The Gender Gap:
Micro Sources and Macro Consequences*

Iacopo Morchio† Christian Moser‡

February 15, 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, Ben Moll, Andreas Mueller, Emi Nakamura, Jesse Schreger, Jón Steinsson, Pierre Yared, Steve Zeldes, 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

¹See 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 quantitative 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. Earnings for men is on average 15 log points.

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Pay gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Monthly earnings</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>7.205</td>
<td>7.053</td>
<td>0.152</td>
</tr>
<tr>
<td>P5</td>
<td>6.251</td>
<td>6.221</td>
<td>0.030</td>
</tr>
<tr>
<td>P10</td>
<td>6.349</td>
<td>6.289</td>
<td>0.060</td>
</tr>
<tr>
<td>P25</td>
<td>6.640</td>
<td>6.512</td>
<td>0.128</td>
</tr>
<tr>
<td>P50</td>
<td>7.035</td>
<td>6.863</td>
<td>0.172</td>
</tr>
<tr>
<td>P75</td>
<td>7.605</td>
<td>7.430</td>
<td>0.175</td>
</tr>
<tr>
<td>P90</td>
<td>8.286</td>
<td>8.137</td>
<td>0.149</td>
</tr>
<tr>
<td>P95</td>
<td>8.786</td>
<td>8.565</td>
<td>0.221</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Pay gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel B. Hourly wage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3.469</td>
<td>3.391</td>
<td>0.078</td>
</tr>
</tbody>
</table>

*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

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

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 individual $i$ in year $t$, denoted $y_{it}$, is as follows:

$$y_{it} = \alpha_{base} - 1 [female_i] \alpha_{gap} + X_{it} \beta + \varepsilon_{it}$$  \hspace{1cm} (1)

where $\alpha_{base}$ is the male pay intercept, $1 [female_i]$ is an indicator that equals 1 if individual $i$ is female and 0 otherwise, $\alpha_{gap}$ is the conditional female pay gap, $X_{it}$ contains other observable worker and characteristics, and $\varepsilon_{it}$ is a residual term. We estimate equation (1) via ordinary least squares (OLS) by imposing the usual strict exogeneity assumption on $\varepsilon_{it}$, namely $E[\varepsilon_{it} | female_i, X_{it}] = 0$. 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:

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

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 $\pm 0.288$. As another means of comparison, we find that the same index is significantly 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 $y_{jt}$:

$$y_{jt} = \text{female employment share}_{jt} \times \gamma + X_{jt} \beta + \varepsilon_{jt} \tag{2}$$

where 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 experience. 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.”
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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women’s employment share</td>
<td>-0.201</td>
<td>-0.534</td>
<td>-0.533</td>
</tr>
<tr>
<td>Industry, occupation, state FEs</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Education × experience FEs</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td># Observations (millions)</td>
<td>35.9</td>
<td>35.3</td>
<td>35.3</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.01</td>
<td>0.37</td>
<td>0.44</td>
</tr>
<tr>
<td>Composition effect</td>
<td>-0.152</td>
<td>-0.294</td>
<td>-0.300</td>
</tr>
<tr>
<td>Net sorting effect</td>
<td>-0.049</td>
<td>-0.240</td>
<td>-0.233</td>
</tr>
</tbody>
</table>

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. 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

3In 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 $y_{ijt}$:

$$y_{ijt} = X_{it} \beta + \alpha_j^{base} - 1 \{female\} \alpha_j^{gap} + \epsilon_{ijt}$$

(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, $\alpha_j^{base}$ is the male firm fixed effect and $\alpha_j^{gap}$ is the intra-firm pay gap, and $\epsilon_{ijt}$ is a residual term satisfying the usual strict exogeneity condition: $E[\epsilon_{ijt}|i,j,t] = 0$. We estimate equation (3) via ordinary least squares (OLS) and focus on estimates of the gender-specific firm fixed effects, $\alpha_j^{base}$ and $\alpha_j^{gap}$. Based on the estimation results, we can decompose the overall gender gap as follows:

$$\alpha^{gap} = E[\alpha_j^{gap} | female] + \left( E[\alpha_j^{base} | male] - E[\alpha_j^{base} | female] \right)$$

where total gender pay differences on the left hand side, $\alpha^{gap}$, 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.\footnote{We obtain similar results in an alternative decomposition where we take expectiations in equation (3) conditional on the distribution of men.}

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
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

![Figure 4](image)

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 $\eta_{jt}$ 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

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

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

![Figure 6](image-url)

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:

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

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
men and 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

As the previous graph showed merely raw wages, one may expect that some of the 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:

$$y_{it} = \sum_{e=0}^{50} [\exp_{it} = e] \left( \alpha_{e}^{base} - 1 \cdot [female] \cdot \alpha_{e}^{gap} \right) + X_{it} \beta + \epsilon_{it}$$

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

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:

\[ y_{it} = \sum_{e=0}^{50} 1[\text{exp}_{it} = e] \left( a^\text{base}_e - 1[female_i] a^\text{gap}_e \right) + X_{it} \beta + a^\text{base}_j - 1[female_i] a^\text{gap}_j + \epsilon_{it} \]

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.

---

5See also Erosa et al. (2016).
Figure 9. The role of firm heterogeneity in explaining life cycle pattern of the gender pay gap

(a) Female earnings gap with and without firm controls

(b) Level and share of female earnings gap due to firms

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 cycle, 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

![Distribution over gender-firm average wage rank](image1)

![Distribution over gender-firm fixed effect rank](image2)

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, we 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 mobility 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

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob. of UE transition</td>
<td>0.165</td>
<td>0.157</td>
</tr>
<tr>
<td>Prob. of EU transition</td>
<td>0.182</td>
<td>0.172</td>
</tr>
<tr>
<td>Prob. of EE transition</td>
<td>0.016</td>
<td>0.012</td>
</tr>
</tbody>
</table>

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 (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

![Graph showing the relationship between male and female firm fixed effects.]

