CESifo WORKING PAPERS

8543
2020

September 2020

Knowledge Intensity and Gender Wage Gaps: Evidence from Linked Employer-Employee Data

Radu Barza, Cristian Jara-Figueroa, César A. Hidalgo, Martina Viarengo

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8543*

Knowledge Intensity and Gender Wage Gaps: Evidence from Linked Employer-Employee Data

Abstract

Do knowledge intense jobs exhibit lower gender gaps in wages? Here we use a linked employer-employee dataset of the entire Brazilian formal labor force to study the relationship between gender wage gaps and the knowledge intensity of industries and occupations. We find that employees in high-skilled occupations and industries experience lower gender wage gaps, and that the effect of knowledge intensity is stronger when the demand for skilled labor is high and the supply of skilled labor is low. We also find evidence that the gender wage gap of skilled workers, but not that of unskilled workers, decreases when knowledge intense industries grow. These effects are robust to controlling for individual, occupation, sector, and location characteristics. To address endogeneity concerns, we use a Bartik instrument based on labor demand shocks. Together, these findings suggest that competition for skilled labor in knowledge intense industries contributes to the reduction of gender wage gaps.

JEL-Codes: J200, O100, I250.

Keywords: knowledge intensity, economic development, labor markets, gender gaps.

Radu Barza
Department of Economics
The Graduate Institute
Geneva / Switzerland
radu.barza@graduateinstitute.ch

Cristian Jara-Figueroa
The MIT Media Lab, Massachusetts Institute of Technology
Cambridge / MA / USA
crisjf@mit.edu

César A. Hidalgo
Artificial and Natural Intelligence Toulouse Institute (ANITI), University of Toulouse / France
cesifoti@gmail.com

*Martina Viarengo**
Department of Economics
The Graduate Institute
Geneva / Switzerland
martina.viarengo@graduateinstitute.ch

*corresponding author

August, 2020 / Preliminary and Confidential / Please do not circulate without the Authors' permission
Barza and Viarengo gratefully acknowledge the support received from the Swiss National Science Foundation. We would like to thank Daron Acemoglu, Tim Swanson, and seminar participants of the Centre for International Environmental Studies (CIES) at The Graduate Institute, Geneva for helpful discussions and suggestions on an earlier draft of the paper.

1. Introduction

Becker's 1957 theory of discrimination suggests that firms that choose workers based on characteristics that are unrelated to their productivity—such as race or gender—must pay a cost. This is the cost of accepting suboptimal matchings, which is larger for knowledge intense economic activities[2]. Since economic development is about the growth of knowledge (Romer 1990), the cost of choosing workers based on characteristics that are unrelated to their productivity should increase with economic development.

Together, Becker and Romer's theories make an interesting prediction about the gender inequalities in wages. They predict gender wage inequalities should decrease with the knowledge intensity of an economy. Moreover, since the effects are driven by the cost of accepting suboptimal workers, we should expect this effect to be more noticeable among high-skilled occupations in knowledge intense economic activities.

Here, we used a linked employer-employee dataset for the universe of the Brazilian labor force to test the relationship between knowledge intensity and gender wage gaps. Our data shows that workers in skilled occupations in knowledge intense industries experience comparatively lower gender wage gaps, and that this effect is larger in Brazilian states with few high-skilled workers and rapidly growing knowledge intense industries. This observation is consistent with the idea that competition for high-skilled labor reduces gender inequality, since competition for skilled labor is larger in states with few skilled workers and rapidly growing knowledge intense industries. To address endogeneity concerns, we introduce a Bartik instrument that uses national level variations in

[2] Unless otherwise specified, throughout the paper we use the term *industries* to refer to the two-digit disaggregate codes that define the economic activities in Brazil. The term *sectors* refer to the aggregate indicator of the 21 sectors of the economic activity.

employment as shocks at the local level. We find these estimates to be consistent with those from hedonic wage regressions. Together, these findings validate the idea that the growth of knowledge intense industries in an economy can contribute to the reduction of gender wage disparities.

2. Gender Gaps and Knowledge Intensity

The process of economic development leads to changes in the structure of the labor market and to the patterns of inequality (Acemoglu 2002; Acemoglu and Autor 2011; Acemoglu 2012). Understanding gender wage gaps is important for both equity-related reasons and efficiency considerations. Our work is related to both of these strands of literature.

Gender equality is known to improve with economic growth, well-being, and efficiency (Becker and Lewis 1974; Galor and Weil 1996; Lagerlöf 2003; Greenwood, Seshadri, and Yorukoglu 2005; Duflo 2012; Bandiera and Natraj 2013; Bandiera et al. 2015; Cavalcanti and Tavares 2016)[3]. Likewise, gender inequalities seem to hinder development (Hill and King 1995; Dollar and Gatti 1999; Tzannatos 1999; Forbes 2000; Knowles, Lorgelly, and Owen 2002). Even though causality is still debated, cross-country studies show a strong correlation between a country's level of development and reductions in gender inequality in outcomes such as labor force participation (Goldin 1995; Goldin 2006; Ganguli, Hausmann, and Viarengo 2013; Jayachandran 2019), human capital investment (Benhabib and Spiegel 1994; Tamura 2006), and wages (Elborgh-Woytek et al. 2014; Olivetti and Petrongolo 2016). Yet, despite these findings, we still

[3] A detailed review of both theoretical and empirical studies is presented in Cuberes and Teignier (2014).

lack a comprehensive understanding of the mechanisms by which economic development leads to a reduction in gender gaps in labor market outcomes.

More than thirty years ago, Claudia Goldin (1990, 1995) prominently provided the theoretical framework and tested the hypothesis of a U-shaped relationship between female labor force participation and economic development.[4] At low levels of income, female participation in productive activities is high. This consists, usually, of young and unmarried women working in factories and paid by the item (Goldin 1994). But as economies and technology develops, female labor force participation decreases together with the need for manual work and with the rise of spousal earnings (Goldin 1991, 1994, 2000). This trend reverses, however, when technology makes some occupations more suitable to women (Goldin 2006; Goldin and Katz 2016)[5], and when development increases women's capacity to plan careers (Goldin and Katz 2002)[6], creating an incentive for human capital investment (Goldin 2005). This U-shaped relationship has been documented in cross-country comparisons and country-specific studies (Goldin 1995; Mammen and Paxson 2000; Luci 2009; Gaddis and Pieters 2012, 2017; Gaddis and Klasen 2014; Olivetti 2014; Verick 2014). This pattern has been documented for the economically advanced economies. Yet, it is unclear if this pattern holds for developing countries. Recent studies find no evidence for the U-shaped relationship or provide arguments for more complicated correlations (World Development Report 2012; Gaddis and Klasen 2014; Jayachandran 2019).

[4] Sinha (1967) first discussed in a descriptive way the U-shaped hypothesis.

[5] Goldin and Katz (2016) explain how transformations within the pharmaceutical profession (lower penalties to part time, democratization of pharmacies in retail and hospital pharmacies) have made this profession one of the most egalitarian, with lower gender gap in wages and with high female share of female employment.

[6] This paper explains how the availability of the contraceptive pill gave women the possibility to plan careers and thus reduce the relative cost of their education, and to increase the age of first marriage.

Beyond labor force participation, scholars have looked at the link between economic development and gender differences in human capital formation and accumulation. Economic development is linked to women's incentive to invest in human capital by extending the service sector (Becker 1992), providing mechanical power (Galor and Weil 2000), improving household technology (Killingsworth and Heckman 1986; Goldin 2006; Bandiera et al. 2015), facilitating birth control, promoting investments in education (Goldin and Katz 2002), and through the development of college preparation for girls (Goldin, Katz, and Kuziemko 2006). As Goldin (2006) explains the expected benefits from increased education encourage women to have careers instead of jobs. Petrongolo and Ronchi (2020) review the literature examining the expansion of the service economy and the increased participation of women in the labor market of developed countries.

Our findings are consistent with this literature and contribute to it by documenting a link between the gender gap in wages and the knowledge intensity of industries and occupations. Specifically, our results imply that as local economies evolve towards a mix of more knowledge intense industries, gender wage gaps are reduced for high-skilled occupations, especially when demand for high-skilled labor increases faster than supply. These findings help bridge the literature on gender inequalities with the literature on knowledge intensity advanced by economic geographers and innovation economists.

Our results are also related to the literature quantifying the knowledge intensity of economic activities. While the role of knowledge has been long recognized by scholars working in evolutionary economics (Nelson and Winter 1982; Dosi et al. 1988), endogenous growth (Romer 1990; Aghion et al. 1998), and innovation (Jaffe, Trajtenberg, and Henderson 1993; Audretsch and Feldman 1996), the recent emergence of metrics of economic complexity (Hidalgo and Hausmann 2009) have motivated a new line of empirical research. These metrics have helped establish the idea that more complex, or knowledge intense economies, grow faster (Hidalgo and Hausmann 2009; Hausmann et al. 2014; Stojkoski, Utkovski, and Kocarev 2016), exhibit lower levels of income inequality (Hartmann et al. 2017), and have lower emission intensities (Garas and Lapatinas 2017; Neagu and Teodoru 2019; Romero and Gramkow 2020). Hence, it is reasonable to ask whether knowledge intense industries and occupations are also connected to lower levels of gender inequality. In fact, a recent paper provides cross-country evidence in this direction (Ben Saâd and Assoumou-Ella 2019).

Here, we use a knowledge intensity metric developed originally for patents by Fleming and Sorenson (2001), to measure the knowledge intensity of industries in Brazil. This metric focuses on how easy it is to re-combine inputs to produce an output (in our case, occupations across industries). We estimate this metrics using a dataset covering 13 years of linked employer-employee data for the universe of the formally employed workers in Brazil and use it to study how gender gaps in labor market outcomes change with economic development.

By using individual level data, we can estimate the contribution of knowledge intensity to gender wage disparities. We find that high-skilled occupations in knowledge intense industries exhibit lower gender wage gaps. In fact, under certain conditions the effect is so strong that the gap reverses (women earn more). Also, we show that states with few high-skilled workers and fast-growing knowledge intense industries are characterized by a reduced gender wage gap. This shows the effect is stronger in places where competition for high-skilled labor is larger, since these are places that are constrained on their supply of high-skilled workers while experiencing growing demand.

Our results support the hypothesis that gender wage gaps for high-skilled occupations are reduced in more knowledge intense industries and provide empirical evidence that the development of knowledge intense industries accelerates the reduction of gender gaps in wages.

The remaining of the paper is organized as follows: Section 3 presents the theoretical foundations of our analysis; Section 4 presents the data and the methods used. Our main results are reported in Section 5, including various robustness checks. Section 6 summarizes the key findings and discusses future perspectives for research on gender gaps in compensation.

3. Theoretical Foundation

Why would average gender wage differences be lower in high-skilled occupations in knowledge intense industries? The potential mechanism we put forward is that of competition for skilled human capital.

When market competition is high, companies need to make sure they hire the best people they can find. If companies are discriminating on any observables not related to productivity, such as gender, they can fail to survive. Supply and demand forces, therefore, imply lower gender wage differences, both at hiring and along the career path, when companies are constrained by the supply of high-skilled individuals.

We define gender inequality in wages as compensation differences among individuals of same observable characteristics (e.g., age, ethnicity, level of educational attainment, experience, company tenure, hours worked, leaves of absence and interruption reasons). This leaves out gender self-selection in occupations and industries, which we document and explain later.

Becker (1957)[7] argues that some employees discriminate against some prejudiced groups (women, racial minorities) and are willing to pay a cost for this. The payment can be direct, such as higher wages, or indirect through reduced income or profits:

> *If an individual has a "taste for discrimination," he must act as if he were willing to pay something, either directly or in the form of a reduced income, to be associated with some persons instead of others. When actual discrimination occurs, he must, in fact, either pay or forfeit income for this privilege. This simple way of looking at the matter gets at the essence of prejudice and discrimination.* (p.14 in Becker 1957).

Let us consider prejudice against female workers. In this setup, an employer that discriminates against women either hires less women than what would be optimal for them or reduces profits by choosing to pay male workers higher wages than those of equivalent female workers. Companies that do not discriminate, hire the best matched individuals for each position regardless of their gender. If markets are competitive (Becker 1957; Arrow 1973), and knowledge is a key input, then non-discriminating companies should outcompete discriminating companies (Becker 1957; Goldberg 1982;

[7] Since Becker (1957), the literature on gender discrimination models has been divided into two branches: (i) competitive models and (ii) collective models . Competitive models include three main categories: taste-based, self-selection, and statistical discrimination. The first category (Becker 1957) relies on a basic model of taste-based discrimination. The second category (Bergmann 1974; G. E. Johnson and Stafford 1995, 1998;) relies on self-selection and explains occupation segregation by pointing out the possibility of differences in the degree of discrimination between certain occupations and look at how members of different groups self-select into different occupations. Yet, this literature fails to propose theoretical models capable of exposing the mechanisms through which social norms or institutional constraints arise and are sustained. A third category looks at statistical discrimination, either through stereotypes models (Phelps 1972; Arrow 1973; Farber and Gibbons 1996; Altonji and Pierret 2001), or information asymmetry on the employer side about the human capital quality in each of the groups (Aigner and Cain 1977; Lundberg and Startz 1983; Lundberg 1991). One drawback of these models is that expectations about stereotypes can be self-fulfilling. More recent literature on discrimination shows that globalization and deregulation trends reduce discrimination against women in some groups of occupations (Black and Strahan 2001; Black and Brainerd 2004). For instance, Black and Juhn (2000) shows that the shift towards a service and skill-intensive economy has increased the proportion of jobs suitable for women. Bertrand (2018) explains that women still face discrimination when they compete for high pay occupations, especially when they require more flexible working hours due to specific household needs. We contribute to this literature by proposing a mechanism reducing the gender wage gap within occupations within industries.

