudamentally different to that for other subpopulations. Figure 19 compares the life cycle pattern of gender versus racial pay gaps. The red line, which is reproduced from a previous figure, shows the life cycle earnings gap for women. In green is the same life cycle earnings gap reproduced for black workers. In constrast with women's experience, while blacks experience a significant average earnings gap of around 5 log points over their life, this gap shows little variation throughout their career. Similarly contrasting to women's experience, the pink line shows that in turn little of the racial pay gap is explained by firm fixed effects or the differential allocation of black workers toward lower-paying firms. In other words, there seems to be something like a distinctly female firm in Brazil, but nothing like a distinctly black firm.

Figure 19. Comparing life cycle pattern of gender versus racial pay gaps

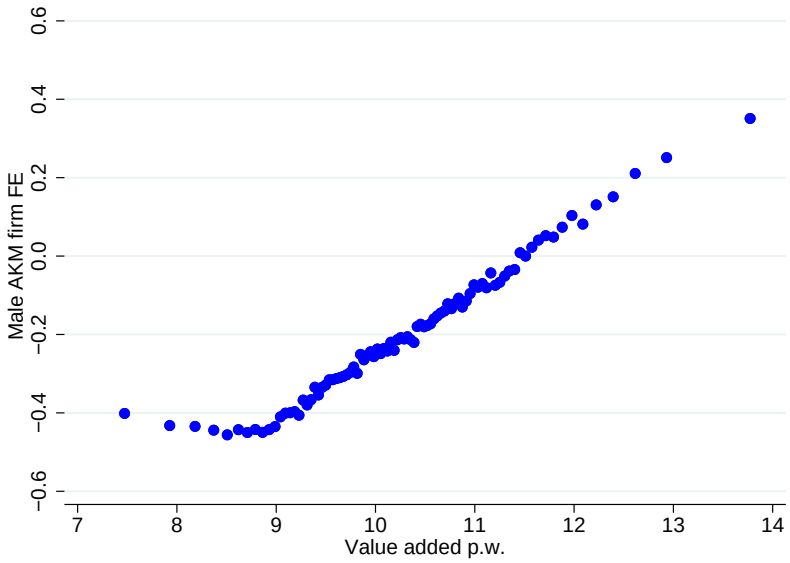


Notes: Brazilian formal sector census data (RAIS) for 2003.

7.3 Estimating Losses from Misallocation

In the previous section, we dissected the micro-sources of the female earnings gap in Brazil, highlighting firm heterogeneity as an important dimension. Although firm pay differences have been studied widely, little connection has been made to the macroeconomic consequences of employer heterogeneity and in particular the female earnings gap. What bridges these two fields is the insight that more productive firms pay more, and co-existence of more and less productive firms may be connected to misallocation of resources in the economy, including the innate talent and human capital embedded in workers of both genders. Indeed, in previous work Alvarez et al. (2018) showed that the most important predictor of firm-level pay is firm-level labor productivity, which they find explains approximately 60% of variation in male firm fixed effects in Brazil's manufacturing and mining sectors. Figure 20 replicates a key graph from their paper, showing the close-to-linear relation between estimated male firm fixed effects in a two-way fixed effects framework due to Abowd et al. (1999a) on the one hand, and firm-level value added per worker on the other hand. The graph shows that sizable labor productivity differences exist in Brazil and that these differences are systematically related to firm-level pay dispersion.

Figure 20. Relationship between estimated firm pay and firm productivity



Notes: 2000-2004 Brazilian formal sector census data (RAIS) and manufacturing & mining firm survey PIA data. Dependent variable is estimated AKM male firm fixed effect from previous regression. Weighted by number of employees. Source: Alvarez et al. (2018).

Given that women systematically work at lower-paying firms and that these lower-paying firms are also less productive, the connection between the gender earnings gap and productivity losses from misallocation becomes evident. A simple back-of-the-envelope calculation then allows us to compute productivity gains from allocating women like men. To be clear, there may be many other reasons not considered here explaining why women work at lower-paying firms on average, besides productivity.[11] With this caveat in mind, we still think that this back-of-the-envelope calculation is a useful tool in guiding our thinking about how large the output losses from the misallocation of female talent can be.

First, recall that the estimated difference in gender-specific firm fixed effects between employers is $\mathbb{E}\left[\alpha_j \middle| male\right] - \mathbb{E}\left[\alpha_j \middle| female\right] = 0.078$. Second, from Figure 20 we see that the male firm fixed effect slope in labor productivity is $\partial \mathbb{E}\left[\alpha_j \middle| male\right] / \partial \left[\text{value added p.w.}\right] = 0.197$. Putting the two pieces together, we can compute the hypothetical potential output gains from allocating women

[11] To mind come differences in firm-specific amenities, costs of switching jobs for workers and costs of recruiting additional employees for firms, and nonlinearities between men and women as well as in total employment in firms' production functions.

across firms to match the male worker distribution to be $0.078/0.197 = 39.6$ percent.

To put this number into context, we can compute the gains from increasing female labor force participation (LFP) to match the male LFP rate. While this calculation is subject to much of the same potential caveats as the previous calculation, comparing the numbers obtained under both calculations provides a useful benchmark. First, recall that the relative female LFP rate is $(\text{Female LFP}) \,/\, (\text{Male LFP}) = 38/62$. Second, note that the female population ratio is close to 0.50 in Brazil. Putting the two pieces together, we can compute the potential output gain from drawing women into the labor force to match the male LFP rate to be $(62/38 - 1) \times (1/2) = 31.6$ percent.

We conclude that the observed gender earnings gap reflects sizable output losses from misallocation of female workers across firms, and that the gains from undoing this type of misallocation are of the same order of magnitude as those from drawing more women into the labor force.

7.4 Comparison with the gender earnings gap in other countries

The gender earnings gap we document for Brazil's formal sector market is striking, and quantitatively larger than in many higher-income economies like the United States amd Portugal, among others. But the general life cycle pattern we document for women relative to men in Brazil accords well with existing evidence from the United States and elsewhere. Our finding on the relation between the female earnings gap and firm heterogeneity qualitatively matches recent findings in the United States (Barth et al., 2017).

8 Conclusion

Large and persistent gender gaps in labor market outcomes remain in almost every country. Studying the gender earnings gap in Brazil, we similarly find sizable pay differences between genders. Our contribution is to establish a new set of facts characterizing gender gaps at the micro-levle and using these estimates to quantify output losses from misallocation at the macro-level. We show that firm heterogeneity is quantitatively important in explaining participation and earnings gaps: 50% of the gender earnings gap is between firms.