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.
What determines 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:

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

where $Z_{jt}$ is again a vector containing the observable characteristics of firm $j$ in year $t$, and $\xi_{jt}$ 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 estimates are not purely mechanical.

Table 7. Regression of female pay gap on its determinants

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

Notes: 5% sample of 2001–2005 Brazilian formal sector census data (RAIS). Dependent variable is within-firm difference in gender-firm FE. 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

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Difference (male - female)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine Manual</td>
<td>0.368</td>
<td>-0.023</td>
<td>0.391</td>
</tr>
<tr>
<td>Routine Cognitive</td>
<td>-0.254</td>
<td>-0.056</td>
<td>-0.198</td>
</tr>
<tr>
<td>Non-routine Manual Physical</td>
<td>0.446</td>
<td>-0.136</td>
<td>0.582</td>
</tr>
<tr>
<td>Non-routine Manual Personal</td>
<td>-0.324</td>
<td>0.269</td>
<td>-0.593</td>
</tr>
<tr>
<td>Non-routine Cogn. Personal</td>
<td>-0.230</td>
<td>0.102</td>
<td>-0.332</td>
</tr>
<tr>
<td>Non-routine Cogn. Analytical</td>
<td>-0.130</td>
<td>0.210</td>
<td>-0.340</td>
</tr>
</tbody>
</table>

# 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:

\[ y_{ijt} = X_{ijt}\beta + \alpha_j^{gender} + \text{taskskill}_{ijt} \times \gamma^{gender} + \epsilon_{ijt} \]

where \( \text{taskskill}_{ijt} \) is the occupation-specific measure derived from the Dictionary of Occupational Titles, and \( \gamma^{gender} \) is a coefficient that we allow to differ across genders. We already described differences in the allocation of women compared to men across task and skill groups, \( \text{taskskill}_{ijt} \), 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 introducing 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

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Routine Manual</td>
<td>-0.078</td>
<td>-0.059</td>
</tr>
<tr>
<td>Routine Cognitive</td>
<td>0.141</td>
<td>0.077</td>
</tr>
<tr>
<td>Non-routine Manual Physical</td>
<td>0.014</td>
<td>0.056</td>
</tr>
<tr>
<td>Non-routine Manual Personal</td>
<td>-0.093</td>
<td>-0.047</td>
</tr>
<tr>
<td>Non-routine Cogn. Personal</td>
<td>-0.083</td>
<td>-0.022</td>
</tr>
<tr>
<td>Non-routine Cogn. Analytical</td>
<td>0.323</td>
<td>0.143</td>
</tr>
<tr>
<td>Edu×exp FEs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>State, industry FEs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Firm FEs</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td># Observations (millions)</td>
<td>21.0</td>
<td>21.0</td>
</tr>
</tbody>
</table>

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:

\[ y_{ijt} = X_{it} \beta + tenure_{ijt} \times \gamma_{gender} + \alpha_{i}^{gender} + \alpha_{t} + \varepsilon_{ijt} \]

where \( tenure_{ijt} \) is tenure of individual \( i \) at firm \( j \) in year \( t \), and \( \gamma_{gender} \) 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

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

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 $a \in [a,\bar{a}]$, 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. Finally, workers are infinitely lived and care for the value of the stream of expected consumption, discounted at rate $\rho$.

A job offer is an opportunity to work at firm $j$ for the wage $w$ and a level of firm-specific
amenity $\pi_j$, 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:

$$\rho S_{a,g}(w + \pi) = w + \pi + \lambda e_{a,g} \int_{w+\pi}^{w+\pi} w' + \pi' S_{a,g}(w' + \pi') - S_{a,g}(w + \pi) dF_{a,g}(w' + \pi')$$

$$+ \delta_{a,g}[W_{a,g} - S_{a,g}(w + \pi)]$$

The problem of the unemployed worker reads

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

(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):

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

(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

$$u_{a,g} = \frac{\delta_{a,g}}{\delta_{a,g} + \lambda e_{a,g}}$$

(7)

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

$$G^\delta(w + \pi) = \frac{F^\delta(w + \pi)}{1 + \kappa_{a,g}(1 - F^\delta(w + \pi))}$$

(8)

where $\kappa_{a,g} = \lambda e_{a,g}/\delta_{a,g}$ 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 a discrimination level 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 gender. 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. In the data, instead, we observe substantial variation in female shares, which are most often than not different from zero and one. The second 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.

---

6 We discuss in detail the proof of this argument in Appendix B.1.
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.
Under the assumption of separate vacancy posting, the firm’s problem reads:

\[
\max_{w_{a,m}, w_{a,f}, v_{a,m}, v_{a,f}} \left\{ \left( p_a - w_{a,m} \right) l_{a,m} \left( w_{a,m} + \pi_{m}, v_{a,m} \right) \right.
\]
\[
\left. + \left( p_a - z - w_{a,f} \right) l_{a,f} \left( w_{a,f} + \pi_{f}, v_{a,f} \right) - c_{a,m}(v_{a,m}) - c_{a,f}(v_{a,f}) \right\}
\]

(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

\[
\max_{\tilde{w}_{a,m}, \tilde{w}_{a,f}, v_{a,m}, v_{a,f}} \left\{ \left( p_a + \pi_{m} - \tilde{w}_{a,m} \right) l_{a,m} \left( \tilde{w}_{a,m}, v_{a,m} \right) \right.
\]
\[
\left. + \left( p_a + \pi_{f} - z - \tilde{w}_{a,f} \right) l_{a,f} \left( \tilde{w}_{a,f}, v_{a,f} \right) - c_{a,m}(v_{a,m}) - c_{a,f}(v_{a,f}) \right\}
\]

(11)