Heckman 1998; Black and Brainerd 2004). This mechanism works when skilled labor supply is low and demand for it is high.

If a firm discriminates based on characteristics that relate weakly to productivity, it creates an opportunity for a second firm to attract any highly productive workers from the prejudiced group. If these workers are in short supply, firms would struggle to fill such positions and the market would select against the first firm. When specific skills are scarce, companies competing for these skills would find it unprofitable to discriminate based on gender or other personal characteristics. This proposed mechanism leads to a reduction of the gender wage gaps when skills are scarce and predicts the gap in wages between observationally identical female and male workers to reduce, or disappear, when competition for skills is important.

4. Data and Methods

A. Data

We use the employer-employee linked dataset, which includes the universe of all formally employed individuals in Brazil between 2003 and 2015. The Annual Social Information Report - *Relação Anual de Informações Sociais* (RAIS) is an administrative register compiled by the Ministry of Labor (MTE) based on the information that all formally registered, public or private companies, have the obligation to provide (Cardoso et al. 2007). RAIS was created to administer and control access to unemployment insurance and other pecuniary benefits to workers, thus creating an additional incentive to report correctly. To further improve data quality and ensure compliance, the MTE cross-tabulates registry information from many other official sources, such as the Ministry of Social Security, the Federal Reserve and the Secretary of Federal Revenues (taxes).

Consequently, MTE estimates that RAIS is annually declared by 98% to 99% of officially existing firms (Cardoso et al. 2007). Information in RAIS is collected annually.

The variables in RAIS are at the individual level, which makes it the most important source of information on the formal labor market dynamics in the country. RAIS contains a unique time-invariant identifying code for each individual. This allows us to follow individuals over time. The same is true about companies, each company has a unique time-invariant code. Data includes information about the monthly wage of an individual, number of weekly hours worked, interruption reasons and duration, gender, age, education, ethnicity, geographical location of the job, the size of the firm and of the establishment, nature of the work contract, termination reason, as well as the juridical nature of the company. Throughout our analysis, we use the average monthly wage specific to each individual-occupation-firm-year adjusted by inflation, with 2010 as the reference year. RAIS spans fine-grained information about individual workers in Brazil, including 5,560 municipalities, 2,500 occupations, and 585 industries for more than 30 million workers each year, with the most important variables for our work being hourly wages, gender, work interruptions, individual characteristics, occupational and sectoral codes, and individuals' work history across establishments.

B. Empirical Method

We divide our empirical analysis into three steps. First, using classical methods of decomposition (Blinder 1973; Oaxaca 1973; Fortin, Lemieux, and Firpo 2011), we document the role that individual-level characteristics, gender-specific occupational and industrial segregation play in explaining the gender wage gap across states. We then estimate the share of the gender wage gap that remains unexplained.

Second, for each state, we estimate the effect of the knowledge intensity of each industry on gender gaps in wages. We find that knowledge intense industries have lower gender gaps in wages in those states with undersupply of high-skilled labor.

Third, because the dynamics of labor markets are endogenous, we correct our estimates with instrumental variable techniques using Bartik instruments for labor demand shocks (Bartik 1991).

a. *Baseline Analysis of Gender Gaps in Labor Markets: Sectoral and Occupational Segregation*

First, we decompose the gender gap in wages between the part explained by segregation within occupations within sectors and the part that remains unexplained. The two main assumptions of our analysis are that wages are a linear function to the covariates, and that the statistical remainder ε_{ig} is conditionally independent of all covariates (i.e. $E\left[\varepsilon_{ig} \middle| X_{ig}\right] = 0$). Our analysis relies on the following hedonic wage function:

$$\omega_{jgriot} = X_{jgriot}\, \beta_g + \varepsilon_{jgriot}, \qquad \text{g} = \text{Female, Male} \tag{1}$$

where the subscript $jgriot$ relates to individual j, of gender g, in region r, sector i, occupation o, in time t, ω stands for log of monthly wages, and X_{jgriot} is a vector of individual characteristics including age, experience (i.e., proxied by age-squared), ethnicity, level of educational attainment, company tenure, hours worked per week, days away from work, and interruption reasons. Then, the gap in wages between female (f) and male (m), conditional on individual observable characteristics, is obtained as the average difference between the two groups:

$$\hat{\Delta} = \overline{\omega_f} - \overline{\omega_m} = E[\omega^f \middle| X^f] - E[\omega^m | X^m] \tag{2}$$

We can think of the effect of gender for each worker on the gender gap in wages: $\hat{\Delta} = \omega_j^f - \omega_j^m$ as the individual treatment effect. To estimate the gender gap in wages, we can rearrange Equation 2 and estimate the following linear equation:

$$\omega_{jriot} = \gamma \times D_j + X_{jriot}\, \beta + \varepsilon_{jriot} \tag{3}$$

where D_j is a dummy variable taking value 1 if the individual j is a female, and value 0 otherwise. Estimates of γ in Equation 3 offer the magnitude of the gender gap in wages conditional on observed individual characteristics such as age, age-squared, company tenure, hours worked per week, days away from work, interruption reasons, attained education level, and ethnicity. This means that we obtain an estimate of the gender gap in wages for observationally identical individuals, the matching being defined by the set of variables captured in the vector X_{jriot} on which we condition the estimation. As our dependent variable is measured in logs, the γ coefficient can be interpreted as percentage difference in wages between observationally identical female and male workers (Wooldridge 2016).

It makes sense to assume that the unobserved error term consists of several components:

$$\varepsilon_{jriot} = \alpha_t + \alpha_i + \alpha_o + \alpha_i \times \alpha_o + \eta_{jriot} \tag{4}$$

where α_t is the time fixed effect; α_i is the sector fixed effect; α_o is the occupation fixed effect, and η_{jriot} is the statistical residual. In this way, we obtain an estimate of the gender gap in wages at sectoral level conditional on individual characteristics. Here we focus on how much of the gender gap in wages is explained, conditional on individual characteristics, by the differences in compensation between sectors and between occupations and how much of the gender difference in wages still remains for observationally identical individuals within occupations within sectors of the economy.

To this end, we estimate the wage Equation 3 and gradually introduce new levels of fixed effects. The percentage change in the estimated coefficient γ associated with the dummy variable D_j shows how much of the gender gap in wages is explained by the newly introduced fixed effects. The new estimated γ coefficients tell us how much of the gender gap in wages is explained by between groups of fixed effects introduced and how much of the gender gap remains to be explained by other mechanisms within groups. Given that the error term specific to each individual for one observation can be correlated across observations in different years and occupations, and given that error terms associated with each company are correlated across individuals, occupations and time, all standard errors reported in this paper are calculated using a double clustering procedure at individual and firm level.

b. The Role of Knowledge Intensity

We define high-skilled occupations as occupations in which with probability one any employee has at least completed some university studies. We measure knowledge intensity of industries using the methodology proposed by Fleming and Sorenson (2001). For each industry, we measure how common is the recombination of occupations it employs. The knowledge intensity of each industry is defined then as the average ease of recombination of the occupations present in that industry, and the ease of recombination of each occupation is defined as the number of occupations that co-appear with that occupation divided by the number of industries that occupation appears in. Specifically,

$$K_i = \frac{1}{\frac{1}{M_i}\sum_o M_{io}E_o} \tag{5}$$

$$E_o = \frac{1}{M_o} \mathbb{1}\left(\sum_i M_{io} M_{io'} > 0\right), \tag{6}$$

where $M_{io} = 1$ if industry i hires occupation o and zero otherwise, $M_i = \sum_o M_{io}$ is the number of occupations hired by industry i, and $M_o = \sum_i M_{io}$ is the number of industries that require occupation o. Intuitively, occupations that are not easily recombined tend to be very specialized. Therefore, industries that hire many specialized occupations will tend to be more knowledge intense because they need to combine a large number of specific inputs. This notion of knowledge intensity is true to the idea that knowledge intense activities are those that require a deep division of knowledge.

To filter out the noise generated by the fact that some industries are larger than others, and hire at least one worker from almost any occupation, we use a ratio of ratios method that mirrors the Revealed Comparative Advantage (RCA) (Balassa and Noland 1989) or Location Quotient idea from Urban Planning and Economic Geography (Miller, Gibson, and Wright 1991). An occupation is said to be present in an industry when it has an RCA of above one.

$$M_{io} = \mathbb{1}(RCA_{io} \geq 1) \tag{7}$$

$$RCA_{io} = \left. \frac{L_{io}}{L_i} \middle/ \frac{L_o}{L} \right. \tag{8}$$

Where L_{io} is the employment of occupation o in industry i, $L_i = \sum_o L_{io}$ is the total employment in industry i, $L_o = \sum_i L_{io}$ is employment in occupation o, and L is the total employment in the country.

Our variable of interest "K" measures the knowledge intensity of an industry. The index is constructed for each 2-digit economic activity at the federal state level, one industry has the same knowledge intensity index across states, for each year and is gender invariant. Higher values of “K” are associated with more knowledge intense industries, while the converse holds true.

For interpretation reasons, we use a normalized measure of the knowledge intensity, and we standardize the initial distribution of the “K” index to a standard normal ~ N (0,1). In Table 1 we report the distribution of the “K” index[8]. In Table 2 we report industries among the top and the bottom 5 percentile of the distribution of the knowledge intensity index.

c. Growth

The annual growth rate of the share of high-skilled occupations defined in Equation 9 is calculated for each region for the current year as the fraction between the share of the high-skilled occupations in that region in the current year ($S_{HS,t,r}$) and the share of the high-skilled occupations in the previous year ($S_{HS,t-1,r}$) minus one. In this way, we can calculate the annual growth rates for each of the years between 2004 and 2015, 2003 being the first year in our dataset, and thus making it impossible to calculate its growth rate.

$$G_{t,r} = \frac{S_{HS,t,r}}{S_{HS,t-1,r}} - 1 \tag{9}$$

d. Expanded Empirical Model

The mechanism outlined in the theory section implies that the gender gap in hourly wages for workers in high-skilled occupations should be lower in knowledge intense industries. To test this, we modify Equation 3 and introduce the following interaction terms:

[8] Density distributions are reported in Figure A.2. of the Appendix.

$$\omega_{jriot} = \gamma_0 D_j + \rho_1 HS_o + \rho_2 K_i + \rho_3 G_{t,r}$$

$$+ \gamma_1 D_j \times HS_o + \gamma_2 D_j \times K_i +$$

$$+ \gamma_3 D_j \times HS_o \times K_i +$$

$$+ X_{jriot}\, \beta + \varepsilon_{jriot} \tag{10}$$

Where HS_o is a dummy equal to 1 if the occupation is a high-skilled occupation, and K_i is our measure of knowledge intensity of the economic activity normalized to a standard normal.

The high-skilled dummy is constructed at the four-digit occupation codes and is taken as constant for each occupation for all years. To estimate the parameters $\gamma_1 \, and \, \gamma_3$ we use occupation fixed effects at the three-digit occupation codes. Estimates also include the intensity of competition for high-skilled labor at 2-digit industry codes by micro-region, measured using the annual growth rate of the share of high-skilled occupations by industry in each micro-region ($G_{t,r}$). As theory of supply and demand suggests, labor market forces have explanatory power on wages, therefore it is necessary to introduce this measure in the wage equation.[9]

Estimating coefficients in Equation (10) gives us two main results. First, the estimate of the parameter γ_3, which is the parameter of the three-way interaction term of females ($D_j = 1$) in high-skilled occupations ($HS_o = 1$) in industries with increasing knowledge intensity (K_i), tells us what happens with the gender gaps in wages for high-skilled occupations when regions develop more knowledge intense industries. The estimate

[9] Indeed, our estimates confirm this. The coefficients associated with the growth rate of the share of high-skilled individuals by industry at micro-region level are all statistically significant. As the focus of this research is on gender gap in wages and for space reasons, we do not report these coefficients here. They are available from the Authors upon request.

captures at state level what the gender gap in wages for high-skilled occupations would be once the economy develops towards more knowledge intense industries by one standard deviation, as the index $K_i \sim N(0,1)$.