The fact that higher-paying firms are also more productive suggests that reallocation of women toward better-paying firms would result in additional productivity and output gains. Naturally, these implications will depend crucially on the sources of firm heterogeneity in the labor mar-

ket. To this end, we develop a model of frictional labor markets with on-the-job search and fit it to the Brazilian linked employer-employee data, RAIS. We find that all three sources of gender gaps—amenities, discrimination, and mobility—play an important role. In particular, differences in the speed of job-to-job transitions between men and women can explain around half of the average gender gap, while more than half of its variation across firms can be explained by heterogeneity in employer tastes for men and women. We find that combining taste-based discrimination with gender-specific amenities is important to capture key features of the data, among them that gender gaps are unrelated to firm-level female shares.

In conclusion, our work provides a new, firm-level perspective on gender gaps in labor market outcomes and highlights that a mix of different sources are important to understand the observed gaps.

References

Abowd, John M, Francis Kramarz, and David N Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Abowd, John M., Hampton Finer, and Francis Kramarz, "Individual and Firm Heterogeneity in Compensation: An Analysis of Matched Longitudinal Employer-Employee Data for the State of Washington," in John C. Haltiwanger, Julia I. Lane, James R. Spletzer, Jules J.M. Theeuwes, and Kenneth R. Troske, eds., *The Creation and Analysis of Employer-Employee Matched Data*, Amsterdam: North Holland, 1999, pp. 3–24.

Acemoglu, D and D Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, 2011, *4*.

Adda, J, C Dustmann, and K Stevens, "The Career Costs of Children," *Journal of Political Economy*, 2017, *125* (2), 293–337.

Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser, "Firms and the Decline of Earnings Inequality in Brazil," *American Economic Journal: Macroeconomics*, 2018, *10* (1), 149–189.

Autor, D, F Levy, and R Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 2003, *118* (4), 1270–1333.

Bagger, Jesper and Rasmus Lentz, "An Empirical Model of Wage Dispersion with Sorting," *Review of Economic Studies*, 2018.

Barth, E, S Kerr, and C Olivetti, "The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data," *Working Paper*, 2017.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman, "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States," *Journal of Labor Economics*, 2016, *34* (S2), S67–S97.

Becker, G, *The Economics of Discrimination*, Economic Research Studies, 1957.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz, "Dynamics of the Gender Gap for Young Professionals in the Corporate and Financial Sectors," *American Economic Journal: Applied Economics*, 2010, *2*, 228–255.

Blau, F and L Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Burdett, Kenneth and Dale T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, *39* (2), 257–273.

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*, 2016, *131* (2), 633–686.

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

Engbom, Niklas and Christian Moser, "Earnings Inequality and the Minimum Wage: Evidence from Brazil," *Federal Reserve Bank of Minneapolis OIGI Working Paper 7*, 2018.

Erosa, A, L Fuster, and D Restuccia, "A Quantitative Theory of the Gender Gap in Wages," *European Economic Review*, 2016, *85* (C), 165–187.

_ , _ , G Kambourov, and R Rogerson, "Hours, Occupations, and Gender Differences in Labor Market Outcomes," *Working Paper*, 2017.

Guner, N, R Kaygusuz, and G Ventura, "Taxation and Household Labour Supply," *Review of Economic Studies*, 2012, *79* (3), 1113–1149.

Guvenen, Fatih, Greg Kaplan, and Jae Song, "The Glass Ceiling and the Paper Floor : Gender Di ff erences among," *Working Paper*, 2014, pp. 1981–2012.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Kleven, H, C Landais, and J Sogaard, "Children and Gender Inequality: Evidence from Denmark," *Working Paper*, 2016.

Kleven, Henrik Jacobsen, Camille Landais, and Jakob Egholt Søgaard, "Children and Gender Inequality: Evidence from Denmark," *Working Paper*, 2016.

Mincer, Jacob, *Schooling, Experience, and Earnings*, New York: NBER Press, 1974.

Moscarini, Giuseppe and Fabien Postel-Vinay, "The cyclical job ladder," *Annual Review of Economics*, 2018, *10*, 165–188.

Ngai, L and B Petrongolo, "Gender Gaps and the Rise of the Service Economy," *American Economic Journal: Macroeconomics*, 2017, *9* (4), 1–44.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality," 2016.

Sorkin, Isaac, "Ranking Firms Using Revealed Preference," *Quarterly Journal of Economics*, 2018.

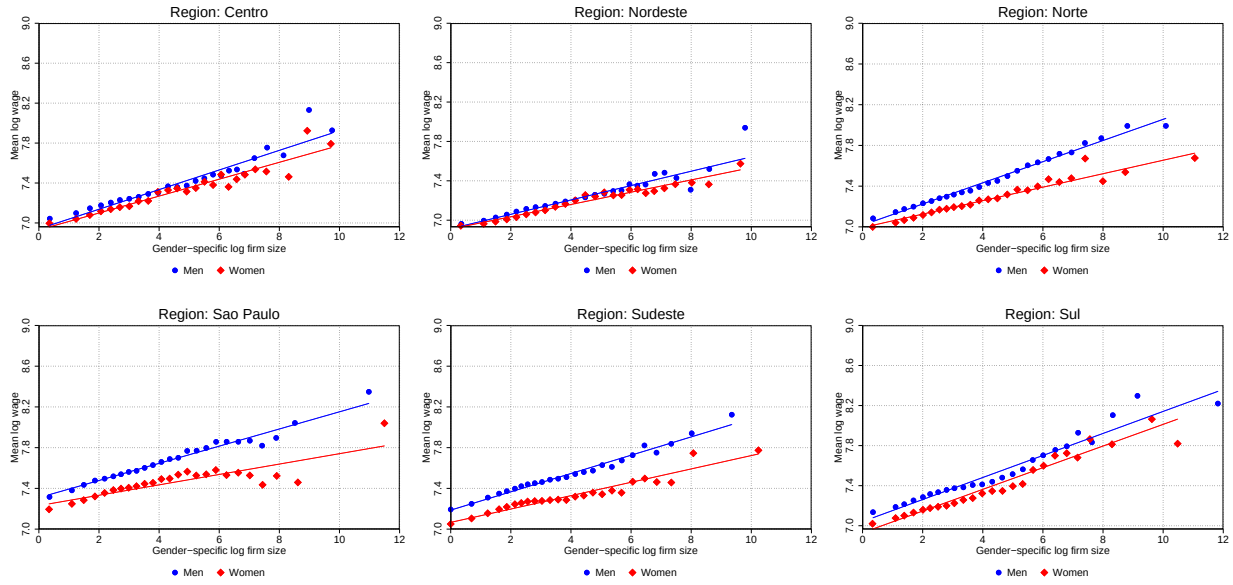
tai Hsieh, Chang, Erik Hurst, Charles I Jones, and Peter J Klenow, "The Allocation of Talent," *Working Paper*, 2016.