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

A firm makes profits if and only if

\[
\phi_{a,g} < \tilde{w} < p_a + \pi_g - z \mathbb{I}[g = \text{female}]
\]

(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:

\[
m_{a,g} = A^\delta [\mu_{a,g}(u_{a,g} + \lambda_{a,g}(1 - u_{a,g})]^a V^{1-a}
\]

(13)

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

\[
V_{a,g} = \int v_{a,g}(p, \pi, z) dD_{a,g}(p, \pi, z)
\]

(14)

4.3 Equilibrium

Define \( \lambda_{a,g} = s_{a,g} \lambda_{a,g}^u \). The following equation represents the law of motion of firm sizes:
\[ \dot{a} = -\delta_a \lambda_{a}^u (1 - F_{a,m}(\bar{w})) \lambda_{a}^c (1 - F_{a,a}(\bar{w})) - s_{a}^d \lambda_{a}^c (1 - F_{a,a}(\bar{w})) \]

\[ \dot{a} = \frac{u_{a,g} \left[ \frac{u_{a,g}}{u_{a,g} + (1 - u_{a,g}) s_{a,g}} + \frac{(1 - u_{a,g}) s_{a,g}}{u_{a,g} + (1 - u_{a,g}) s_{a,g}} \right] G_{a,g}(\bar{w}) }{v_{a,g}} \]  

Solving for the stationary solution:

\[ l_{a,g}(\bar{w}, v_a) = \left( \frac{1}{\delta_{a,g} + s_{a,g} \lambda_{a,g}^u (1 - F_{a,g}(\bar{w}))} \right)^2 \frac{v_{a,g} u_{a,g} \lambda_{a,g}^u (\delta_{a,g} + s_{a,g} \lambda_{a,g}^u)}{V_{a,g}} \]

\[ l_{a,g}(\bar{w}_{a,g}, v_a) = \left( \frac{1}{\delta_{a,g} + s_{a,g} \lambda_{a,g}^u (1 - F_{a,g}(\bar{w}_{a,g}))} \right)^2 \frac{v_{a,g} u_{a,g} \lambda_{a,g}^u (\delta_{a,g} + s_{a,g} \lambda_{a,g}^u)}{V_{a,g}} \]

### 4.4 Equilibrium Characterization

To find the firm policy functions, define \( T_{a,g} = \mu_{a,g} \left[ \mu_{a,g} \lambda_{a,g}^u (\delta_{a,g} + s_{a,g} \lambda_{a,g}^u) \right] / V_{a,g} \). We rewrite the firm’s problem as a function of the steady state mass of employed workers as follows:

\[ \max_{\bar{w}_{a,m}, \bar{w}_{a,f}, \bar{w}_{a,f}} \left\{ T_{a,m} v_{a,m} (a p + \pi_m - \bar{w}_{a,m}) \left( \frac{1}{\delta_{a,m} + s_{a,m} \lambda_{a,m}^u (1 - F_{a,m}(\bar{w}_{a,m}))} \right)^2 + T_{a,f} v_{a,f} (a p + \pi_f - z - \bar{w}_{a,f}) \left( \frac{1}{\delta_{a,f} + s_{a,f} \lambda_{a,f}^u (1 - F_{a,f}(\bar{w}_{a,f}))} \right)^2 - c_{a,m}(v_{a,m}) - c_{a,f}(v_{a,f}) \right\} \]

The associated FOC read:

\[ c_{m}^{m}(v_{a,m}) = T_{a,m} (a p + \pi_m - \bar{w}_{a,m}) \left( \frac{1}{\delta_{a,m} + s_{a,m} \lambda_{a,m}^u (1 - F_{a,m}(\bar{w}_{a,m}))} \right)^2 \]

\[ c_{f}^{f}(v_{a,f}) = T_{a,f} (a p + \pi_f - z - \bar{w}_{a,f}) \left( \frac{1}{\delta_{a,f} + s_{a,f} \lambda_{a,f}^u (1 - F_{a,f}(\bar{w}_{a,f}))} \right)^2 \]

\[ 1 = (a p + \pi_m - \bar{w}_{a,m}) \frac{2s_{a,m} \lambda_{a,m}^u f_{a,m}(\bar{w}_{a,m})}{\delta_{a,m} + s_{a,m} \lambda_{a,m}^u (1 - F_{a,m}(\bar{w}_{a,m}))} \]

\[ 1 = (a p + \pi_f - z - \bar{w}_{a,f}) \frac{2s_{a,f} \lambda_{a,f}^u f_{a,f}(\bar{w}_{a,f})}{\delta_{a,f} + s_{a,f} \lambda_{a,f}^u (1 - F_{a,f}(\bar{w}_{a,f}))} \]

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

The model parameters are identified by making the model replicate data features that are closely related to our parameters. In this section we discuss in detail the calibration strategy. Consistent with our empirical analysis that controls for worker-level heterogeneity and to focus on firm-level determinants of gender gaps, we study employer level heterogeneity within a single ability type \( a = 1 \) for both genders. Also, we set \( \mu_f = 0.386 \), which is equal to the female share of the labor force in RAIS data.

We first fix the functional forms of the joint and marginal distributions of \( \{p, \pi_m, \pi_f, z\} \). First, we assume that the joint distribution is normal, with covariance matrix \( \Sigma \). We assume \( p \) is distributed as a Pareto with scale parameter \( \mu_p \) and tail parameter \( \sigma_p \). Furthermore, we assume that the marginal distribution of \( \pi_m, \pi_f \) and \( z \) is normal, so that for each firm-specific characteristic we must identify a mean parameter and a dispersion parameter.