Second, as discussed in the methods section, the gender gaps in wages can be estimated by looking at the coefficient of the female dummy in the hedonic wage equation. In this specification, the gap has the form:

$$\hat{\Delta} = \gamma_0 + \gamma_1 HS_o + \gamma_2 K_i + \gamma_3 HS_o \times K_i \tag{11}$$

which for high-skilled workers and for low-skilled workers are reduced to:

$$\hat{\Delta}_{HS} = \gamma_0 + \gamma_1 + (\gamma_2 + \gamma_3) K_i \tag{12}$$

$$\hat{\Delta}_{LS} = \gamma_0 + \gamma_2 K_i \tag{13}$$

e. Endogenous Labor Market Demand

Finally, we use a Bartik instrument to address the possible smearing effect in the coefficient estimates since changes in employment level and labor demand are endogenous (implying that the growth rate of the share of high-skilled labor is endogenous with wages). Bartik shocks rely on interacting the industrial structure of a region with the estimated growth rate of each industry based on the national level trend. To make the estimated national level trend independent of the region in question, we remove the region to calculate the national level growth. Here we use Bartik shocks for high-skilled and for low-skilled workers, following Diamond (2016). Bartik shocks are defined as:

$$B_{r,t_o,t_f} \quad = \sum_i \frac{L_{ir,t_o}}{L_{r,t_0}} g_{i,t_0,t_f;r} \tag{14}$$

$$g_{i,t_0,t_f;r} \; = \log L_{i,t_f;r} - \log L_{i,t_0;r} \tag{15}$$

where $g_{i,t_0,t_f;r}$ is the observed growth in employment of industry i after removing region r, and $L_{i,t_f;r}$ is employment in industry i in year t_f everywhere but in region r. We instrument the growth rates of the shares of high-skilled workers by industry in the local labor markets by the Bartik labor demand shocks.

5. Results

A. Baseline Findings on Gender Gaps in Wages.

We start by documenting the role of individual-level characteristics and occupational and sectoral segregation in explaining gender wage gaps across states. Table 4 reports the coefficients of the gender dummy for each state in a regression of individual wages on different covariates (the γ coefficient of D_j variable in Equation 3). Overall, our analysis shows that roughly half of the gender gap in compensation across states in Brazil, after controlling for individual-level characteristics and year fixed effects, is explained by differences in compensation among occupations and economic sectors[10]. This suggests that on average men sort into higher-paying jobs (occupations and sectors) and women into lower-paying ones. It also leaves unexplained a gender difference in compensation

[10] Economic sectors (or sectors interchangeably throughout the text) refer to the 21 sectors of the economy as defined by the Ministry of the Economy of Brazil. They include: accommodation and food; administrative activities and complementary services; agriculture, livestock, forestry, fishing and aquaculture; arts, culture, sport and recreation; construction; domestic services; education; electricity and gas; extractive industries; financial, insurance and related services activities; human, health and social services; information and communication; IOs and other extraterritorial activities; manufacturing industries; other service activities; professional, scientific and technical activities; public administration, defense and social security; real estate activities; trade, repair of motor vehicles and motorcycles; transport, storage and mail; water, sewage, waste management and decontamination activities.

between 5.5% and 16.6% for observationally equivalent individuals within the same occupation in the same sector.

Occupational segregation explains about 40% of the gender gap in wages. This represents a large share of the gender gap in wages and varies across states between 5 and 15 log points of differences between the two genders. In absolute terms, this represents a gender difference in monthly wages between 27 and 83 US dollars (for an average monthly wage of 505 US dollars adjusted for 2010 inflation). Gender sorting across occupations within economic sectors explain roughly half of the gender gap in wages.

The residual gender gap in wages is the coefficient of the gender dummy once we account for all individual-level observables and for occupations and sectors fixed effects. This gap varies between 5.5% and 16.6% across states. These stylized facts are generally consistent with findings in the existing literature (e.g., Bergmann 1974; Blau, Simpson, and Anderson 1998; G. Johnson and Stafford 1998; Blau and Kahn 2006; Card, Mas, and Rothstein 2008; Pan 2015; Coudin, Maillard, and Tô 2018).

Figure 3 summarizes our findings by showing the share of explained/unexplained gender gap (Panel A), the share of the gap explained by differences across economic sectors (Panel B), occupations (Panel C), and occupations within sectors (Panel D).

B. The Role of Knowledge Intensity

Next, we document the role of knowledge intensity of an economic activity in explaining gender gaps in wages (Table 5), following equations 12 and 13. A positive sign of the coefficient γ_3—the triple interaction term between female, high-skilled, and knowledge intensity—means that the gender wage gap is lower in knowledge intense industries. This would suggest that as states move towards a more knowledge intense mix of industries the gender gap in wages for high-skilled occupations could be reduced. We find positive estimates for the coefficient γ_3 for all states, except for Serpige where the coefficient is not statistically different from zero, and for Piauí where the estimated coefficient implies an increase in the gender gap in wages.

Following on from the hypotheses derived from the underlying theory, the effect of knowledge intensity should be stronger when the local labor market is constrained on high-skilled labor and when the demand for it is increasing rapidly. Our results support this hypothesis and show that workers in high-skilled occupations in knowledge intense economic sectors experience a reduction of the gender wage gap. When the demand for high-skilled labor increases much faster than the supply, the gender gaps in wages is even reversed.

However, the gender wage gap is not reduced for all workers in complex industries. For example, in Rio de Janeiro, while the coefficient associated with a female in high-skilled occupations in knowledge intense industries indicates a reduction of the gender wage gap by 6 log points, females working in low-skilled occupations in these industries face an increased gender gap by 2.7 log points. In Acre, the average female working in a high-skilled occupation in a knowledge intense industry experiences a gender gap that is lower by 3.8%. In relative terms, the females in these occupations and industries face a reduction of 48.71% of the residual gender gap in wages. Again, low-skilled women in knowledge intense industries face an increase in the gender gap. The gap increases in absolute value by 1.1%, which represents a 14% increase of the gap. If we focus on São Paulo, Brazil's largest state, knowledge intensity reduces by very little the gender wage gap for high-skilled workers in knowledge intense industries (the coefficient equals 0.4%). We interpret these results as implying that even though São Paulo has an important share of knowledge intense industries these are not constrained by the supply of high-skilled workers.

C. Endogenous Labor Market Demand

As labor market forces are endogenous, we use Bartik labor demand shocks to instrument for the annual growth rate of the share of high-skilled occupations to correct our estimates. Results for the IV estimates are reported in Table 6. The interpretation of these coefficients is similar to that presented in section 5.B. Corrected results reported in Table 6 follow the same patterns as results in Table 5. The magnitude of some of the estimated coefficients changes slightly, which shows that indeed the IV estimates more precisely the coefficients of interest.

All estimated coefficients γ_3 are either positive or not statistically different from zero at 5% significance level, except of Piauí. These coefficients show that females working in high-skilled occupations have positive benefits from growth of knowledge intense industries. This demonstrates that the development of more knowledge intense industries reduces the gender gaps in wages. Interestingly, the gender gap in wages is reversed for women working in high-skilled occupations in knowledge intense industries in states such as Acre, Amapa, Amazonas, Espirito_Santo, Goias, Groso, Mato Grosso do Sul, Paraiba, Rondônia, Roraima or Tocantis. These are the economies that experienced a rapid development towards knowledge intense industries over the period examined. Rondônia ends up being one of the states with the highest share of employment in knowledge intense industries. The driving forces explaining the results for these states is the initial low supply of high-skilled workers, which characterizes the early stages of the process of economic development.

One could think of states such as Rio de Janeiro and São Paolo as having more knowledge intense industries. Indeed, if we look particularly at these two states, we observe that the share of individuals working in the industries in the top 10 percentiles of the distribution of knowledge intensity index is the highest in São Paulo, of 6.61% of total population, and the fourth highest in Rio de Janeiro, of 5.56%. However, both states have the largest dispersion of the index of knowledge intensity (with a standard deviation of 0.42), showing that both extremes of the knowledge intensity index are present. Our theory states that the gender gap in wages for high-skilled occupations would be reduced if companies are competing for high-skilled labor. But in Rio de Janeiro, the share of the high-skilled labor is of 6.50% of the total population, being the state with the highest share of high-skilled labor. Female workers are more skilled than male workers, and the share of high-skilled female workers in Rio de Janeiro is of 8.81%, being the highest in Brazil. São Paulo has a share of high-skilled labor of 4.98% of total working population, being the second highest share of skilled labor in the country. This means that in these two states companies are not constrained on skilled labor in the same way as in other states where the demand for high-skilled labor is positive and the availability of skilled labor is much more limited.

D. Robustness Checks

In line with the existing literature, we carry out two sets of robustness checks. First, we introduce firm level fixed effects. This accounts for firm specific unobservables, and reveals how much of the gender wage gap is due to differences in pay between firms and how much within firms. Second, we examine other labor market outcomes, such as female labor force participation, and check whether the gender gap is reduced in knowledge intense industries for workers in high-skilled occupations.

D.1.Firm Level Analysis

The economic literature argues for the important role of companies and their heterogeneity in wage setting (Sorkin 2017; Card et al. 2018). To assess how much of the gender gaps in wages remains within companies between observationally identical individuals, we introduce firm level fixed effects in Equation 3. Different treatment between companies explain an additional 13.81% of the gender wage gap in Rondônia and almost half of the remaining gender wage gap in states such as Piauí and Maranhão (Figure 4, Panel B). In Rio, the differences in treatment between companies explain a further 36.22% of the gender wage gap, reducing the unexplained gender gap from 11.60% to 8.10%. In the case of Minas Gerais, firm level fixed effects explain 27.15% of the gender wage gap reducing it from 15.140% to 11.00%. This leaves more than half of the residual gender wage gap documented in Subsection 5.A. unexplained. We find that the remaining gender wage gap within companies varies between 3.6% in Acre and 15.6% in Randônia (Figure 4, Panel A). Since the estimation of the relevant coefficients in Equation 10 relies on variations in knowledge intensity across industries, we cannot consistently estimate them using firm level fixed effects[11]. Introducing firm level fixed

[11] Specifically, we cannot estimate the coefficient associated with the knowledge intensity variable as firm level fixed effects and the value of knowledge intensity are collinear, as knowledge intensity varies little and only in few cases over the years.

effects would work only if there were no companies that hire only male or only female, but this is not the case. To go around this issue and to investigate how the gender gap in wages varies across companies of different knowledge intensity, we construct a rough measure of knowledge intensity at firm level[12]. In the existing literature on growth and human capital, average years of schooling are a common indicator to proxy knowledge (Benhabib and Spiegel 1994; Temple 1999; Barro 2000, 2001). Education has been shown to be a key determinant for economic growth, employment, and earnings in knowledge-based economies (Woessmann 2016). Across regions, across and within countries, Gennaioli et al. (2013) show that years of schooling determine differences in regional development. Years of schooling are also shown to account for cross-country differences in the levels of development (Hanushek and Woessmann 2012, 2015).

RAIS data offers information on levels of educational attainment of the employed population. Over the period 2006 – 2015 we have detailed information on the highest level of completed education within tertiary education. We therefore expand the factor variable of attained level of education into years of education and we calculate for each company the average level of years of education of the employed workforce over the entire period. We first calculate for all companies the average of years of education over the period 2006 – 2015. Then we normalize this measure to a standard normal [i.e. ~N(0,1)] using all companies in the Brazilian Economy. In this way we can interpret the coefficients as one standard deviation increase in the knowledge intensity measure across

[12] If we calculate the index of knowledge intensity as defined in Subsection B of Section 3 at firm level, this will only exacerbate the importance of company tenure as more time spent in the company would be rewarded more as the company needs specific human capital. It will also index low knowledge intensity for highly specialized firms that employ only few occupations. This measure would therefore reflect the size of the company (as small companies would have two few occupations to be knowledge intense). Therefore, we need to think of a more appropriate measure for the knowledge intensity at firm level.

states. Results equivalent to Table 6 are reported in Panel C of Figure 4. All results are consistent with our previous findings described in Section 5, Subsections A-C. In line with our theory, as the knowledge intensity increases, companies cannot afford suboptimal matching for high-skilled occupations, and thus they reduce gender discrimination, or any other type of discrimination based on characteristics not associated with productivity. The gender gap in wages is reduced for high-skilled occupations as companies become more knowledge intense, in line with the main findings reported in Section 5.C.

D.2. Labor Force Participation

Most of the existing literature on the link between economic development and gender disparities in labor market outcomes has focused on female labor force participation. We replicate our analysis by examining labor force participation. We estimate a linear probability model by regressing the female dummy on all the other covariates. This form of estimation gives us a simple measure of the reduction of the gender gap in labor force participation in high-skilled occupations in knowledge intense industries.

Results are reported in Panel D of Figure 4. We notice that with few exceptions, the gender gaps in labor force participation are reduced. For the six states where the coefficient is negative, for three of them the coefficient is not significantly different from zero (Maranhão, Piaui, and Tocantis). For Randônia and Amapá, the main analysis in our paper shows for these two states two the highest gender gap reversal in wages. This might imply that women have higher bargaining power in negotiating wages and time spent at work. We cannot account of self-selection in (out of) the workforce. In these two states, female working in high-skilled occupations in knowledge intense industries earn more than their male counterpart. It might be also because the labor force participation of women is much lower than that of men, and for key positions females are better qualified and can better negotiate their wages. These results are in line with the literature on the link between economic development and female labor force participation. Other robustness checks carried out with labor market outcomes such as full-time employment and hours worked are available from the Authors upon request.