Appendix

A Data Appendix

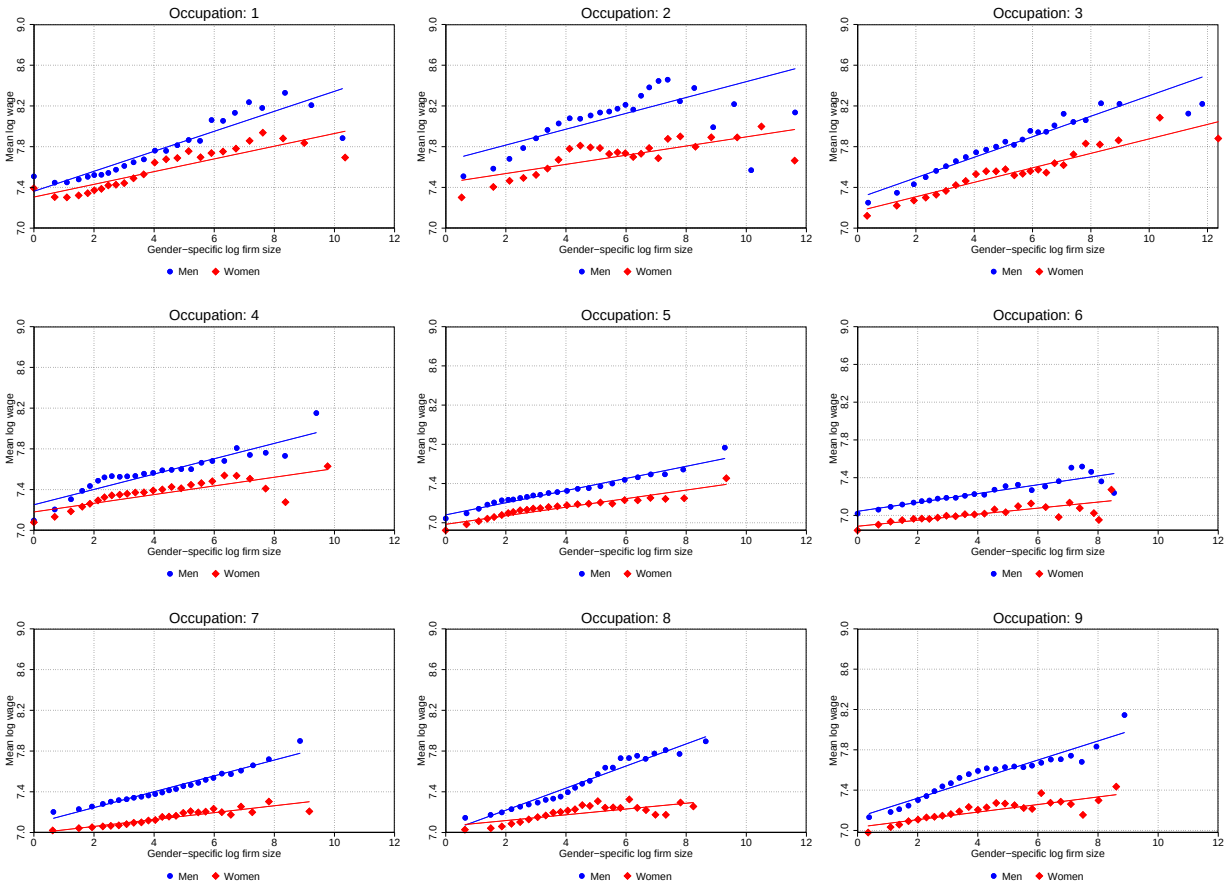
A.1 Gender-specific firm size-pay premium by different subgroups

Figure 21. Binscatter plot of gender-specific mean log wage against log firm size, by region



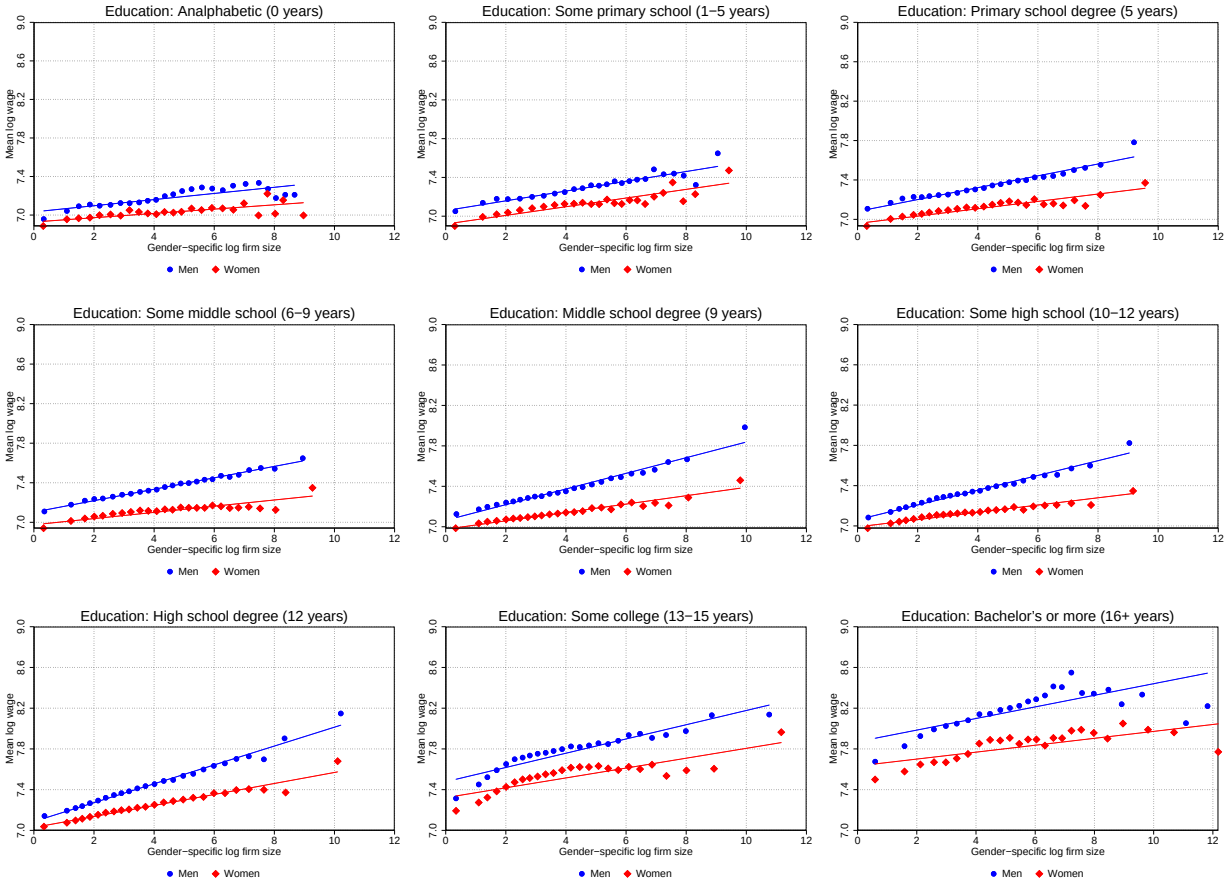
Notes: Brazilian formal sector census data (RAIS) for 2012.