Regarding the covariance matrix \( \Sigma \), in principle we have to identify:

\[
\begin{bmatrix}
1 & \rho_{p,\pi_m} & \rho_{p,\pi_f} & \rho_{p,z} \\
\rho_{p,\pi_m} & 1 & \rho_{\pi_m,\pi_f} & \rho_{\pi_m,z} \\
\rho_{p,\pi_f} & \rho_{\pi_m,\pi_f} & 1 & \rho_{\pi_f,z} \\
\rho_{p,z} & \rho_{\pi_m,z} & \rho_{\pi_f,z} & 1
\end{bmatrix}
\] (19)

To simplify the analysis we focus only on three unique correlations: \( \rho_{p,\pi_m} \), the correlation between productivity and male amenities; \( \rho_{\pi_m,\pi_f} \), the correlation between male and female amenities; and \( \rho_{p,z} \), the correlation between productivity and the negative taste for women. We fix the other correlations to be equal to the cross-product of these three unique correlations.

One consequence of our distributional assumptions is that it is possible that, for some jobs, the monetary wage \( \tilde{w} - \pi_g \) is negative. This happens because the job ladder is in terms of effective wages \( \tilde{w} = w + \pi \). Thus, if a firm is endowed with low productivity but very high amenities, it is possible that the equilibrium monetary wage \( w = \tilde{w} - \pi \) will be negative. We throw out these observations in our simulations, and we interpret these as jobs that we would not observe in our data anyway, in the sense that they are held by individuals who work for free because the job has a high amenity value for them. An example of this would be a worker volunteering in an NGO,
which is not part of our data on formal sector workers.\footnote{Nevertheless, fewer than 1 out of 100,000 firms in our simulations exhibit negative wages.}

Without loss of generality, we normalize the scale parameter $\mu_p = 1$. As is standard in the literature, we target the tail parameter $\sigma_p$ to match the variance of firm fixed effects in the RAIS data.

We find that the parameters $\mu_{\pi_m}$ and $\mu_{\pi_f}$ cannot be separately identified from the outside options $b_m$ and $b_f$. The reason is that the model produces identical wages, firm sizes and transition rates if we increase $\mu_{\pi_g}$ and $b_g$ by the same amount, because when all firms can offer higher amenities and the outside option is higher, the two cancel each other out, as is easy to see from equations 4 and 5. Therefore, we fix the means of amenities for both genders to 0 and, as is standard, we calibrate the outside options $b_m$ and $b_f$ to match the median-to-minimum wage ratio in the data.

In general, our identification argument to separate the role of amenities and discrimination relies on the relationship between wages and firm sizes. Intuitively, for a given level of productivity, both higher amenities and higher discrimination imply a lower wage. However, the two fundamentals have exactly opposite effects on firm sizes: while higher amenities lead to a larger firm size, because the firm can place itself higher up in the ladder, higher discrimination implies lower female firm size, because discrimination acts as a dampener of the firm’s utility. Therefore, given productivity, we can distinguish a high-amenities firm from a high-discrimination firm by looking at firm size. Table 11 sketches the idea behind the identification argument.

<table>
<thead>
<tr>
<th>low discrimination</th>
<th>high discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>low amenities</td>
<td>high pay, medium firm size</td>
</tr>
<tr>
<td>high amenities</td>
<td>medium pay, small firm size</td>
</tr>
<tr>
<td></td>
<td>medium pay, large firm size</td>
</tr>
<tr>
<td></td>
<td>low pay, medium firm size</td>
</tr>
</tbody>
</table>

An increase in $\mu_z$ naturally implies a larger gender wage gap, because it makes all firms have a stronger distaste for women. Therefore, we use $\mu_z$ to match the gender wage gap in firm fixed effects observed in the data. Since the distribution of $z$ enters all women’s wage determination, we use $\sigma_z$ to match the variance of women’s wages in the data and in the model.

A higher variance in amenities implies that workers are more likely to experience monetary
wage losses upon switching jobs (as in Sorkin 2018). This is because, when the variance in amenities is higher relative to that of productivity, it becomes more likely that a worker is attracted to a different job not because of a wage increase, but because of better amenities. Therefore, we use $\sigma_{\pi_m}$ and $\sigma_{\pi_f}$ to match the fraction of job-to-job transitions that are accompanied by a monetary wage decrease.

$\rho_{p,\pi_m}$, the correlation between productivity and male amenities, is strictly tied to the firm size wage premium. Intuitively, when productivity is the only determinant of a firm’s placement in the ladder, wages and firm sizes are strictly related, therefore the firm size wage premium is high. When amenities are accounted for, this is not necessarily true anymore, as firms can be large because they are highly productive or because they have high amenities, thus lowering the firm size wage premium. In this case, a stronger correlation between productivity and amenities determines a lower firm size wage premium, because as productivity increases amenities also increase in the distribution, having a negative effect on the wage and thus reducing the premium associated with larger firms.

$\rho_{\pi_m,\pi_f}$, the correlation between male and female amenities, determines how strongly correlated male and female firm sizes are within firms. Suppose for simplicity that there is no variation in $z$: if $\rho_{\pi_m,\pi_f} = 1$, then male and female sizes are strongly related. The lower the correlation is, the less male and female sizes are related. Therefore, we use this correlation to target the correlation between male size ranks and female size ranks observed in the data.

$\rho_{p,z}$, the correlation between productivity and distaste for women, is identified by matching the correlation between the gender gap and the male wage rank in the data. Intuitively, if more productive firms dislike women more, then gender gaps will be higher for more productive firms, thereby increasing the gender gap at those firms.