6. Conclusion

Gender gaps in labor market outcomes decline with economic development. Yet, the underlying mechanisms are not fully understood. In this study, we provide evidence of a mechanism through which gender gaps in labor market outcomes change with economic development. Specifically, we focus on the expansion of knowledge intense industries and occupations, and the subsequent change in the demand for skilled labor. By relying on an employer-employee linked dataset that covers the entire formal labor force in Brazil, we find that employees in high-skilled occupations and industries experience lower gender wage gaps, and that the effect of knowledge intensity is stronger when the demand for skilled labor is high and the supply of skilled labor is low. Our analysis also shows that gender wage gap of skilled workers, but not that of unskilled workers, decreases when knowledge intense industries grow. These results are conditional on controlling for individual, occupation, sector, and location characteristics. Our results are robust to the use of a Bartik instrument based on labor demand shocks, to address endogeneity concerns.

Our study builds on the existing literature, which has studied the role that the expansion of the service sector has played in closing gender gaps in labor market outcomes. We examine the role of the expansion of knowledge intensity of each industry and each occupation within industries. We find that the expansion of knowledge intense industries contributes to the reduction of the gender wage gap.

This study also establishes a clear link between gender inequality in compensation and the knowledge intensity of an industry. For the first time in the literature focusing on inequality and knowledge intensity, we bring in the gender dimension and document a reduction of gender inequality in compensation for individuals employed in high-skilled occupations in knowledge intense industries. These results open an agenda for future research. We study an emerging economy, Brazil, which includes states at different stages of development. It will be important to study the relationship between knowledge intense activities and gender gaps in other countries as well. It is of paramount importance to understand what is the knowledge intensity level that needs to be reached for countries to start experiencing a reduction in gender differences in labor market outcomes, and how long it will take for countries at different stages of development to close these gaps.

Bibliography

Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40 (1): 7–72.

Acemoglu, Daron. 2012. "Introduction to Economic Growth." *Journal of Economic Theory* 147 (2): 545–50.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, 4:1043–1171. Elsevier.

Aghion, Philippe et al. 1998. *Endogenous Growth Theory*. MIT press.

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30(2): 175–87.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116(1): 313–50.

Arrow, Kenneth. 1973. Discrimination in Labor Markets *The Theory of Discrimination*. Princeton.

Audretsch, David B, and Maryann P Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *The American economic review* 86(3): 630–40.

Balassa, Bela, and Marcus Noland. 1989. "'Revealed' Comparative Advantage in Japan and the United States." *Journal of International Economic Integration*: 8–22.

Bandiera, Oriana et al. 2015. *Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa*. International Labor Office | Geneva.

Bandiera, Oriana, and Ashwini Natraj. 2013. "Does Gender Inequality Hinder Development and Economic Growth? Evidence and Policy Implications." *World Bank Research Observer* 28(1): 2–21.

Barro, Robert J. 2000. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth* 5(1): 5–32.

———. 2001. "Human Capital and Growth." *American Economic Review: Papers and Proceedings* 91(2): 12–17.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* WE Upjohn Institute for Employment Research.

Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Becker, Gary S., and H. Gregg Lewis. 1974. "Interaction between Quantity and Quality of Children." *Economics of the Family: Marriage, Children, and Human Capital* ISBN: 81–90.

Becker, Gary S. 1992. "The Economic Way of Looking at Life." *Nobel Lecture*.

Benhabib, Jess, and Mark M Spiegel. 1994. "The Role of Human Capital in Economic

Development Evidence from Aggregate Cross-Country Data." *Journal of Monetary Economics* 34(2): 143–73.

Ben Saâd, Myriam, and Giscard Assoumou-Ella. 2019. "Economic Complexity and Gender Inequality in Education: An Empirical Study." *Economics Bulletin* 39(1): 321–34.

Bergmann, Barbara R. 1974. "Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex." *Eastern Economic Journal* 1(2): 103–10.

Bertrand, Marianne. 2018. "Coase Lecture – The Glass Ceiling." *Economica* 85(338): 205–31.

Black, Sandra E., and Elizabeth Brainerd. 2004. "Importing Equality? The Impact of Globalization on Gender Discrimination." *Industrial and Labor Relations Review* 57(4): 540–59.

Black, Sandra E., and Philip E. Strahan. 2001. "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry." *American Economic Review* 91(4): 814–31.

Black, Sandra E, and Chinhui Juhn. 2000. "The Rise of Female Professionals: Are Women Responding to Skill Demand?" *American Economic Review* 90(2): 450–55.

Blau, Francine D., and Lawrence M. Kahn. 2006. "The U. S. Gender Pay Gap in the 1990s : Slowing Convergence." *Industrial and Labor Relations Review* 60(1): 45–66.

Blau, Francine D, Patricia Simpson, and Deborah Anderson. 1998. "Continuing Progress? Trends in Occupational Segregation in the United States over the 1970s and 1980s." *Feminist economics* 4(3): 29–71.

Blinder, Alan S. 1973. "Wage Discrimination : Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4): 436–55.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36(S1): S13--S70.

Card, David, ALexandre Mas, and Josse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics* 123(1): 177–218.

Cardoso, A. et al. 2007. "International Microdata Scoping Studies Project: Brazil." *Rio de Janeiro: Economic and Social Research Council (ESRC)* 19.

Cavalcanti, Tiago, and José Tavares. 2016. "The Output Cost of Gender Discrimination: A Model-Based Macroeconomics Estimate." *The Economic Journal* 126(590): 109–34.

Correia, Sergio. 2016. "Reghdfe: Estimating Linear Models with Multi-Way Fixed Effects." *Stata Conference*.

Coudin, Elise, Sophie Maillard, and Maxime Tô. 2018. "Family, Firms and the Gender Wage Gap in France." *IFS Working Paper* W18/01.

Cuberes, David, and Marc Teignier. 2014. "Gender Inequality and Economic Growth: A Critical Review." *Journal of International Development* 26: 260–76.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review* 106(3): 479–524.

Dollar, David, and Roberta Gatti. 1999. "Gender Inequality, Income, and Growth: Are Good Times Good for Women?" *Washington, DC: Development Research Group, The World Bank.*

Dosi, Giovanni et al. 1988. *Technical Change and Economic Theory*.

Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature* 50(4): 1051–79.

Elborgh-Woytek, Katrin et al. 2014. *Women, Work, and the Economy: Macroeconomic Gains from Gender Equity*.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics* 111(4): 1007–47.

Fleming, Lee, and Olav Sorenson. 2001. "Technology as a Complex Adaptive System: Evidence from Patent Data." *Research Policy* 30(7): 1019–39.

Forbes, Kristin J. 2000. "A Reassessment of the Relationship between Inequality and Growth." *American Economic Review* 90(4): 869–87.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition Methods in Economics." *Handbook of Labor Economics* 4(PART A): 1–102.

Gaddis, Isis, and Stephan Klasen. 2014. "Economic Development, Structural Change, and Women's Labor Force Participation." *Journal of Population Economics* 27(3): 639–81.

Gaddis, Isis, and Janneke Pieters. 2012. "Trade Liberalization and Female Labor Force Participation: Evidence from Brazil."

———. 2017. "The Gendered Labor Market Impacts of Trade Liberalization Evidence from Brazil." *Journal of Human Resources* 52(2): 457–90.

Galor, Oded, and David N. Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86(3): 374–87.

Galor, Oded, and David N Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90(4): 806–28.

Ganguli, Ina, Ricardo Hausmann, and Martina Viarengo. 2014. "Closing the Gender Gap in Education: What Is the State of Gaps in Labor Force Participation for Women, Wives and Mothers?" *International Labour Review* 153(2): 173–207.

Garas, Antonios, and Athanasios Lapatinas. 2017. "The Role of Consumer Networks in Firms' Multi-Characteristics Competition and Market Share Inequality." *Structural Change and Economic Dynamics* 43: 76–86.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics* 128(1): 105–64.

Goldberg, Matthew S. 1982. "Discrimination, Nepotism, and Long-Run Wage

Differentials." *Quarterly Journal of Economics* 97(2): 307–19.

Goldin, Claudia. 1999. "The gender gap: An economic history of American women." Cambridge UniversityPress, New York, United States.

———. 1991. "The Role of World War II in the Rise of Women's Employment." *American Economic Review*: 741–56.

———. 1994. "Understanding the Gender Gap: An Economic History of American Women." In *Equal Employment Opportunity: Labor Market Discrimination and Public Policy,* ed. Paul Burstein. Aldine de Gruyter New York, 17–26.

———. 1995. "The U-Shaped Female Labor Force Function in Economic Development and Economic History." In *Investment in Women's Human Capital and Economic Development.,* ed. T. P. Schultz. University of Chicago Press, 61–90.

———. 2000. "Labor Markets in the Twentieth Century." In *The Cambridge Economic History of the United States, III,* eds. Stanley L. Engerman and Robert E. Gallman. Cambridge: Cambridge University Press, 549–623.

———. 2005. "From the Valley to the Summit: A Brief History of the Quiet Revolution That Transformed Women's Work." *Regional Review* Q1: 5–12.

———. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96(2): 1–21.

Goldin, Claudia, and Lawrence F Katz. 2002. "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy* 110(4): 730–70.

———. 2016. "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation." *Journal of Labor Economics* 34(3): 705–46.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20(4): 133–56.

Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu. 2005. "Engines of Liberation." *Review of Economic Studies* 72(1): 109–33.

Hanushek, Eric A, and Ludger Woessmann. 2012. "Schooling, Educational Achievement, and the Latin American Growth Puzzle." *Journal of Development Economics* 99(2): 497–512.

———. 2015. *The Knowledge Capital of Nations: Education and the Economics of Growth*. MIT press.

Hartmann, Dominik et al. 2017. "Linking Economic Complexity, Institutions, and Income Inequality." *World Development* 93: 75–93.

Hausmann, Ricardo et al. 2014. *The Atlas of Economic Complexity, Mapping Paths to Prosperity*. MIT Press.

Heckman, James. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12(2): 101–16.

Hidalgo, César A., and Ricardo Hausmann. 2009. "The Building Blocks of Economic Complexity." *Proceedings of the National Academy of Sciences* 106(26): 10570–75.

Hill, M Anne, and Elizabeth King. 1995. "Women's Education and Economic Well-Being." *Feminist Economics* 1(2): 21–46.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *the Quarterly journal of Economics* 108(3): 577–98.

Jayachandran, Seema. 2019. "Social Norms as a Barrier to Women's Employment in Developing Countries." *Northwestern Working Paper*.

Johnson, George E., and Frank P. Stafford. 1995. "A Model of Occupational Choice with Distributions of Relative Abilities." *Mimeo.(Department of Economics, University of Michigan, Ann Arbor, MI)*.

———. 1998. *Alternative Approaches to Occupational Exclusion*. Routledge Research in Gender and Society.

Killingsworth, Mark R., and James J. Heckman. 1986. "Female Labor Supply: A Survey." *Handbook of Labor Economics* 1(C): 103–204.

Knowles, Stephen, Paula K Lorgelly, and P Dorian Owen. 2002. "Are Educational Gender Gaps a Brake on Economic Development? Some Cross-Country Empirical Evidence." *Oxford Economic Papers* 54(1): 118–49.

Lagerlöf, Nils-Petter. 2003. "Gender Equality and Long-Run Growth." *Journal of Economic Growth* 8(4): 403–26.

Luci, Angela. 2009. "Female Labour Market Participation and Economic Growth." *International Journal of Innovation and Sustainable Development* 4(2/3): 97–108.

Lundberg, Shelly J., and Richard Startz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Market." *The American Economic Review* 73(3): 340–47.

Lundberg, Shelly J. 1991. "The Enforcement of Equal Opportunity Laws under Imperfect Information: Affirmative Action and Alternatives." *Quarterly Journal of Economics* 106(1): 309–26.

Mammen, Kristin, and Christina Paxson. 2000. "Women's Work and Economic Development." *Journal of Economic Perspectives* 14(4): 141–64.

Miller, Mark M, Lay James Gibson, and N Gene Wright. 1991. "Location Quotient: A Basic Tool for Economic Development Analysis." *Economic Development Review* 9(2): 65.

Neagu, Olimpia, and Mircea Constantin Teodoru. 2019. "The Relationship between Economic Complexity, Energy Consumption Structure and Greenhouse Gas Emission: Heterogeneous Panel Evidence from the EU Countries." *Sustainability* 11(2): 497.

Nelson, Richard R, and Sidney G Winter. 1982. "The Schumpeterian Tradeoff Revisited." *American Economic Review* 72(1): 114–32.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3): 693–709.

Olivetti, Claudia. 2014. "The Female Labor Force and Long-Run Development:The

American Experience in Comparative Perspective.” *Human Capital in History: The American Record*: 161.