Figure 22. Binscatter plot of gender-specific mean log wage against log firm size, by occupation



Notes: Brazilian formal sector census data (RAIS) for 2012.

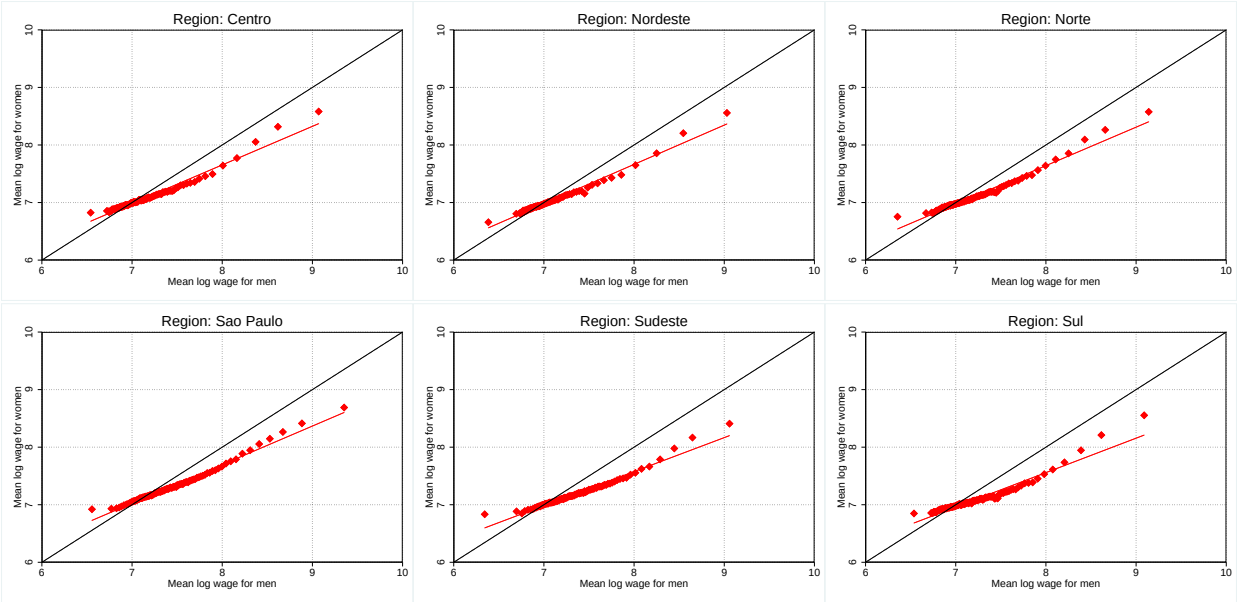
Figure 23. Binscatter plot of gender-specific mean log wage against log firm size, by education



Notes: Brazilian formal sector census data (RAIS) for 2012.

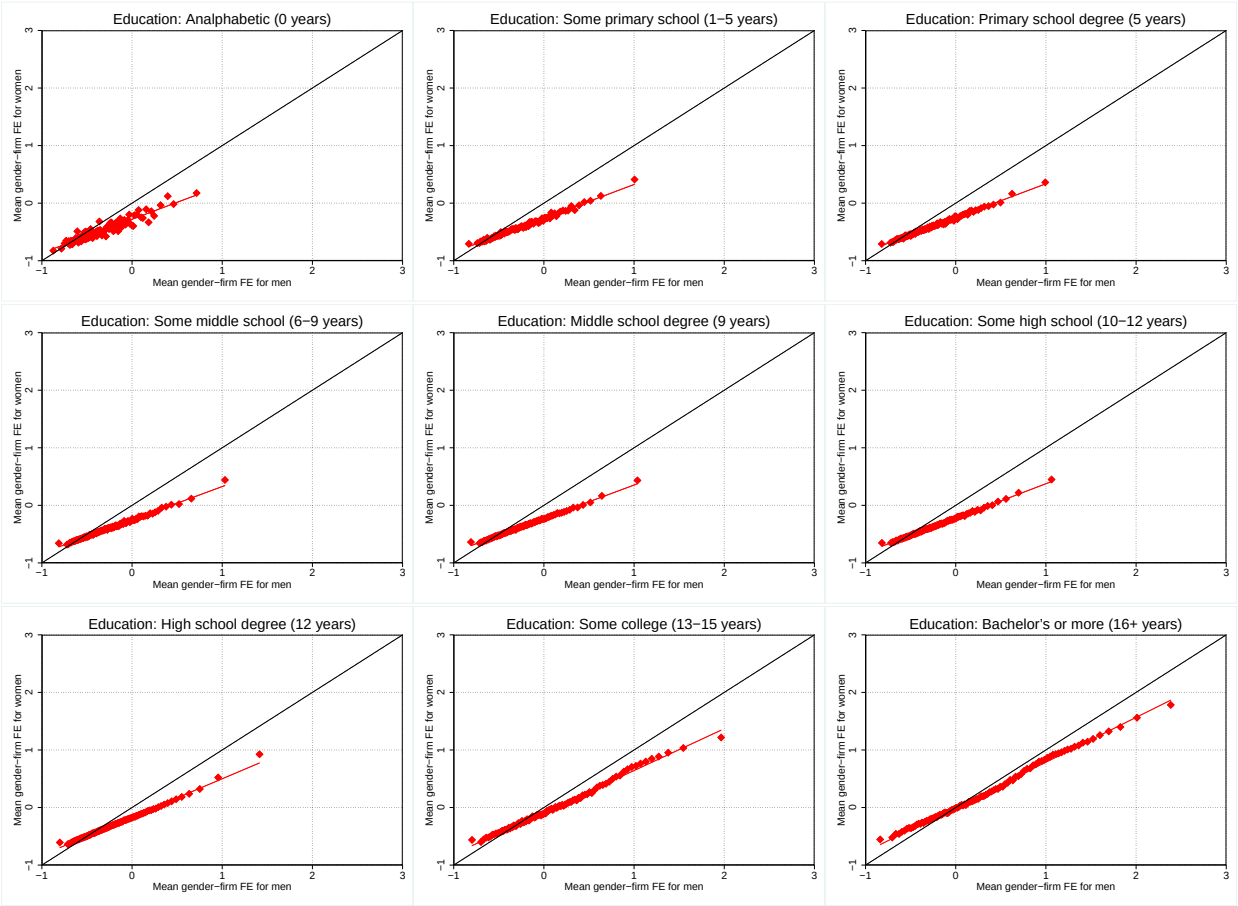
A.2 Gender-firm fixed effect for women versus men by different subgroups