Turning to transition rates, we can feed $\lambda^u_{m}$, $\lambda^u_{f}$, $\delta^u_{m}$ and $\delta^u_{f}$ directly into our model. However, we must endogenously calibrate $s^u_{m}$ and $s^u_{f}$, the on-the-job search effort levels. In models of this type, it is common to invert the wage distribution in order to directly feed these parameters to the model. However, this is possible only if productivity is the only determinant of wages, as in that case the wage distribution and the offer distribution are one and the same thing. In our case, the offer distribution is extremely different from the monetary wage distribution because it includes the value of amenities. Therefore, we calibrate $s^e$ for each gender by matching the model-generated job-to-job transition rate, calculated as follows:
Table 12. Baseline Calibration Table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Moment</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distributions, Marginals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_p )</td>
<td>1.0000</td>
<td>Normalization</td>
<td>0.1600</td>
<td>0.1500</td>
</tr>
<tr>
<td>( \sigma_p )</td>
<td>24.7622</td>
<td>Var of wages, M</td>
<td>0.1570</td>
<td>0.1547</td>
</tr>
<tr>
<td>( \mu_{\pi_m} )</td>
<td>0.0100</td>
<td>Normalization</td>
<td>0.4000</td>
<td>0.4170</td>
</tr>
<tr>
<td>( \sigma_{\pi_m} )</td>
<td>1.3653</td>
<td>Frac of J2J with ( w ) cut, M</td>
<td>0.4000</td>
<td>0.4099</td>
</tr>
<tr>
<td>( \mu_{\pi_f} )</td>
<td>0.0100</td>
<td>Normalization</td>
<td>0.4000</td>
<td>0.4099</td>
</tr>
<tr>
<td>( \sigma_{\pi_f} )</td>
<td>0.9709</td>
<td>Frac of J2J with ( w ) cut, F</td>
<td>0.4000</td>
<td>0.4099</td>
</tr>
<tr>
<td>( \mu_{z} )</td>
<td>0.1085</td>
<td>Gender Wage Gap</td>
<td>0.1570</td>
<td>0.1547</td>
</tr>
<tr>
<td>( \sigma_{z} )</td>
<td>0.9236</td>
<td>Var of wages, F</td>
<td>0.1600</td>
<td>0.1608</td>
</tr>
<tr>
<td><strong>Distributions, Correlations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \rho(p, \pi_m) )</td>
<td>0.9176</td>
<td>Beta(w, size), M</td>
<td>0.0800</td>
<td>0.0853</td>
</tr>
<tr>
<td>( \rho(\pi_m, \pi_f) )</td>
<td>0.9352</td>
<td>Rank Corr(size M, size F)</td>
<td>0.8500</td>
<td>0.9586</td>
</tr>
<tr>
<td>( \rho(p, z) )</td>
<td>0.2243</td>
<td>Beta(gender gap, rank w M)</td>
<td>0.2850</td>
<td>0.2793</td>
</tr>
<tr>
<td><strong>Transition Rates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta_m )</td>
<td>0.0300</td>
<td>Exogenous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta_f )</td>
<td>0.0300</td>
<td>Exogenous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda^u_m )</td>
<td>0.1600</td>
<td>Exogenous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda^u_f )</td>
<td>0.1600</td>
<td>Exogenous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( s^m )</td>
<td>0.1629</td>
<td>J2J rate, M</td>
<td>0.0120</td>
<td>0.0103</td>
</tr>
<tr>
<td>( s_f )</td>
<td>0.1072</td>
<td>J2J rate, F</td>
<td>0.0080</td>
<td>0.0073</td>
</tr>
<tr>
<td><strong>Outside Options</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b_m )</td>
<td>0.8426</td>
<td>log(Median/Min wage), M</td>
<td>1.3000</td>
<td>1.3783</td>
</tr>
<tr>
<td>( b_f )</td>
<td>0.0524</td>
<td>log(Median/Min wage), F</td>
<td>1.3000</td>
<td>1.2996</td>
</tr>
</tbody>
</table>

\[
J2J = \lambda^e \int (1 - F(\tilde{w}_g)) g(\tilde{w}_g) d\tilde{w}_g
\]

where \( g(\tilde{w}_g) \) is the model-generated density function of the effective wage distribution.

As can be seen from Table 12, the model has no trouble getting very close to the data moments.

6 Results

In this section we first interpret the calibration results in terms of their implications for what the distributions of productivity, amenities and discrimination look like. Next, we explore the quantitative performance of our model. In particular, we try to understand if the model is qualitatively and quantitatively consistent with some noncalibrated facts on the relationship between gender gaps and firm ranks in the wage ladder, as well as firm size premia and the relationship between
gender gaps and firm sizes.

6.1 Calibration Results

The calibration implies that dispersion in productivity across firms is substantially larger than dispersion in amenities for either gender or discrimination. Also, we find that dispersion in male amenities is larger than dispersion in female amenities, and that discrimination is relatively less dispersed than amenities for either gender:

Figure 13. Model Results. Marginal distribution of $p$ (productivity), $\pi^m$ (amenities for men), $\pi^f$ (amenities for women) and $z$ (discrimination) under the baseline calibration.

Under the baseline calibration, productivity and amenities are strongly positively correlated: $\rho_{p,\pi^m} \simeq 0.92$, implying that high-productivity firms also tend to offer better amenities to their male employees. Also, amenities are strongly correlated across genders ($\rho_{\pi^m,\pi^f} \simeq 0.93$), implying that high-amenities firms for men tend also to be high-amenity firms for women. Finally, discrimination appears to be positively, but weakly, correlated to productivity ($\rho_{p,z} \simeq 0.2$).

6.2 Model Implications

The model can replicate rather well the distribution of firm fixed effects by gender. The inability to replicate the right tail of fixed effects distributions is a caveat that this model shares with other on-the-job random search models.

The model implies that the wage ladder is considerably steeper for men than for women. In particular, while the difference between the top and the bottom of the ladder for men is around
Figure 14. Model Results. Distribution of Male and Female Firm Fixed Effects under the baseline calibration.

Figure 15. Model Results. Average men’s firm fixed effects (red) and average women’s firm fixed effects (blue) by rank in the men’s ladder under the baseline calibration.
150 log points, the difference for women is only 100 log points. This replicates well the fact that, in the data, while men and women earn approximately the same at the bottom of the ladder, the firm-specific gender gap increases with male wage ranks, until the gender gap is approximately 50 log points at the top of the wage ladder.