Olivetti, Claudia, and Barbara Petrongolo. 2016. “The Evolution of Gender Gaps in Industrialized Countries.” *Annual Review of Economics* 8(1): 405–34.

Pan, Jessica. 2015. “Gender Segregation in Occupations: The Role of Tipping and Social Interactions.” *Journal of Labor Economics* 33(2): 365–408.

Petrongolo, Barbara, and Maddalena Ronchi. 2020. "Gender gaps and the structure of local labor markets." *Labour Economics* 64: 101819.

Phelps, Edmund S. 1972. “The Statistical Theory of Racism and Sexism.” *American Economic Review* 62(4): 659–61.

Romer, Paul M. 1990. “Endogenous Technological Change.” *Journal of Political Economy* 98(5, Part 2): S71--S102.

Romero, João P, and Camila Gramkow. 2020. “Economic Complexity and Greenhouse Gas Emission Intensity.”

Sinha, Jania N. 1967. “Dynamics of Female Participation in Economic Activity in a Developing Economy.” In *Proceedings of the World Population Conference. Belgrade, 1965*: 336–37.

Sorkin, Isaac. 2017. “The Role of Firms in Gender Earnings Inequality: Evidence from the United States.” *American Economic Review* 107(5): 384–87.

Stojkoski, Viktor, Zoran Utkovski, and Ljupco Kocarev. 2016. “The Impact of Services on Economic Complexity: Service Sophistication as Route for Economic Growth.” *PloS one* 11(8): e0161633.

Tamura, Robert. 2006. “Human Capital and Economic Development.” *Journal of Development Economics* 79(1): 26–72.

Temple, Jonathan. 1999. “The New Growth Evidence.” *Journal of Economic Literature* 37(1): 112–56.

Tzannatos, Zafiris. 1999. “Women and Labor Market Changes in the Global Economy: Growth Helps, Inequalities Hurt and Public Policy Matters.” *World development* 27(3): 551–69.

Verick, Sher. 2014. “Female Labor Force Participation in Developing Countries.” *IZA World of Labor*.

Woessmann, Ludger. 2016. “The Economic Case for Education.” *Education Economics* 24(1): 3–32.

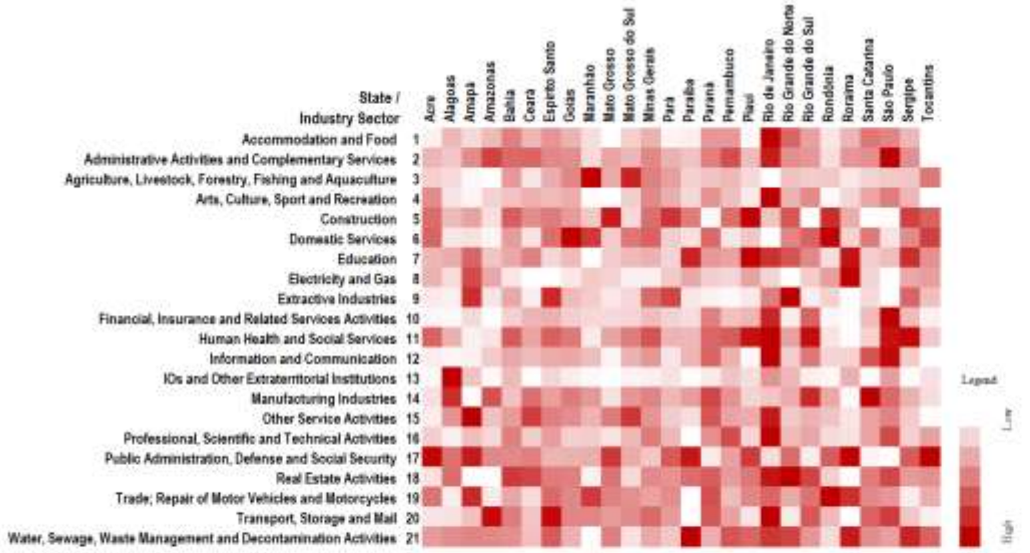
Wooldridge, Jeffrey M. 2016. *Introductory Econometrics: A Modern Approach*. Nelson Education.

World Bank. 2012. *World Development Report 2012: Gender Equality and Development*. The World Bank Washington, DC.

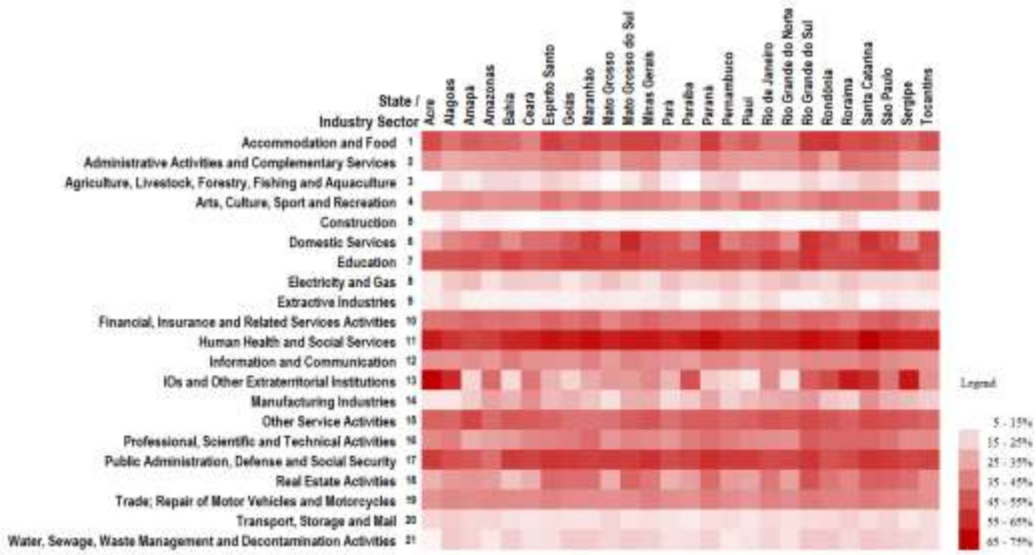
Figures and Tables

Figure 1. Brazilian States Patterns

A. Economic Sector Composition



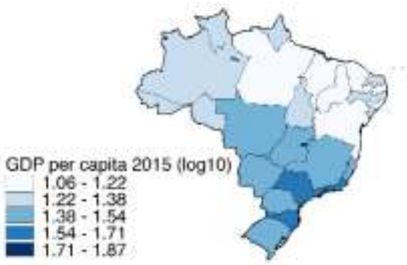
A. Gender Segregation by Economic Sector



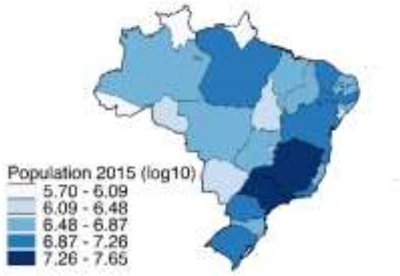
Note: A. *Shows the heterogeneity between states in terms of sector composition. Color intensities are calculated across states within economic sectors. Each square represents the share of employment in that specific sector. The intensity of the color compares the employment share by sectors across states.* ***B.*** *Shows the gender segregation by sectors within states in Brazil. Color intensity varies across sectors within state. Each square represents the female employment share by sector.*