Figure 24. Average female firm fixed effect across male firm fixed effects distribution, by region



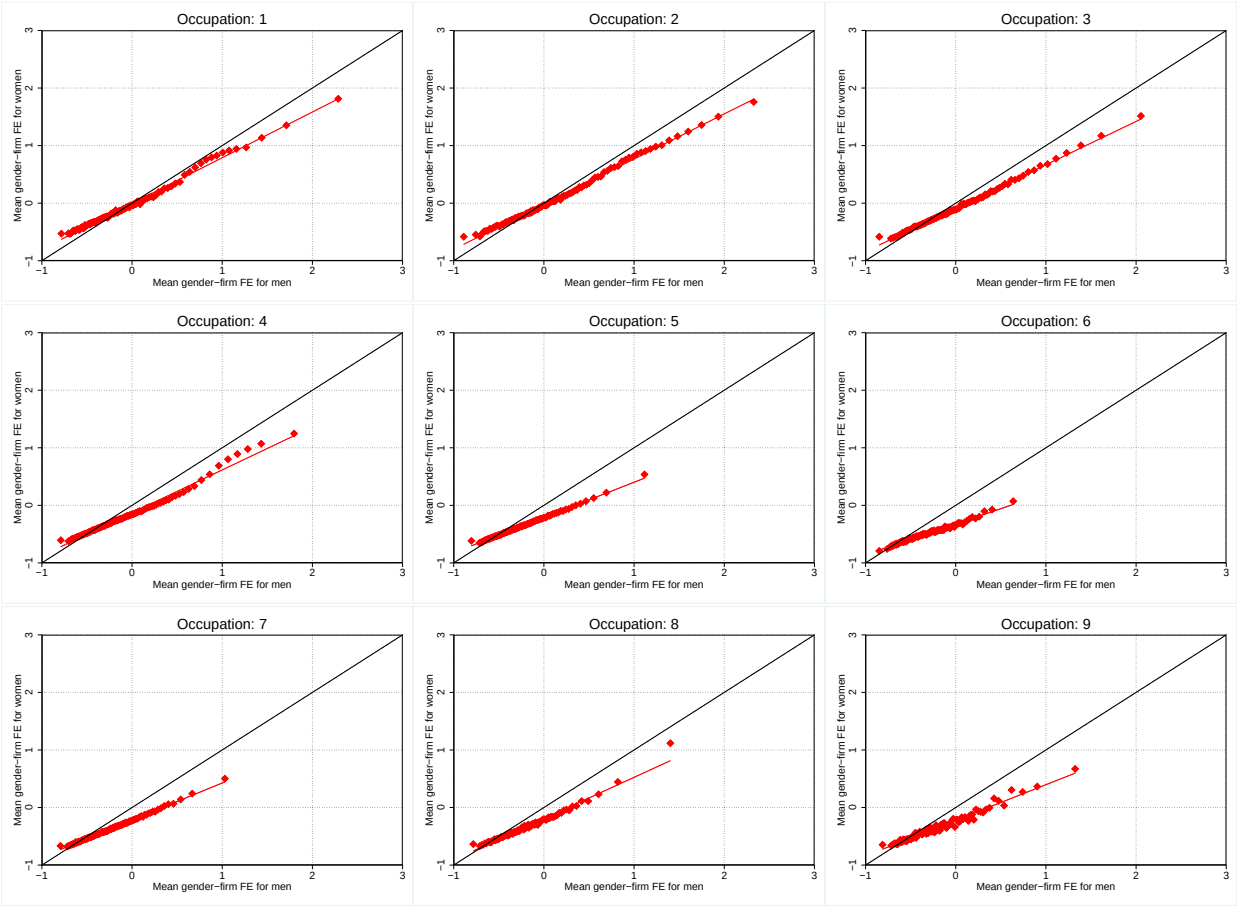
Notes: Brazilian formal sector census data (RAIS) for 2012.

Figure 25. Average female firm fixed effect across male firm fixed effects distribution, by education group



Notes: Brazilian formal sector census data (RAIS) for 2012.

Figure 26. Average female firm fixed effect across male firm fixed effects distribution, by occupation



Notes: Brazilian formal sector census data (RAIS) for 2012.

B Model Appendix

B.1 Implications of Alternative Assumptions on Vacancy Posting

Suppose that, rather than assuming that the cost of vacancy posting is separable in male and female vacancies, we assume that the vacancy posting cost reads

$$c(v_{a,m}, v_{a,f}) = c(v_{a,m} + v_{a,f})$$

where the function c retains the properties discussed in the paper: $c(0) = 0$, $c' > 0$, $c'' > 0$. It is easy to prove that, in this case, any firm will employ only men, only women, or be indifferent

between the two. To see this, notice that the problem of the firm now can be written as

$$\max_{w_{a,m},w_{a,f},v_{a,m},v_{a,f}} \quad \Big\{ (p\,a - w_{a,m}) l_{a,m}(w_{a,m} + \pi_m, v_{a,m}) + (p\,a - z - w_{a,f}) l_{a,f}(w_{a,f} + \pi_f, v_{a,f}) - c(v_{a,m} + v_{a,f}) \Big\} \tag{22}$$

The FOC for vacancy posting now read:

$$c'(v_{a,m}) = T_{a,m}(ap + \pi_m - \tilde{w}_{a,m}) \left(\frac{1}{\delta_{a,m} + s_{a,m}\lambda_a^{m,u}(1 - F_{a,m}(\tilde{w}_{a,m}))} \right)^2$$
$$c'(v_{a,f}) = T_{a,f}(a\,p + \pi_f - z - \tilde{w}_{a,f}) \left(\frac{1}{\delta_{a,f} + s_{a,f}\lambda_a^{f,u}(1 - F_{a,f}(\tilde{w}_{a,f}))} \right)^2$$

The left-hand side is the marginal cost of an additional vacancy, which is equated to the marginal benefit of an additional vacancy (the right-hand side): an increase in the labor force of that gender for the firm multiplied by the profits made by that worker. The right-hand side of both expressions is independent of the amount of vacancies posted and can be treated as a constant from the vacancy posting perspective. This is because wages are set according to other first-order conditions, which do not depend on the amount of vacancies posted by that firm. Therefore, it is clear that it is not possible for the firm to equate at the same time the two expressions, except for the rare case in which the two marginal benefits are equal. Thus, in virtually all cases, any solution to the firm's problem must be a corner solution in which the firm hires only men or only women. The intuition is that the amount of vacancies of either gender does not affect the marginal benefit of a male or female worker, and as a consequence the firm will only hire the gender that gives the highest marginal benefit. This implication is clearly counterfactual as most firms in our data have a mixed-gender labor force composition.