Figure 16. Model Results. Average female wages against male wages, under the baseline calibration.

The male and female ladder are strongly related, with the important difference that women climb the ladder slower. However, while higher male rungs roughly correspond to higher female rungs, this is not an exact correspondence. Firms that have the same ranking for men can have different rankings for women because they have different amenities or tastes for women.

Figure 17. Model Results. Average female wage ranks against male ranks, under the baseline calibration.
The model can generate substantial variation in female shares, replicating around 60% of the variance of female shares in the data. Under the current calibration, the model also implies that the gender gap is unrelated to the female share, consistently with the data. This is because amenities and discrimination almost exactly offset each other as sources of gender gaps: in the model, firms that hire more women tend to either place themselves at the bottom of the ladder because they discriminate women, or at the top by offering high amenities. Thus, gender gaps end up being unrelated to the female share, because firms pay women less either because of amenities or because of discrimination, and the two are weakly related in the model.

### 6.3 Decomposing the Gender Pay Gap

In this section we simulate what happens in the model when we shut down differences between men and women. We do so to understand what is the quantitative relevance of the differences across genders in firm-level fundamentals in accounting for the gender wage gap and other differences. All models are solved again in general equilibrium in order to neutralize the effects that changing such fundamentals would have on other outcomes, such as arrival rates, making the comparison more complicated and less transparent. Results are summarized in table 13.

Table 13. 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), discrimination shut down (3), different rates of J2J transition shut down (4) and all combinations (5-8).

<table>
<thead>
<tr>
<th>Counterfactual Simulations</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amenities</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Discrimination</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>J2J arrival rate</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gender Wage Gap</td>
<td>0.15</td>
<td>0.71</td>
<td>0.06</td>
<td>-0.20</td>
<td>0.54</td>
<td>0.14</td>
<td>-0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>VAR((w^f))</td>
<td>0.16</td>
<td>0.62</td>
<td>0.11</td>
<td>0.12</td>
<td>0.41</td>
<td>0.28</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>(\beta(\text{GWG}, \text{rank } w^m))</td>
<td>0.28</td>
<td>-0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>-0.42</td>
<td>-0.18</td>
<td>0.41</td>
<td>0.00</td>
</tr>
<tr>
<td>(\beta(\text{GWG, female share}))</td>
<td>-0.01</td>
<td>-2.79</td>
<td>1.51</td>
<td>0.06</td>
<td>-2.57</td>
<td>-2.15</td>
<td>2.54</td>
<td>0.10</td>
</tr>
<tr>
<td>Std. Dev.(female share)</td>
<td>0.18</td>
<td>0.13</td>
<td>0.13</td>
<td>0.15</td>
<td>0.07</td>
<td>0.11</td>
<td>0.09</td>
<td>0.00</td>
</tr>
</tbody>
</table>

We find that discrimination is an important source of gender gaps in the model, accounting for approximately two-thirds of the gender gap, and for one-third of overall variation in women’s wages (column 3). However, the importance of discrimination is dwarfed by that of the lower arrival rate of offers during employment \(\lambda^e\) (column 4): in a world in which women received as
many offers as men during employment, the gender gap would actually be negative (i.e. women would be paid more than men), because women enjoy lower amenities under the baseline calibration and therefore would receive higher monetary wages overall if they could transit faster to higher-ranked jobs.

Turning to amenities, we find that, if women enjoyed the same level of amenities as men, their monetary wages would be lower by approximately 46% on average, increasing the gender gap more than four-fold. Interestingly, if women enjoyed the same amenities as men, the relationship between the gender gap at the firm level and the male wage rank would be negative rather than positive, a fact strongly at odds with our data, in which the gender gap is clearly increasing in the male wage rank. Also, if women and men enjoyed the same amenities, variation in female shares across firms would be generated only by variation in discrimination: as a result, more discriminatory firms would hire fewer women, and the gender gap would be decreasing in the female share, another fact at odds with the data. When amenities and the arrival rate of offers are equalized, the composition of the increase in the gender gap induced by amenities and the decrease in the gap induced by the increase in the arrival rate imply that, overall, the gender gap would be lowered by approximately 2 percentage points.

In general our simulation results suggest that, when discrimination dominates amenities as a source of gender gaps, the gender gap and the female share are negatively related, because firms that hire more women must be less discriminatory. The opposite happens when amenities dominate discrimination. Under our baseline calibration, the two forces almost exactly compensate each other, consistently with the relationship between gender gaps and female shares found in the data.

Obviously, if all sources of differences between men and women are turned off, the gender gap becomes zero, the wage variance of women becomes identical to that of men and the model generates no variation in the female share, as women become identical to men except that they are present in the economy in lower numbers.
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 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 around the time of maternity leave in this event study approach. If anything the wage growth of mothers, conditional on remaining employed 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.
7.1 7.2 7.3 7.4 7.5 7.6
Mean monthly earnings (log real BRL 2016)
control treatment (maternity leave in 2007)

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 fundamentally 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 contrast 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.
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.\(^9\) 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 \(E \left[ \alpha_j | male \right] - E \left[ \alpha_j | female \right] = 0.078\). Second, from Figure 20 we see that the male firm fixed effect slope in labor productivity is \(\partial E \left[ \alpha_j | male \right] / \partial [value added p.w.] = 0.197\). Putting the two pieces together, we can compute the hypothetical potential output gains from allocating women

\(^9\)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 and 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-level 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, taste-based discrimination among employers can explain around two-thirds of the gender gap. 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


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

\[
\max_{w_{a,m}, w_{a,f}, v_{a,m}, v_{a,f}} \left\{ (p_a - w_{a,m}) l_{a,m}(w_{a,m} + \pi_m, v_{a,m}) + (p_a - z - w_{a,f}) l_{a,f}(w_{a,f} + \pi_f, v_{a,f}) - c(v_{a,m} + v_{a,f}) \right\}
\]