Figure 2. Structural Differences across States

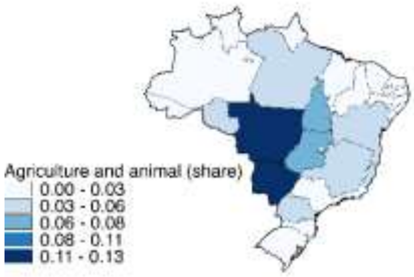
A. GDP per Capita



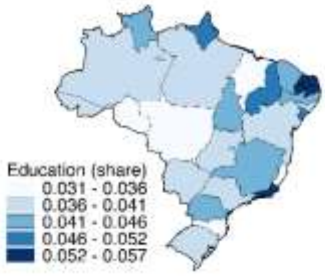
B. Population



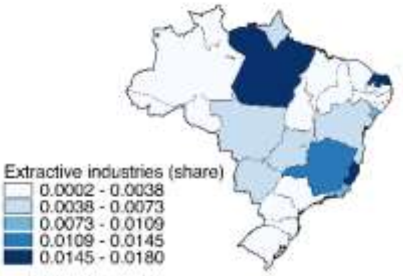
C. Agriculture



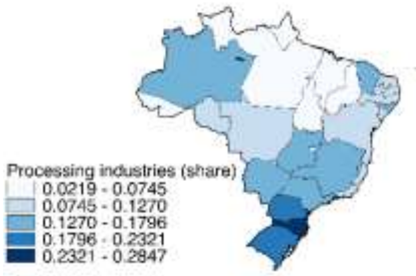
D. Education



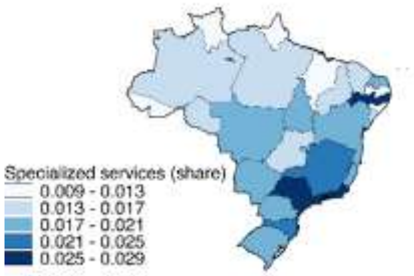
E. Extractive Industries



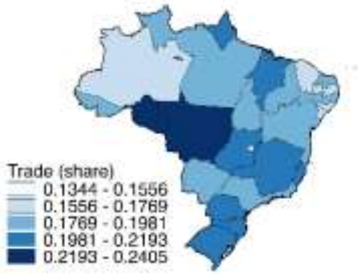
F. Processing Industries



G. Specialized Services

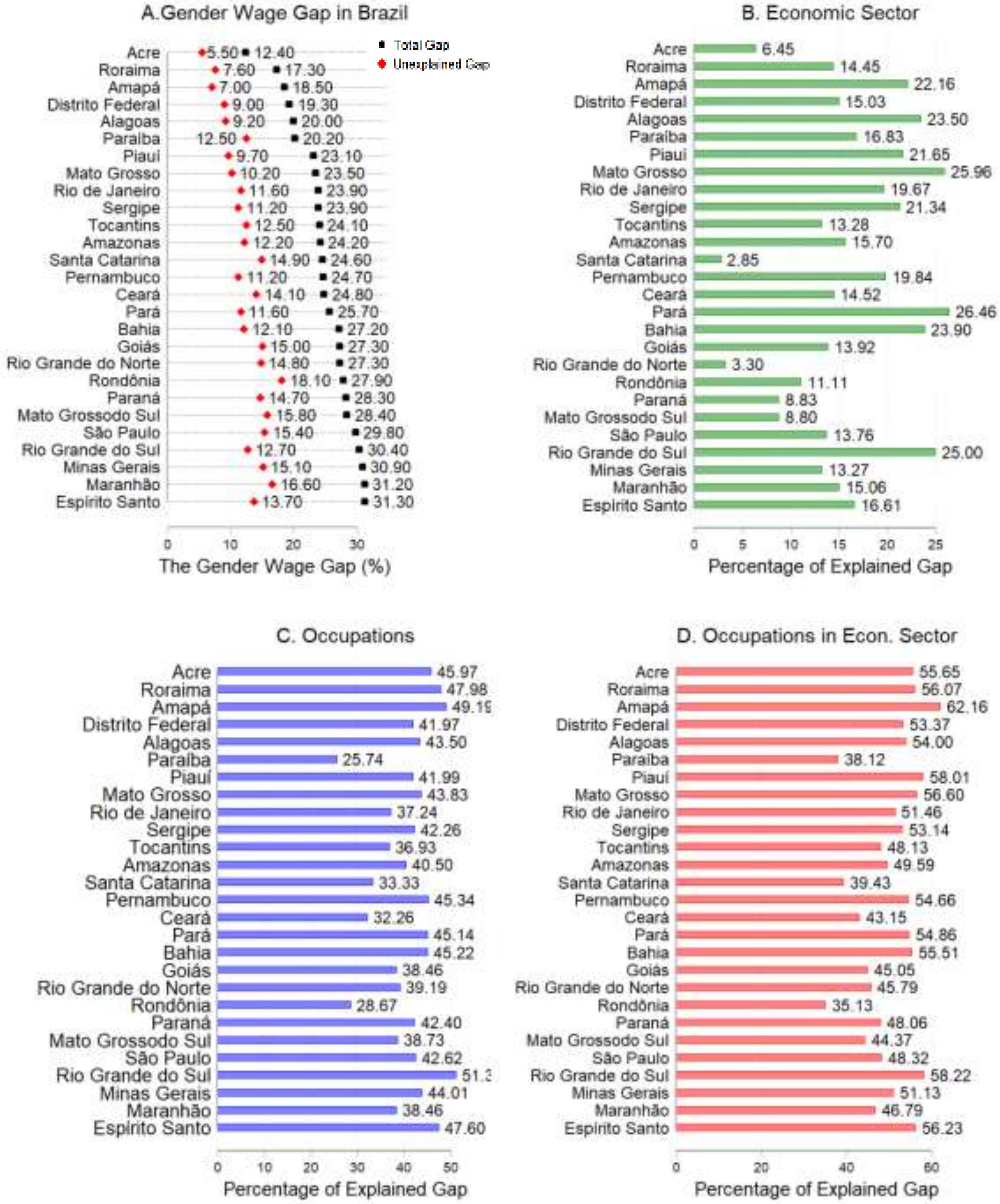


H. Trade Industries



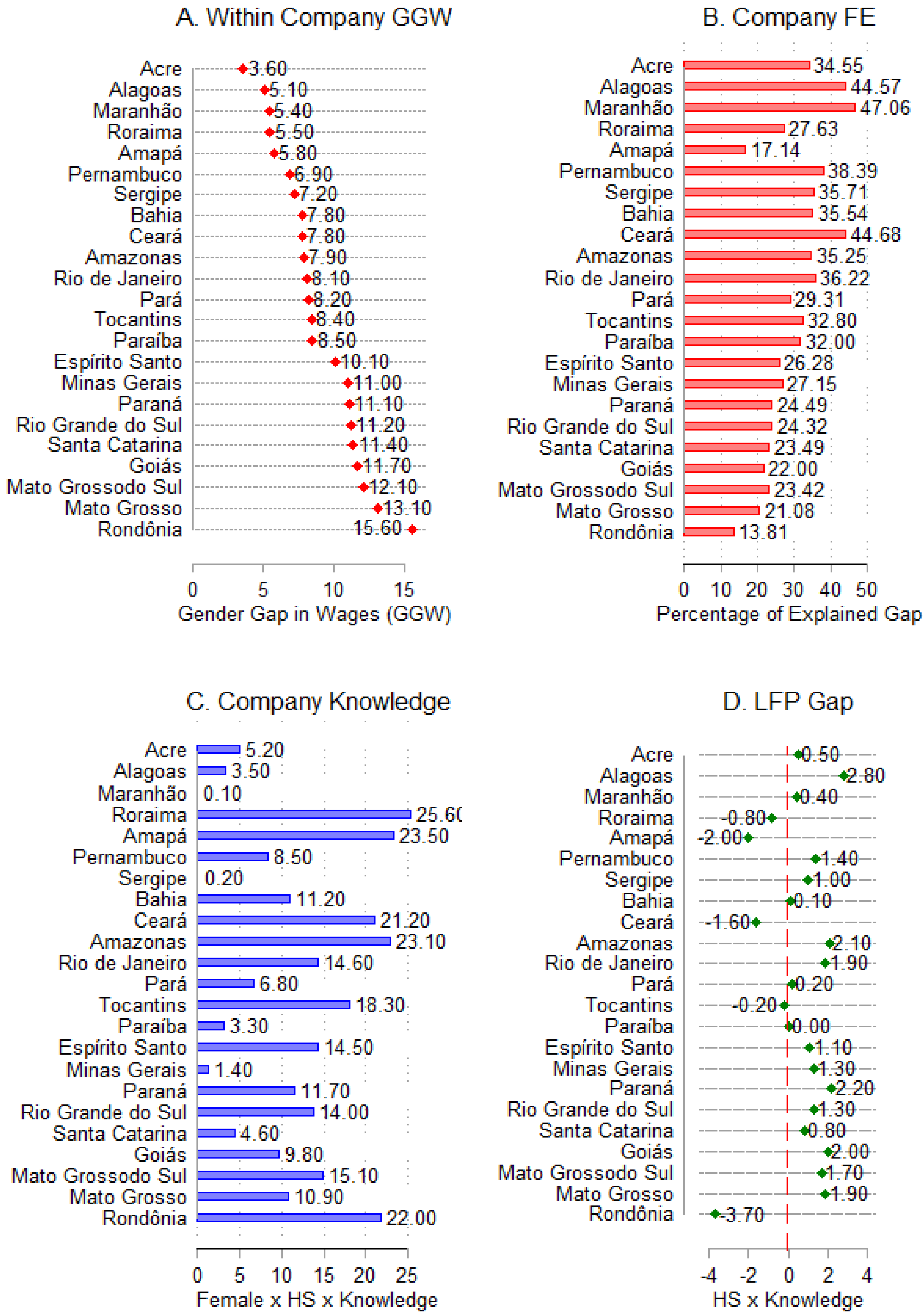
Note: *Data for panels A and B come from the 2015 Census Estimates from the Brazilian Institute of Geography and Statistics. Panels C-H use RAIS data from the year 2015*

Figure 3. Decomposition of the Gender Gap in Wages



__Note:__ Panel A presents the gender gap in wages conditional on individual characteristics by state (black dots) and the unexplained gender gap in wages within occupations within economic sectors conditional on individual characteristics by state (red diamonds). The difference between the unconditional and the unexplained is the part of the gender gap in wages explained by the gender segregation within occupations within sectors. Panel B presents by state how much of the total gender gap in wages is explained by gender differences in wages between economic sectors. Panel C presents by state how much of the total gender gap in wages is explained by gender differences in wages between occupations. Panel D presents how much of the gender gap in wages is explained by segregation within occupations within economic sectors (how much of the total gender gap in wages is explained by differences between occupations within sectors).

Figure 4. Robustness Checks Results



Note: *Please refer to notes of Tables 4-6 for the specifications of the data used to produce the above estimates. Panel A presents the residual gender gap in wages within company (Estimates of the gender coefficient in Equation 3 including company fixed effects). Panel B explains how much of the gender gap is further on explained by the differences between companies. Panel C presents the results of the estimate of the coefficient of female x high-skilled occupations x knowledge intensity term in Equation 9 when the knowledge intensity is calculated at firm level. Panel D presents results for Labor Force Participation analysis. The full set of results is available from the Authors upon request.*

Table 1. Distribution of the Knowledge Intensity Measure

	Mean	Median	S.D.	Min	Max	75 percentiles	90 percentiles
Knowledge Intensity Measure	1.09	1.07	0.42	0.12	2.37	1.39	1.61
Normalized Measure of Knowledge Intensity	0.00	-0.04	1.00	-2.29	3.04	0.72	1.25

Table 2. Top and Bottom Knowledge Intense Industries

Industries	**Normalized Index of Knowledge Intensity**
In the bottom 10 percentile of the index	
Surveillance Activities, Security and Research	-2.25
Mail Delivery and Other Activities	-1.84
Food	-1.81
Construction	-1.75
Veterinary Activities	-1.47
In the top 10 percentile of the index	
Manufacture of Information Technology Equipment	1.28
Manufacture of Chemical Products	1.38
Technical Equipment Manufacturing, Installation and Service	1.63
Asset Management	1.92
Architecture and Engineering Services, Technical Analysis	2.5

Note: *In top 10% of the index of knowledge intense industries we also have Scientific Research and Development, Business Consulting, Manufacturing of other electrical equipment, medical equipment, and computer accessories. In the bottom 10% of the index we have industries such as Clothing manufacturing, terrestrial transportation, agricultural activities, domestic activities.*

Table 3. Descriptive Statistics of Main Variables by State

State	Total Number of Observations	Share of Female Workers	Share of High-Skilled Labor			Growth Rate of the Share of High-Skilled Individuals		
			Male	Female	Total	Male	Female	Total
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Acre	1,797,718	*44.60*	*1.21*	*1.66*	*1.36*	*-3.13*	*-1.16*	*-2.50*
Alagoas	7,571,239	*35.30*	*1.60*	*2.62*	*1.83*	*6.16*	*5.04*	*5.90*
Amapa	1,719,914	*41.60*	*1.13*	*5.03*	*1.97*	*7.05*	*15.78*	*8.87*
Amazonas	9,431,379	*37.00*	*2.29*	*4.34*	*2.84*	*3.25*	*3.38*	*3.28*
Bahia	35,083,280	*39.80*	*3.23*	*6.12*	*3.84*	*1.19*	*0.75*	*1.10*
Ceara	21,602,420	*41.80*	*2.27*	*4.47*	*2.81*	*-2.90*	*-3.01*	*-2.92*
Espirito Santo	15,969,110	*39.90*	*2.83*	*5.73*	*3.36*	*4.65*	*4.73*	*4.67*
Goias	23,890,220	*38.80*	*2.30*	*3.94*	*2.70*	*3.49*	*3.34*	*3.45*
Grosso	13,718,200	*35.40*	*1.79*	*3.54*	*2.20*	*3.41*	*4.14*	*3.59*
Maranhao	9,658,814	*39.70*	*2.12*	*4.88*	*2.57*	*6.18*	*6.92*	*6.30*
Mato Grosso do Sul	10,529,180	*38.30*	*1.85*	*2.77*	*2.08*	*3.50*	*3.75*	*3.56*
Minas Gerais	84,676,450	*40.10*	*3.60*	*5.56*	*4.07*	*2.17*	*2.69*	*2.30*
Para	15,784,850	*36.30*	*1.77*	*4.72*	*2.26*	*4.22*	*3.74*	*4.14*
Paraiba	8,689,095	*40.60*	*1.26*	*3.00*	*1.63*	*2.45*	*2.78*	*2.52*
Parana	50,864,130	*41.80*	*2.75*	*4.21*	*3.14*	*2.95*	*3.14*	*3.00*
Pernambuco	24,895,640	*38.20*	*3.03*	*6.51*	*3.88*	*2.38*	*1.73*	*2.22*
Piaui	5,775,740	*41.50*	*2.50*	*3.99*	*2.83*	*3.76*	*2.48*	*3.48*
Rio de Janeiro	69,294,990	*40.40*	*5.68*	*8.81*	*6.50*	*2.01*	*1.75*	*1.94*
Rio Grande do Norte	9,218,621	*39.80*	*1.89*	*3.71*	*2.26*	*9.42*	*14.25*	*10.39*
Rio Grande do Sul	49,881,680	*43.40*	*2.46*	*3.79*	*2.82*	*1.80*	*2.03*	*1.86*
Rondonia	5,498,377	*39.40*	*1.75*	*2.26*	*1.88*	*3.54*	*3.76*	*3.60*
Roraima	1,109,034	*46.40*	*1.82*	*1.63*	*1.76*	*8.85*	*6.46*	*8.03*
Santa Catarina	36,915,480	*43.20*	*2.59*	*3.46*	*2.82*	*2.70*	*2.81*	*2.73*
Sao Paolo	228,584,400	*41.00*	*4.42*	*6.52*	*4.98*	*1.93*	*1.82*	*1.91*
Serpige	5,838,900	*39.60*	*2.30*	*3.63*	*2.61*	*-4.25*	*-5.23*	*-4.48*
Tocantins	3,930,338	*40.60*	*1.47*	*2.99*	*1.76*	*1.67*	*0.72*	*1.49*

Note: *Calculations carried out by the Authors using RAIS data between 2003 and 2015. Here and thorough the analysis in this paper we refer only to the active labor force with legal working age, i.e. 16-65 y.o.*

Table 4. Determinants of Gender Wage Differences in Brazil

State	Dependent variable: Log of Average Monthly Wage in 2010 Real				
	(1)	(2)	(3)	(4)	(5)
Acre	-0.124***	-0.116***	-0.067***	-0.064***	-0.055***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Alagoas	-0.200***	-0.153***	-0.113***	-0.101***	-0.092***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Amapá	-0.185***	-0.144***	-0.094***	-0.087***	-0.070***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Amazonas	-0.242***	-0.204***	-0.144***	-0.139***	-0.122***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Bahia	-0.272***	-0.207***	-0.149***	-0.134***	-0.121***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ceará	-0.248***	-0.212***	-0.168***	-0.156***	-0.141***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Distrito Federal	-0.193***	-0.164***	-0.112***	-0.104***	-0.090***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Espírito Santo	-0.313***	-0.261***	-0.164***	-0.152***	-0.137***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Goiás	-0.273***	-0.235***	-0.168***	-0.162***	-0.150***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Grosso	-0.312***	-0.265***	-0.192***	-0.185***	-0.166***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Maranhão	-0.235***	-0.174***	-0.132***	-0.121***	-0.102***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mato Grosso do Sul	-0.284***	-0.259***	-0.174***	-0.168***	-0.158***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Minas Gerais	-0.309***	-0.268***	-0.173***	-0.165***	-0.151***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pará	-0.257***	-0.189***	-0.141***	-0.126***	-0.116***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Paraíba	-0.202***	-0.168***	-0.150***	-0.132***	-0.125***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Paraná	-0.283***	-0.258***	-0.163***	-0.157***	-0.147***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pernambuco	-0.247***	-0.198***	-0.135***	-0.123***	-0.112***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Piauí	-0.231***	-0.181***	-0.134***	-0.113***	-0.097***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rio de Janeiro	-0.304***	-0.228***	-0.148***	-0.139***	-0.127***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Rio Grande do Norte	-0.239***	-0.192***	-0.150***	-0.136***	-0.116***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rio Grande do Sul	-0.273***	-0.264***	-0.166***	-0.164***	-0.148***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Rondônia	-0.279***	-0.248***	-0.199***	-0.196***	-0.181***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Roraima	-0.173***	-0.148***	-0.090***	-0.082***	-0.076***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Santa Catarina	-0.246***	-0.239***	-0.164***	-0.161***	-0.149***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
São Paulo	-0.298***	-0.257***	-0.171***	-0.163***	-0.154***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Serpige	-0.239***	-0.188***	-0.138***	-0.125***	-0.112***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tocantins	-0.241***	-0.209***	-0.152***	-0.141***	-0.125***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Individual Level Covariates	✓	✓	✓	✓	✓
Economic Sector FE	-	✓	-	✓	-
Occupation FE	-	-	✓	✓	-
Economic Sector- Occupation FE	-	-	-	-	✓
Year FE	✓	✓	✓	✓	✓