B.2 The Joint Vacancy Posting Model (JVPM)

Under joint vacancy posting, firms cannot post gender-specific vacancies but only type-specific ones. In that case, their problem becomes:

$$\max_{w_{a,m},w_{a,f},v_a} \quad \Big\{ (p\,a - w_{a,m}) l_{a,m}(w_{a,m} + \pi_m, v_a) + (p\,a - z - w_{a,f}) l_{a,f}(w_{a,f} + \pi_f, v_a) - c_a(v_a) \Big\} \tag{23}$$

Similarly to the separate vacancy posting model, we can rewrite the problem in terms of offering effective wages $\tilde{w} = w + \pi$:

$$\begin{aligned}\max_{\tilde{w}_{a,m},\tilde{w}_{a,f},v_a} \quad & \Big\{(p\,a + \pi_m - \tilde{w}_{a,m})l_{a,m}(\tilde{w}_{a,m}, v_a) \\ & + (p\,a + \pi_f - z - \tilde{w}_{a,f})l_{a,f}(\tilde{w}_{a,f}, v_a) - c_a(v_a)\Big\}\end{aligned} \tag{24}$$

Therefore, firms place themselves on a ladder according to $\tilde{w}$.

Notice that we do not impose that firms hire both genders in each submarket: it is always possible for a firm to offer an effective wage $\tilde{w}_{a,g} < \phi_{a,g}$ such that no worker of gender g will accept it. The consequence is that, while a total of V_a vacancies are posted in each submarket in the aggregate, only $V_{a,g} \leq V_a$ vacancies are available to workers of each gender.

This implies that the matching function becomes

$$m_{a,g} = A^g[\mu_{a,g}(u_{a,g} + \lambda^e_{a,g}(1 - u_{a,g}))]^\alpha V^{1-\alpha}\frac{V_{a,g}}{V} \tag{25}$$

which includes the probability that a worker of gender g will meet a vacancy which has too low a wage, therefore automatically rejecting it. It is straightforward to show that this matching function exhibits all the properties of standard matching functions, i.e. $f_{a,g}/q_{a,g} = V/[u_{a,g} + \lambda^e_{a,g}(1 - u_{a,g})]$, where $f_{a,g} = m_{a,g}/[u_{a,g} + \lambda^e_{a,g}(1 - u_{a,g})]$ and $q_{a,g} = m_{a,g}/V$.

Thus, the total number of vacancies posted in submarket $\{a\}$ for each gender g is equal to

$$V_{a,g} = \int v_a(p, \pi, z)\, dD_{a,g}(p, \pi, z) \tag{26}$$

B.3 Equilibrium of the JVPM

Define $\lambda^e_{a,g} = s_{a,g}\lambda^u_{a,g}$. The following equation represents the law of motion of firm sizes:

$$\begin{aligned}\dot{l}_{a,g}(\tilde{w}, v) = & -\delta_{a,g}l^g_a(\tilde{w}, v) - s^g_a\lambda^e_{a,g}(1 - F_{a,g}(\tilde{w}))l^g_a(\tilde{w}, v) + \\ & vq_{a,g}\left[\frac{u_{a,g}}{u_{a,g} + (1 - u_{a,g})s_{a,g}} + \frac{(1 - u_{a,g})s_{a,g}}{u_{a,g} + (1 - u_{a,g})s_{a,g})}G_{a,g}(\tilde{w})\right]\end{aligned} \tag{27}$$

Solving for the stationary solution:

$$l_{a,g}(\tilde{w}_{a,g}, v_a) = \left(\frac{1}{\delta_{a,g} + s_{a,g}\lambda^u_{a,g}(1 - F_{a,g}(\tilde{w}_{a,g}))}\right)^2 \frac{v_a}{V_a}\mu_{a,g}u_{a,g}\lambda^u_{a,g}(\delta_{a,g} + s_{a,g}\lambda^u_{a,g}) \tag{28}$$

B.4 Equilibrium Characterization of the JVPM

To find the firm's policy functions, define $T_{a,g} = \mu_{a,g}[u_{a,g}\lambda^u_{a,g}(\delta_{a,g} + s_{a,g}\lambda^u_{a,g})]/V_a$. we rewrite the firm's problem as a function of the steady state mass of employed workers as follows:

$$\begin{aligned}\max_{\tilde{w}_{a,m},\tilde{w}_{a,f},v_a} \quad & \left\{ T_{a,m} v_a (a\,p + \pi_m - \tilde{w}_{a,m}) \left(\frac{1}{\delta_{a,m} + s_{a,m}\lambda_a^{M,u}(1 - F_{a,m}(\tilde{w}_{a,m}))} \right)^2 \right. \\ & \left. + \left\{ T_{a,f} v_a (a\,p + \pi_f - z - \tilde{w}_{a,f}) \left(\frac{1}{\delta_{a,f} + s_{a,f}\lambda_a^{W,u}(1 - F_{a,f}(\tilde{w}_{a,f}))} \right)^2 - c_a(v_a) \right\} \right.\end{aligned} \tag{29}$$

The associated FOC read:

$$\begin{aligned} c'(v_a) &= T_{a,m}(ap + \pi_m - \tilde{w}_{a,m}) \left(\frac{1}{\delta_{a,m} + s_{a,m}\lambda_a^{m,u}(1 - F_{a,m}(\tilde{w}_{a,m}))} \right)^2 \\ &\quad + T_{a,f}(a\,p + \pi_f - z - \tilde{w}_{a,f}) \left(\frac{1}{\delta_{a,f} + s_{a,f}\lambda_a^{f,u}(1 - F_{a,f}(\tilde{w}_{a,f}))} \right)^2 \\ 1 &= (a\,p + \pi_m - \tilde{w}_{a,m}) \frac{2 s_{a,m}\lambda_a^{m,u} f_{a,m}(\tilde{w}_{a,m})}{\delta_{a,m} + s_{a,m}\lambda_a^{m,u}(1 - F_{a,m}(\tilde{w}_{a,m}))} \\ 1 &= (a\,p + \pi_f - z - \tilde{w}_{a,f}) \frac{2 s_{a,f}\lambda_a^{f,u} f_{a,f}(\tilde{w}_{a,f})}{\delta_{a,f} + s_{a,f}\lambda_a^{f,u}(1 - F_{a,f}(\tilde{w}_{a,f}))} \end{aligned} \tag{30}$$

B.5 Limitations of JVPM: Variation in Female Shares

In the data, we observe that female shares are a) quite dispersed; the variance in female shares in the data is an order of magnitude larger than what the model generates and ranging from 0 to 1 (as opposed to 0.41 to 0.65 in the model), and b) the female share is weakly correlated to the gender wage gap. Why does the model fail in replicating these features of the data?