(21)

The FOC for vacancy posting now read:

\[
c'(v_{a,m}) = T_{a,m}(a p + \pi_m - \bar{w}_{a,m}) \left( \frac{1}{\delta_{a,m} + s_{a,m} \lambda^{m,u}_{a}} \right) \left( \frac{1}{1 - F_{a,m} \bar{w}_{a,m}} \right)^2
\]

\[
c'(v_{a,f}) = T_{a,f}(a p + \pi_f - z - \bar{w}_{a,f}) \left( \frac{1}{\delta_{a,f} + s_{a,f} \lambda^{f,u}_{a}} \right) \left( \frac{1}{1 - F_{a,f} \bar{w}_{a,f}} \right)^2
\]

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:

\[
\max_{w_{a,m}, w_{a,f}, v_a} \left\{ (p_a - w_{a,m}) l_{a,m}(w_{a,m} + \pi_m, v_{a,m}) + (p_a - z - w_{a,f}) l_{a,f}(w_{a,f} + \pi_f, v_{a,f}) - c_a(v_a) \right\}
\]

(22)
Similarly to the separate vacancy posting model, we can rewrite the problem in terms of offering effective wages \( \tilde{w} = w + \pi \):

\[
\max_{w_{a,m},\tilde{w}_{a,m},v_a} \left\{ (p a + \pi_m - \tilde{w}_{a,m})l_{a,m}(\tilde{w}_{a,m}, v_a) + (p a + \pi_f - z - \tilde{w}_{a,f})l_{a,f}(\tilde{w}_{a,f}, v_a) - c_a(v_a) \right\}
\]

(23)

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

\[
m_{a,g} = A^\delta [\mu_{a,g} (u_{a,g} + \lambda_{a,g}^e (1 - u_{a,g}))]^a V^{1-a} \frac{V_{a,g}}{V}
\]

(24)

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. \( f_{a,g} / q_{a,g} = V / [u_{a,g} + \lambda_{a,g}^e (1 - u_{a,g})] \), where \( f_{a,g} = m_{a,g} / [u_{a,g} + \lambda_{a,g}^e (1 - u_{a,g})] \) 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

\[
V_{a,g} = \int_v v_a(p, \pi, z) dD_{a,g}(p, \pi, z)
\]

(25)

B.3 Equilibrium of the JVPM

Define \( \lambda_{a,g}^e = s_{a,g} \lambda_{a,g}^u \). The following equation represents the law of motion of firm sizes:

\[
l_{a,g}(\tilde{w}, v) = - \delta_{a,g} l_{a,g}^S(\tilde{w}, v) - s_{a,g} \lambda_{a,g}^e (1 - F_{a,g}(\tilde{w})) l_{a,g}^S(\tilde{w}, v) + \left[ \frac{u_{a,g}}{u_{a,g} + (1 - u_{a,g}) s_{a,g}} + \frac{(1 - u_{a,g}) s_{a,g}}{u_{a,g} + (1 - u_{a,g}) s_{a,g}} G_{a,g}(\tilde{w}) \right] q_{a,g}
\]

(26)

Solving for the stationary solution:

\[
l_{a,g}(\tilde{w}_{a,g}, v_a) = \left( \frac{1}{\delta_{a,g} + s_{a,g} \lambda_{a,g}^u (1 - F_{a,g}(\tilde{w}_{a,g}))} \right)^2 \frac{v_a}{V_a} \mu_{a,g} u_{a,g} \lambda_{a,g}^u (\delta_{a,g} + s_{a,g} \lambda_{a,g}^u)
\]

(27)
B.4 Equilibrium Characterization of the JVPM

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

$$\max_{\bar{w}_{a,m},\bar{w}_{a,f},v_{a}} \left\{ T_{a,m}v_{a}(a p + \pi_{m} - \bar{w}_{a,m}) \left( \frac{1}{\delta_{a,m} + s_{a,m}\lambda_{a}^{M,u} (1 - F_{a,m}(\bar{w}_{a,m}))} \right)^{2} \right. $$

$$\left. + T_{a,f}v_{a}(a p + \pi_{f} - z - \bar{w}_{a,f}) \left( \frac{1}{\delta_{a,f} + s_{a,f}\lambda_{a}^{M,u} (1 - F_{a,f}(\bar{w}_{a,f}))} \right)^{2} - c_{a}(v_{a}) \right\} \quad (28)$$

The associated FOC read:

$$c'(v_{a}) = T_{a,m}(a p + \pi_{m} - \bar{w}_{a,m}) \left( \frac{1}{\delta_{a,m} + s_{a,m}\lambda_{a}^{M,u} (1 - F_{a,m}(\bar{w}_{a,m}))} \right)^{2} $$

$$+ T_{a,f}(a p + \pi_{f} - z - \bar{w}_{a,f}) \left( \frac{1}{\delta_{a,f} + s_{a,f}\lambda_{a}^{M,u} (1 - F_{a,f}(\bar{w}_{a,f}))} \right)^{2} $$

$$1 = (a p + \pi_{m} - \bar{w}_{a,m}) \frac{2s_{a,m}\lambda_{a}^{M,u} f_{a,m}(\bar{w}_{a,m})}{\delta_{a,m} + s_{a,m}\lambda_{a}^{M,u} (1 - F_{a,m}(\bar{w}_{a,m}))} $$

$$1 = (a p + \pi_{f} - z - \bar{w}_{a,f}) \frac{2s_{a,f}\lambda_{a}^{M,u} f_{a,f}(\bar{w}_{a,f})}{\delta_{a,f} + s_{a,f}\lambda_{a}^{M,u} (1 - F_{a,f}(\bar{w}_{a,f}))} \quad (29)$$