Note: *The results reported in this table and in the subsequent tables use the full sample dataset for each reported state between 2003 and 2015 for the legal working age population, i.e. 16 – 65 y.o. We report for each state the estimated coefficient associated with the dummy variable for gender. All results are obtained with linear regression models using fixed effects for* *the specified variables. We employ the* Correia (2016) *proposed methodology for the multi-way fixed effects in column 5. Each regression from left to right differ in the levels of fixed effects as reported in the bottom of the table. All regressions include the following individual level covariates: age, age-squared, level of educational attainment, ethnic background, hours worked, leaves of absence, and company tenure. Age is measured in years. Leaves of absence is the number of days missed from work in each year between 2003 and 2015. Company tenure is a variable that represents the years spent by an individual within the same company (defined as continuous, based on the number of weeks worked within the same company divided by 12). Individual covariates include indicators for the level of educational attainment: basic education 1 incomplete; basic education 1 completed; basic education 2 incomplete; basic education 2 completed; high school incomplete; high school completed; college incomplete; college education completed; indicators for the seven groups of the ethnic background: White, Indian, Black, Asian, multiracial, unidentified, and unreported, and indicators for the reason of work interruption (distinct types of medical reasons, voluntary leave, etc.) Economic sectors refer to the 21 sectors of the economy as defined by the Ministry of Economy in Brazil. Occupations fixed effects uses the four-digit occupations codes as defined by the Brazilian Occupational Codes from the Ministry of Labor in Brazil. Standard errors are in parenthesis. We calculate them using a two-way clustering procedure at individual and company level.*

*Significance level. * p <0.1; ** p<0.05; *** p<0.01.*

Table 5. The Role of Knowledge Intensity of Economic Activity by State

	Dependent variable: Log of Average Monthly Wage in 2010 Real						
	Acre	Alagoas	Amapá	Amazonas	Bahia	Ceará	Espírito Santo
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.078***	-0.120***	-0.110***	-0.159***	-0.162***	-0.172***	-0.194***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
High-Skilled Occupation (HS)	0.700***	0.471***	0.621***	0.565***	0.595***	0.626***	0.578***
	(0.003)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)
Knowledge Intensity (K)	0.031***	0.035***	0.069***	0.053***	0.069***	0.037***	0.091***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Female x HS	-0.039***	-0.011***	-0.034***	-0.058***	-0.040***	-0.142***	-0.021***
	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Female x K	-0.011***	-0.023***	-0.027***	-0.019***	-0.030***	-0.039***	-0.056***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Female x HS x K	0.088***	0.026***	0.144***	0.083***	0.028***	0.060***	0.089***
	(0.004)	(0.002)	(0.004)	(0.002)	(0.001)	(0.001)	(0.001)

	Goiás	Grosso	Maranhão	Mato Grosso do Sul	Minas Gerais	Pará	Paraíba
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Female	-0.194***	-0.228***	-0.141***	-0.194***	-0.197***	-0.150***	-0.144***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
High-Skilled Occupation (HS)	0.579***	0.566***	0.598***	0.617***	0.570***	0.484***	0.602***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)
Knowledge Intensity (K)	0.031***	0.043***	0.070***	0.031***	0.066***	0.078***	0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female x HS	-0.029***	0.017***	-0.117***	-0.028***	-0.058***	-0.037***	-0.055***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)
Female x K	-0.010***	-0.009***	-0.014***	-0.010***	-0.039***	-0.035***	-0.007***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Female x HS x K	0.080***	0.063***	0.013***	0.096***	0.085***	0.053***	0.105***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.001)	(0.002)
Individual Level Covariates	✓	✓	✓	✓	✓	✓	✓
Economic Sector FE	✓	✓	✓	✓	✓	✓	✓
Occupation FE	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓

Table 5 – (cont'd)

	Dependent variable: Log of Average Monthly Wage in 2010 Real					
	Paraná	**Pernambuco**	**Piauí**	**Rio de Janeiro**	**Rio Grande do Norte**	**Rio Grande do Sul**
	(15)	(16)	(17)	(18)	(19)	(20)
Female	-0.187***	-0.147***	-0.129***	-0.157***	-0.153***	-0.196***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
High-Skilled Occupation (HS)	0.600***	0.706***	0.512***	0.423***	0.490***	0.615***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.002)	(0.001)
Knowledge Intensity (K)	0.060***	0.059***	0.026***	0.054***	0.053***	0.056***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female x HS	-0.033***	-0.053***	-0.105***	-0.086***	-0.065***	-0.007***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)
Female x K	-0.031***	-0.035***	0.002**	-0.027***	-0.024***	-0.030***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Female x HS x K	0.049***	0.062***	-0.055***	0.060***	0.078***	0.029***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.002)	(0.001)

	Rondônia	**Roraima**	**Santa Catarina**	**São Paulo**	**Serpige**	**Tocantins**
	(21)	(22)	(23)	(24)	(25)	(26)
Female	-0.229***	-0.103***	-0.181***	-0.179***	-0.147***	-0.174***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
High-Skilled Occupation (HS)	0.530***	0.699***	0.531***	0.556***	0.725***	0.639***
	(0.002)	(0.004)	(0.001)	(0.001)	(0.002)	(0.002)
Knowledge Intensity (K)	0.068***	0.042***	0.051***	0.042***	0.029***	0.061***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Female x HS	-0.009***	-0.007*	-0.066***	-0.069***	-0.114***	0.032***
	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)
Female x K	-0.055***	0.000	-0.028***	-0.012***	-0.012***	-0.025***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Female x HS x K	0.221***	0.179***	0.062***	0.004***	0.000	0.159***
	(0.003)	(0.006)	(0.001)	(0.001)	(0.002)	(0.003)
Individual Level Covariates	✓	✓	✓	✓	✓	✓
Economic Sector FE	✓	✓	✓	✓	✓	✓
Occupation FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Note: *Please refer to notes of Table 4 for sample and variable descriptions. All regressions include the following individual level covariates: age, age-squared, indicators for the level of educational attainment, indicators for ethnic background, hours worked per week, leaves of absence, and company tenure. The knowledge intensity measure is constructed based on combination of occupations. This index is normalized to a standard normal with mean zero and standard deviation 1. Each estimate includes the annual growth rate of the share of high-skilled occupations by micro-region as a measure of competition for high-skilled occupations. Because of computational reasons, for the state of São Paulo the estimates are obtained with dummies for the specified fixed effects instead of Correia (2016) methodology. Standard errors are in parenthesis, clustered at individual and company level. Significance level. * p <0.1; ** p<0.05; *** p<0.01.*

Table 6. Endogenous Labor Market Forces

	Dependent variable: Log of Average Monthly Wage in 2010 Real						
	Acre	**Alagoas**	**Amapá**	**Amazonas**	**Bahia**	**Ceará**	**Espírito Santo**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.077***	-0.120***	-0.109***	-0.105***	-0.162***	-0.111***	-0.192***
	(0.001)	(0.001)	(0.002)	(0.006)	(0.000)	(0.005)	(0.001)
High-Skilled Occupation (HS)	0.694***	0.469***	0.654***	0.739***	0.624***	0.513***	0.581***
	(0.010)	(0.005)	(0.013)	(0.022)	(0.002)	(0.013)	(0.003)
Knowledge Intensity (K)	0.030***	0.035***	0.060***	0.036***	0.062***	-0.019***	0.094***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.004)	(0.000)
Female x HS	-0.040***	-0.009*	-0.025*	-0.020	-0.039***	-0.106***	-0.023***
	(0.007)	(0.004)	(0.010)	(0.011)	(0.002)	(0.009)	(0.003)
Female x K	-0.011***	-0.023***	-0.022***	-0.005*	-0.032***	-0.029***	-0.055***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)	(0.002)	(0.001)
Female x HS x K	0.085***	0.028***	0.179***	0.096***	0.030***	0.008	0.088***
	(0.009)	(0.005)	(0.012)	(0.008)	(0.002)	(0.007)	(0.003)
	Goiás	**Grosso**	**Maranhão**	**Mato Grosso do Sul**	**Minas Gerais**	**Pará**	**Paraíba**
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Female	-0.194***	-0.228***	-0.141***	-0.194***	-0.200***	-0.150***	-0.145***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
High-Skilled Occupation (HS)	0.573***	0.565***	0.661***	0.617***	0.569***	0.484***	0.601***
	(0.003)	(0.004)	(0.005)	(0.004)	(0.002)	(0.004)	(0.006)
Knowledge Intensity (K)	0.025***	0.043***	0.067***	0.031***	0.065***	0.078***	0.020***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Female x HS	-0.042***	0.017***	-0.096***	-0.028***	-0.058***	-0.037***	-0.051***
	(0.003)	(0.004)	(0.005)	(0.004)	(0.002)	(0.004)	(0.005)
Female x K	-0.012***	-0.009***	-0.013***	-0.010***	-0.037***	-0.035***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Female x HS x K	0.073***	0.063***	0.017***	0.096***	0.093***	0.053***	0.108***
	(0.003)	(0.003)	(0.005)	(0.004)	(0.001)	(0.003)	(0.005)
Individual Level Covariates	✓	✓	✓	✓	✓	✓	✓
Economic Sector FE	✓	✓	✓	✓	✓	✓	✓
Occupation FE	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓

Table 6 – (cont'd)

	Dependent variable: Log of Average Monthly Wage in 2010 Real					
	Paraná	**Pernambuco**	**Piauí**	**Rio de Janeiro**	**Rio Grande do Norte**	**Rio Grande do Sul**
	(15)	(16)	(17)	(18)	(19)	(20)
Female	-0.187***	-0.148***	-0.111***	-0.157***	-0.159***	-0.196***
	(0.000)	(0.001)	(0.002)	(0.000)	(0.001)	(0.000)
High-Skilled Occupation (HS)	0.605***	0.708***	0.662***	0.422***	0.434***	0.615***
	(0.002)	(0.003)	(0.012)	(0.001)	(0.006)	(0.002)
Knowledge Intensity (K)	0.061***	0.059***	0.027***	0.052***	0.048***	0.056***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Female x HS	-0.033***	-0.050***	-0.013	-0.086***	-0.100***	-0.007***
	(0.002)	(0.003)	(0.009)	(0.001)	(0.005)	(0.002)
Female x K	-0.031***	-0.036***	0.010***	-0.027***	-0.018***	-0.030***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)
Female x HS x K	0.053***	0.064***	-0.029***	0.060***	0.021***	0.029***
	(0.001)	(0.002)	(0.008)	(0.001)	(0.005)	(0.001)
	Rondônia	**Roraima**	**Santa Catarina**	**São Paulo**	**Serpige**	**Tocantins**
	(21)	(22)	(23)	(24)	(25)	(26)
Female	-0.230***	-0.103***	-0.180***	-0.171***	-0.141***	-0.174***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
High-Skilled Occupation (HS)	0.529***	0.703***	0.534***	0.553***	0.741***	0.643***
	(0.006)	(0.011)	(0.002)	(0.001)	(0.006)	(0.005)
Knowledge Intensity (K)	0.068***	0.042***	0.051***	0.046***	0.016***	0.062***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
Female x HS	-0.009	-0.007	-0.066***	-0.077***	-0.109***	0.032***
	(0.006)	(0.008)	(0.002)	(0.001)	(0.006)	(0.004)
Female x K	-0.054***	0.002	-0.027***	-0.021***	-0.007***	-0.025***
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
Female x HS x K	0.224***	0.177***	0.060***	0.006***	-0.013*	0.161***
	(0.006)	(0.014)	(0.002)	(0.001)	(0.005)	(0.007)
Individual Level Covariates	✓	✓	✓	✓	✓	✓
Economic Sector FE	✓	✓	✓	✓	✓	✓
Occupation FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Note: *Please refer to notes of Table 4 for sample and variable descriptions. All regressions include the following individual level covariates: age, age-squared, indicators for the level of educational attainment, indicators for ethnic background, hours worked per week, leaves of absence, and company tenure. Regressions include knowledge intensity measure as combinations of occupations. Bartik labor demand instrument for the measure of the competition for high-skilled labor: the annual growth rate of share of high-skilled occupations at micro-region level, which is the dependent variable in the 1st stage ($G_{t,r}$). All results are obtained with linear regression models using fixed effects for the specified variables. Standard errors are clustered at individual-company level. Significance level. * p <0.1; ** p<0.05; *** p<0.01.*

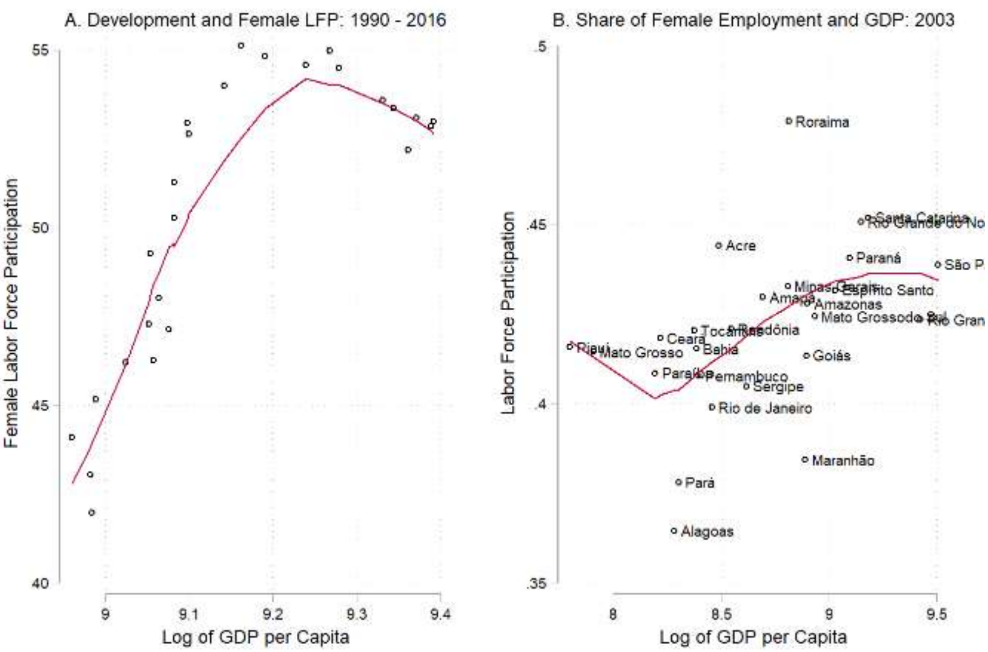
First stage results are available from the Authors upon request.