Regarding the variation in female shares, it can be proved analytically that the lowest and highest female shares that the model generates are functions of the parameters in the model. Using

equation 28 we can write the female share of a firm as

$$
\begin{aligned}
s_f =& \frac{l_{a,f}(\tilde{w}_{a,f}, v_a)}{l_{a,f}(\tilde{w}_{a,f}, v_a) + l_{a,m}(\tilde{w}_{a,m}, v_a)} \\
=& \frac{\left(\frac{1}{\delta_{a,f}+s_{a,f}\lambda_a^{f,u}(1-F_{a,f}(\tilde{w}_{a,f}))}\right)^2 \frac{v_a}{V_a} u_{a,f}\lambda_a^{f,u}(\delta_{a,f}+s_{a,f}\lambda_a^{f,u})}{\left(\frac{1}{\delta_{a,f}+s_{a,f}\lambda_a^{f,u}(1-F_{a,f}(\tilde{w}_{a,f}))}\right)^2 \frac{v_a}{V_a} u_{a,f}\lambda_a^{f,u}(\delta_{a,f}+s_{a,f}\lambda_a^{f,u}) + \left(\frac{1}{\delta_{a,m}+s_{a,m}\lambda_a^{m,u}(1-F_{a,m}(\tilde{w}_{a,m}))}\right)^2 \frac{v_a}{V_a} u_{a,m}\lambda_a^{m,u}(\delta_{a,m}+s_{a,m}\lambda_a^{m,u})} \\
=& \frac{1}{1+\frac{\left(\frac{1}{\delta_{a,m}+s_{a,m}\lambda_a^{m,u}(1-F_{a,m}(\tilde{w}_{a,m}))}\right)^2 u_{a,m}\lambda_a^{m,u}(\delta_{a,m}+s_{a,m}\lambda_a^{m,u})}{\left(\frac{1}{\delta_{a,f}+s_{a,f}\lambda_a^{f,u}(1-F_{a,f}(\tilde{w}_{a,f}))}\right)^2 u_{a,f}\lambda_a^{f,u}(\delta_{a,f}+s_{a,f}\lambda_a^{f,u})}}
\end{aligned}
$$

However, notice that in our data E-U transition rates and U-E transition rates are substantially identical between the two genders. That is, $\delta_{a,f} = \delta_{a,m} = \delta$ and $\lambda_a^{f,u} = \lambda_a^{m,u} = \lambda_a^u$. Therefore unemployment rates are also identical across genders, so that the expression for the female share simplifies to

$$
s_f = \frac{1}{1+\frac{\left(\frac{1}{\delta+s_{a,m}\lambda_a^{m,u}(1-F_{a,m}(\tilde{w}_{a,m}))}\right)^2 (\delta+s_{a,m}\lambda_a^u)}{\left(\frac{1}{\delta+s_{a,f}\lambda_a^u(1-F_{a,f}(\tilde{w}_{a,f}))}\right)^2 (\delta+s_{a,f}\lambda_a^u)}}
$$

Since firm sizes are monotonically increasing in the effective wage $\tilde{w}$ offered by the firm, we can obtain the minimum female share $\underline{s}_f$ by considering a firm that is at the bottom of the ladder for women and at the top of the ladder for men. That is, $F_{a,f} = 0$ and $F_{a,m} = 1$.[12] Then we can

[12] This firm might not actually exist in our simulations as male and female draws are correlated, thereby making it extremely unlikely that such a firm is drawn. Then, we will have that our simulations produce a minimum female share that is slightly larger than this theoretical lower bound.

write

$$\begin{aligned}
\underline{s}_f =& \frac{1}{1+\frac{\left(\frac{1}{\delta}\right)^2(\delta+s_{a,m}\lambda_a^u)}{\left(\frac{1}{\delta+s_{a,f}\lambda_a^u}\right)^2(\delta+s_{a,f}\lambda_a^u)}} \\
=& \frac{1}{1+\frac{\frac{1}{\delta^2}(\delta+s_{a,m}\lambda_a^u)}{\frac{1}{\delta+s_{a,f}\lambda_a^u}}} \\
=& \frac{1}{1+\frac{(\delta+s_{a,m}\lambda_a^u)(\delta+s_{a,f}\lambda_a^u)}{\delta^2}} \\
=& \frac{1}{2+\frac{s_{a,f}\lambda_a^u}{\delta}+\frac{s_{a,m}\lambda_a^u}{\delta}+\frac{s_{a,m}s_{a,f}(\lambda_a^u)^2}{\delta^2}}
\end{aligned}$$

In our data we find roughly that $s_{a,m}\lambda_a^u \simeq 0.4\,\delta$ and $s_{a,f}\lambda_a^u \simeq 0.27\,\delta$. If we apply these numbers, we find the minimum possible female share predicted by our model:

$$\underline{s}_f = \frac{1}{2+0.27+0.4+0.108} = 0.36$$

which is clearly inconsistent with our data. Similarly, we can write the maximum female share $\bar{s}_f$ as

$$\begin{aligned}
\bar{s}_f =& \frac{1}{1+\frac{\left(\frac{1}{\delta+s_{a,m}\lambda_a^u}\right)^2(\delta+s_{a,m}\lambda_a^u)}{\left(\frac{1}{\delta}\right)^2(\delta+s_{a,f}\lambda_a^u)}} \\
=& \frac{1}{1+\frac{\frac{1}{\delta+s_{a,m}\lambda_a^u}}{\frac{(\delta+s_{a,f}\lambda_a^u)}{\delta^2}}} \\
=& \frac{1}{1+\frac{\delta^2}{(\delta+s_{a,m}\lambda_a^u)(\delta+s_{a,f}\lambda_a^u)}} \\
=& \frac{1}{1+\frac{\delta^2}{\delta^2+\delta s_{a,m}\lambda_a^u+\delta s_{a,f}\lambda_a^u+s_{a,m}s_{a,f}(\lambda_a^u)^2}} \\
\simeq& \frac{1}{1+\frac{\delta^2}{\delta^2+0.4\,\delta^2+0.27\,\delta^2+0.108\delta^2}} \simeq 0.64
\end{aligned}$$

which is again very far from the maximum female share of 0.99 found in the data.