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 27 we can write the female share of a firm as

\[ s_f = \frac{l_{a,f}(\tilde{w}_{a,f}, v_a)}{l_{a,f}(\tilde{w}_{a,f}, v_a) + l_{a,m}(\tilde{w}_{a,m}, v_a)} \]

\[ = \frac{\left( \frac{1}{\delta_{a,f} + s_{a,f} \lambda_a^{f,m}(1 - F_{a,f}(\tilde{w}_{a,f}))} \right)}{1 + \left( \frac{1}{\delta_{a,m} + s_{a,m} \lambda_a^{m,u}(1 - F_{a,m}(\tilde{w}_{a,m}))} \right)} \]

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 \( \lambda_a^{f,m} = \lambda_a^{m,u} = \lambda_a^u \). Therefore unemployment rates are also identical across genders, so that the expression for the female share simplifies to

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

Since firm sizes are monotonically increasing in the effective wage \( \tilde{w} \) offered by the firm, we can obtain the minimum female share \( s_f \) 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 \( F_{a,m} = 1 \). Then we can

\[ 10 \text{This firm might not actually exist in our simulations as male and female draws are correlated, thereby making it extremely unlikely that such a firm is drawn. Then, we will have that our simulations produce a minimum female share that is slightly larger than this theoretical lower bound.} \]
write

\[ s_f = \frac{1}{1 + \left( \frac{1}{\delta + s_{a,m} \lambda^u_2} \right)^2 (\delta + s_{a,f} \lambda^u_2)} = \frac{1}{1 + \frac{1}{\delta + s_{a,m} \lambda^u_2}(\delta + s_{a,f} \lambda^u_2)} = \frac{1}{1 + \frac{(\delta + s_{a,m} \lambda^u_2)(\delta + s_{a,f} \lambda^u_2)}{\delta^2}} = \frac{1}{2 + \frac{s_{a,f} \lambda^u_2}{\delta} + \frac{s_{a,m} \lambda^u_2}{\delta} + \frac{s_{a,m} s_{a,f} (\lambda^u_2)^2}{\delta^2}} \]

In our data we find roughly that \( s_{a,m} \lambda^u_2 \simeq 0.4 \delta \) and \( s_{a,f} \lambda^u_2 \simeq 0.27 \delta \). If we apply these numbers, we find the minimum possible female share predicted by our model:

\[ s_f = \frac{1}{2 + 0.27 + 0.4 + 0.108} = 0.36 \]

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

\[ \bar{s}_f = \frac{1}{1 + \left( \frac{1}{\delta + s_{a,m} \lambda^u_2} \right)^2 (\delta + s_{a,f} \lambda^u_2)} = \frac{1}{1 + \frac{1}{\delta + s_{a,m} \lambda^u_2}(\delta + s_{a,f} \lambda^u_2)} = \frac{1}{1 + \frac{(\delta + s_{a,m} \lambda^u_2)(\delta + s_{a,f} \lambda^u_2)}{\delta^2}} = \frac{1}{1 + \frac{\delta^2}{\delta^2 + s_{a,m} \lambda^u_2 + s_{a,f} \lambda^u_2 + s_{a,m} s_{a,f} (\lambda^u_2)^2}} \]

\[ \simeq 1 + \frac{1}{\frac{\delta^2}{\delta^2 + 0.4 \delta^2 + 0.27 \delta^2 + 0.108 \delta^2}} \simeq 0.64 \]

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:

\[
p_{a,m} = a p + \pi_m
\]
\[
p_{a,f} = a p + \pi_f - z
\]

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

\[
v_a = \frac{T_{a,m}}{c} (p_{a,m} - \bar{w}_{a,m}) \left( \frac{1}{\delta_{a,m} + s_{a,m} \lambda_{a,m}^{m,m} (1 - F_{a,m}(\bar{w}_{a,m}))} \right)^2 + \frac{T_{a,f}}{c} (p_{a,f} - \bar{w}_{a,f}) \left( \frac{1}{\delta_{a,f} + s_{a,f} \lambda_{a,f}^{m,f} (1 - F_{a,f}(\bar{w}_{a,f}))} \right)^2 \tag{31}
\]

By definition:

\[
V_{a,m} = \int \int v_a [p_{a,m} > \Phi_{a,m}] \gamma(p_{a,m}, p_{a,f}) dp_{a,m} dp_{a,f}
\]
\[
V_{a,f} = \int \int v_a [p_{a,f} > \Phi_{a,f}] \gamma(p_{a,m}, p_{a,f}) dp_{a,m} dp_{a,f}
\]
\[
V_a = \int \int v_a [p_{a,m} > \Phi_{a,m} \text{ OR } p_{a,f} > \Phi_{a,f}] \gamma(p_{a,m}, p_{a,f}) dp_{a,m} dp_{a,f} \tag{32}
\]

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

\[
h(p) = F(\bar{w}(p))
\]
\[
h'(p) = f(w(p)) w'(p), \text{ therefore}
\]
\[
f(w(p)) = h'(p) / w'(p)
\]
\[
\bar{v}_a(p_{a,m}) = \frac{V_{a,m} h'(p_{a,m})}{\gamma(p_{a,m})}, \text{ therefore}
\]
\[
h'(p_{a,m}) = \frac{\bar{v}_a(p_{a,m})}{V_{a,m}} \gamma(p_{a,m})
\]

Thus, the wage FOC can be rewritten as follows:
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 $\tilde{\omega}(p_{a,g})$, $F(\tilde{\omega}(p_{a,g}))$ 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 33.

5. Use $h'(p_{a,g})$ to compute the new function $h(p_{a,g}) = F(\tilde{\omega}(p_{a,g}))$.

6. Use $h'$ and $h$ to solve the ODE in B.6.

7. Update the wage function $\tilde{\omega}(p_{a,g})$ and the CDF $F(\tilde{\omega}(p_{a,g}))$.

8. Go back to step 2. Repeat until convergence.