B.6 Developing the Solution Algorithm for the JVPM

The joint vacancy posting model is challenging because it does not allow us to solve men and women as two separate differential equations. Thus, we rely on a different algorithm to solve the joint vacancy posting model. Define the composite productivity, amenity and discrimination for each gender (firm-specific) as:

$$\begin{aligned} p_{a,m} =& ap + \pi_m \\ p_{a,f} =& ap + \pi_f - z \end{aligned} \tag{31}$$

Assume $c(v_a) = c\frac{v_a^2}{2}$. Then,

$$\begin{aligned} v_a = &\frac{T_{a,m}}{c}(p_{a,m} - \tilde{w}_{a,m})\left(\frac{1}{\delta_{a,m} + s_{a,m}\lambda_a^{m,u}(1 - F_{a,m}(\tilde{w}_{a,m}))}\right)^2 \\ &+ \frac{T_{a,f}}{c}(p_{a,f} - \tilde{w}_{a,f})\left(\frac{1}{\delta_{a,f} + s_{a,f}\lambda_a^{f,u}(1 - F_{a,f}(\tilde{w}_{a,f}))}\right)^2 \end{aligned} \tag{32}$$

By definition:

$$\begin{aligned} V_{a,m} &= \int\int v_a \, \mathbb{I}[p_{a,m} > \phi_{a,m}] \, \gamma(p_{a,m}, p_{a,f}) \, dp_{a,m} \, dp_{a,f} \\ V_{a,f} &= \int\int v_a \, \mathbb{I}[p_{a,f} > \phi_{a,f}] \, \gamma(p_{a,m}, p_{a,f}) \, dp_{a,m} \, dp_{a,f} \\ V_a &= \int\int v_a \, \mathbb{I}[p_{a,m} > \phi_{a,m} \;\text{ OR }\; p_{a,f} > \phi_{a,f}] \, \gamma(p_{a,m}, p_{a,f}) \, dp_{a,m} \, dp_{a,f} \end{aligned} \tag{33}$$

Start by defining $\gamma(p_{a,m})$ as the marginal distribution of $p_{a,m}$. Also, define $\bar{v}_a(p_{a,m}) = \int v(p_{a,m}, p_{a,f})\gamma(p_{a,m}, p_{a,f}$ as the average vacancies posted by firms with male productivity $p_{a,m}$. Then we can write

$$\begin{aligned} h(p) &= F(\tilde{w}(p)) \\ h'(p) &= f(w(p)) \, w'(p), \text{therefore} \\ f(w(p)) &= h'(p)/w'(p) \\ \bar{v}_a(p_{a,m}) &= \frac{V_{a,m}h'(p_{a,m})}{\gamma(p_{a,m})}, \text{therefore} \\ h'(p_{a,m}) &= \frac{\bar{v}_a(p_{a,m})}{V_{a,m}}\gamma(p_{a,m}) \end{aligned} \tag{34}$$

Thus, the wage FOC can be rewritten as follows:

$$1 = (p_{a,g} - \tilde{w}_{a,m}) \frac{2 s_{a,g} \lambda^u_{a,g} f_{a,g}(\tilde{w}_{a,g})}{\delta_{a,g} + s_{a,g} \lambda^u_{a,g} (1 - F_{a,g}(\tilde{w}_{a,g}))}$$

$$\tilde{w}'(p_{a,g}) = (p_{a,g} - \tilde{w}_{a,g}) \frac{2 s_{a,g} \lambda^u_{a,g} h'(p_{a,g})}{\delta_{a,g} + s_{a,g} \lambda^u_{a,g} (1 - h(p_{a,g}))}$$

We start from the case in which $\gamma(p_{a,m}, p_{a,f})$ is an analytical function and all marginal and conditional distributions associated are easy to compute (therefore also the marginals $\gamma(p_{a,g})$ are known). Using the previous intuitions, the algorithm works as follows:

1. Start with a guess for $\tilde{w}(p_{a,g})$, $F(\tilde{w}(p_{a,g})$ and $V_{a,g}$ for each gender.
2. Calculate transition rates $\lambda^u_{a,g}$, $\lambda^e_{a,g}$ using the guess for $V_{a,g}$.
3. Using the guess for the wage function, compute $v_a(p_{a,m}, p_{a,f})$ on a large grid over values of $p_{a,m}$ and $p_{a,f}$.
4. Normalize v_a to get $V_{a,g}$. Use the function v_a to compute $h'(p_{a,g})$ for both genders as in 34.
5. Use $h'(p_{a,g})$ to compute the new function $h(p_{a,g}) = F(\tilde{w}(p_{a,g})$.
6. Use h' and h to solve the ODE in B.6.
7. Update the wage function $\tilde{w}(p_{a,g})$ and the CDF $F(\tilde{w}(p_{a,g})$.
8. Go back to step 2. Repeat until convergence.

B.7 Identifying Productivity and Amenities: Monte Carlo Simulations

We solve our model and simulate firm-level data on wages, amenities, ranks and vacancies. We use this data to construct our estimates of rank r, density f_r and CDF F_r, and use them to estimate amenities and productivity at the firm level to test whether our algorithm is successful at uncovering the true firm-specific parameters. Our results are summarized in Table 15.

Table 15. Monte Carlo Simulations: estimation results using simulated data, under different parametrizations of the initial underlying amenities distribution.

Properties of true π					
Variance(π)	0.632	0.904	1.061	1.092	0.942
Corr(π,w)	-0.840	-0.849	-0.850	-0.932	-0.659
Corr(π,rank)	0.386	0.503	0.556	0.404	0.830
Corr(w,rank)	0.074	-0.083	-0.164	-0.126	-0.409
Properties of true p					
Variance(p)	0.904	0.956	0.984	1.015	0.762
Corr(p,w)	0.823	0.767	0.719	0.874	0.148
Corr(p,rank)	0.623	0.571	0.561	0.367	0.832
Corr(p,π)	-0.422	-0.355	-0.295	-0.664	0.553
Estimation Results:					
Variance($\hat{\pi}$)	0.577	0.721	0.741	0.892	0.274
Corr($\hat{\pi}$,w)	-0.905	-0.950	-0.980	-1.000	-0.985
Corr($\hat{\pi}$,rank)	0.304	0.357	0.336	0.150	0.544
Variance($\hat{p}$)	0.602	0.638	0.662	0.881	0.221
Corr($\hat{p}$,w)	0.951	0.978	0.992	1.000	0.994
Corr($\hat{p}$,rank)	0.353	0.054	-0.079	-0.102	-0.353
Corr($\hat{p}$,$\hat{\pi}$)	-0.777	-0.906	-0.965	-0.999	-0.976
Goodness of fit					
Corr($\hat{\pi}$, π)	0.989	0.970	0.936	0.942	0.777
Mean Error	-0.679	-0.347	-0.193	-0.608	0.546
Mean Squared Error	0.305	0.168	0.181	0.528	0.707
Corr($\hat{p}$, p)	0.940	0.833	0.766	0.885	0.199
Mean Error	0.103	0.301	0.636	0.718	1.427
Mean Squared Error	0.129	0.380	0.807	0.745	2.372