APPENDIX

A. Female Labor Force Participation and Log GDP per Capita in Brazil between 1990 and 2015.

As female labor force participation data starts in 1990, we capture only the upward part of the U-shaped curve: as GDP increases the female labor force participation increases as well. We report in the appendix also the link between share of female employment and log GDP per capita by sate for 2003. Female Labor Force Participation data comes from the World Bank. GDP data comes from the Brazilian Institute of Geography and Statistics, while the share of female employment is calculated by the Authors using RAIS data.

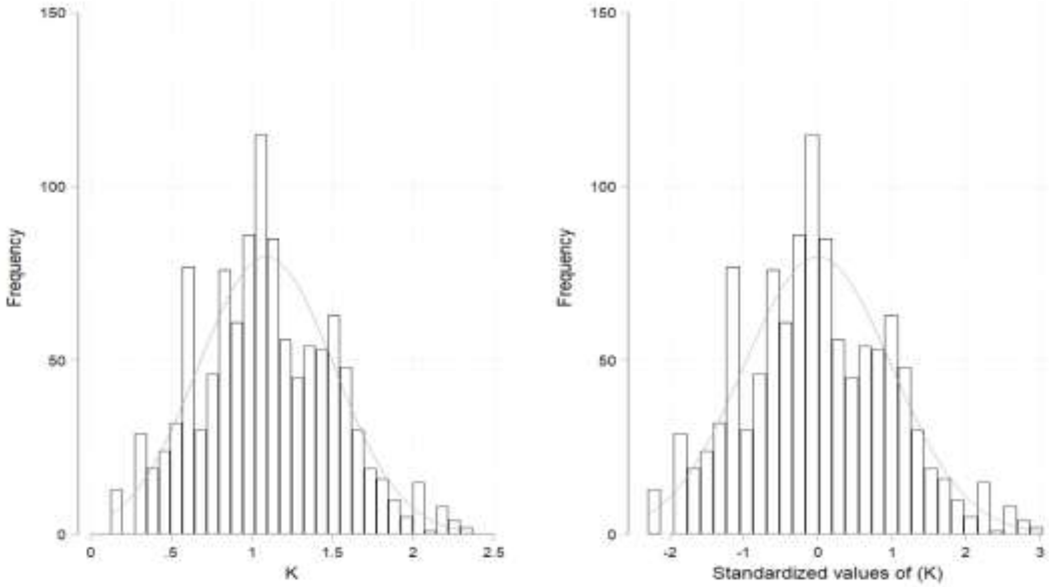
Figure A. 1. Female Labor Force Participation and Development



Note: Data for panel A comes from the World Bank. For panel B, the GDP per capita data comes from the Brazilian Institute of Geography and Statistics, while the share of female employment comes from RAIS data.

B. Distribution of the Measure of Knowledge Intensity of Economic Activity

Figure A. 2. Distribution of the Knowledge Intensity Index Before Normalization



C. Summary Statistics

Table A. 1. Average Number of Worked Hours by State

State	Average Number of Weekly Worked Hours											
	All Employment				75p K				90p K			
	Male	Female	Difference	Total	Male	Female	Difference	Total	Male	Female	Difference	Total
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Acre	*41.50*	*39.37*	*-2.13****	*40.55*	*43.11*	*42.74*	*-0.38****	*42.97*	*43.24*	*43.65*	*0.41**	*43.37*
Alagoas	*41.85*	*38.29*	*-3.56****	*40.59*	*42.66*	*41.45*	*-1.21****	*42.27*	*43.18*	*42.93*	*-0.25**	*43.12*
Amapá	*41.58*	*40.53*	*-1.06****	*41.15*	*43.34*	*42.65*	*-0.69****	*43.10*	*43.64*	*42.74*	*-0.90****	*43.44*
Amazonas	*41.02*	*39.53*	*-1.50****	*40.47*	*43.07*	*42.74*	*-0.33****	*42.96*	*43.43*	*43.36*	*-0.07*	*43.41*
Bahia	*42.30*	*39.64*	*-2.67****	*41.24*	*42.84*	*41.28*	*-1.57****	*42.43*	*42.97*	*41.99*	*-0.98****	*42.76*
Ceará	*41.80*	*39.04*	*-2.76****	*40.64*	*43.05*	*41.28*	*-1.77****	*42.51*	*43.37*	*41.96*	*-1.41****	*43.02*
Espírito Santo	*42.23*	*39.55*	*-2.68****	*41.16*	*42.92*	*41.39*	*-1.53****	*42.57*	*43.32*	*42.66*	*-0.66****	*43.20*
Goiás	*42.46*	*40.18*	*-2.28****	*41.58*	*42.92*	*41.22*	*-1.71****	*42.45*	*43.42*	*42.72*	*-0.70***	*43.25*
Grosso	*42.69*	*40.13*	*-2.55****	*41.78*	*43.36*	*41.52*	*-1.84****	*42.89*	*43.61*	*42.83*	*-0.77**	*43.42*
Maranhão	*41.34*	*37.11*	*-4.23****	*39.66*	*42.96*	*40.65*	*-2.32****	*42.43*	*43.37*	*43.12*	*-0.25*	*43.33*
Mato Grosso do Sul	*42.42*	*39.14*	*-3.28****	*41.16*	*43.25*	*42.18*	*-1.07****	*42.92*	*43.54*	*42.96*	*-0.58**	*43.40*
Minas Gerais	*42.36*	*39.08*	*-3.28****	*41.05*	*42.87*	*41.38*	*-1.49****	*42.48*	*43.10*	*42.31*	*-0.79**	*42.91*
Para	*41.49*	*37.91*	*-3.59****	*40.19*	*42.99*	*41.56*	*-1.43****	*42.67*	*43.30*	*43.16*	*-0.14*	*43.27*
Paraíba	*41.29*	*37.89*	*-3.40****	*39.91*	*43.36*	*41.72*	*-1.64****	*42.96*	*43.65*	*43.12*	*-0.53**	*43.54*
Paraná	*42.52*	*39.88*	*-2.63****	*41.41*	*43.06*	*41.07*	*-1.99****	*42.48*	*43.46*	*42.62*	*-0.84***	*43.24*
Pernambuco	*42.28*	*39.61*	*-2.67****	*41.26*	*43.26*	*41.72*	*-1.54****	*42.85*	*43.50*	*43.01*	*-0.49**	*43.39*
Piauí	*41.66*	*39.19*	*-2.48****	*40.63*	*43.22*	*41.75*	*-1.48****	*42.86*	*43.25*	*42.85*	*-0.40**	*43.16*
Rio de Janeiro	*41.97*	*40.06*	*-1.91****	*41.20*	*42.74*	*41.19*	*-1.56****	*42.28*	*43.13*	*42.27*	*-0.86****	*42.91*
Rio Grande do Norte	*42.22*	*40.08*	*-2.13****	*41.37*	*43.21*	*41.77*	*-1.44****	*42.87*	*43.71*	*43.13*	*-0.58**	*43.60*
Rio Grande do Sul	*42.29*	*39.37*	*-2.92****	*41.02*	*43.14*	*41.43*	*-1.72****	*42.66*	*43.27*	*42.43*	*-0.84**	*43.04*
Rondônia	*42.45*	*40.82*	*-1.63****	*41.81*	*43.26*	*41.85*	*-1.42****	*42.83*	*43.75*	*43.44*	*-0.31*	*43.67*
Roraima	*40.81*	*39.06*	*-1.74****	*40.00*	*42.93*	*42.05*	*-0.88****	*42.64*	*43.15*	*42.21*	*-0.95***	*42.83*
Santa Catarina	*42.57*	*40.56*	*-2.01****	*41.70*	*43.29*	*41.68*	*-1.61****	*42.83*	*43.52*	*42.47*	*-1.04****	*43.23*
São Paulo	*42.60*	*40.75*	*-1.86****	*41.84*	*43.18*	*41.92*	*-1.26****	*42.81*	*43.43*	*42.79*	*-0.64***	*43.26*
Serpige	*41.70*	*39.60*	*-2.10****	*40.87*	*42.81*	*41.58*	*-1.24****	*42.50*	*43.19*	*42.54*	*-0.64***	*43.04*
Tocantins	*42.01*	*39.55*	*-2.46****	*41.01*	*43.22*	*41.43*	*-1.79****	*42.78*	*43.76*	*43.38*	*-0.38*	*43.69*

Note: *K stands for the normalized index of the knowledge intensity of an economic sector. 75p represents the 75 percentiles of the index of knowledge intensity of an economic activity. 90p represents the 90 percentiles of the index of knowledge intensity of an economic activity. Significance level. * p <0.1; ** p<0.05; *** p<0.01 for the test of the null that the difference is not different from zero.*

Table A. 2. Average Number of Years of Company Tenure by State

State	Average Company Tenure in Years											
	All Employment				75p K				90p K			
	Male	Female	Difference	Total	Male	Female	Difference	Total	Male	Female	Difference	Total
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Acre	*4.34*	*6.93*	*2.59****	*5.50*	*2.17*	*2.37*	*0.20*	*2.24*	*1.32*	*1.42*	*0.10*	*1.35*
Alagoas	*3.88*	*6.34*	*2.46****	*4.75*	*2.44*	*2.63*	*0.19*	*2.50*	*1.94*	*2.11*	*0.17*	*1.98*
Amapá	*3.86*	*5.77*	*1.90****	*4.65*	*1.83*	*4.56*	*2.73****	*2.78*	*0.90*	*1.48*	*0.58**	*1.03*
Amazonas	*3.31*	*3.96*	*0.65****	*3.55*	*2.37*	*2.21*	*-0.16*	*2.32*	*2.17*	*2.32*	*0.14*	*2.21*
Bahia	*3.06*	*5.00*	*1.94****	*3.83*	*2.51*	*2.91*	*0.40**	*2.62*	*2.18*	*2.10*	*-0.08*	*2.17*
Ceará	*3.21*	*5.07*	*1.87****	*3.99*	*2.23*	*2.53*	*0.30**	*2.32*	*1.95*	*1.96*	*0.01*	*1.95*
Espírito Santo	*2.62*	*3.15*	*0.53****	*2.83*	*2.08*	*2.22*	*0.14*	*2.11*	*1.72*	*1.67*	*-0.04*	*1.71*
Goiás	*2.40*	*3.85*	*1.45****	*2.97*	*1.81*	*2.14*	*0.33*	*1.90*	*1.53*	*1.51*	*-0.02*	*1.52*
Grosso	*2.00*	*2.93*	*0.93****	*2.33*	*1.44*	*1.82*	*0.38***	*1.54*	*1.28*	*1.43*	*0.15*	*1.32*
Maranhão	*3.03*	*5.28*	*2.26****	*3.92*	*2.84*	*2.64*	*-0.20*	*2.80*	*1.71*	*1.88*	*0.17*	*1.74*
Mato Grosso do Sul	*2.29*	*3.08*	*0.79****	*2.59*	*1.69*	*1.92*	*0.23*	*1.76*	*1.29*	*1.31*	*0.01*	*1.30*
Minas Gerais	*2.76*	*3.80*	*1.05****	*3.18*	*2.68*	*2.43*	*-0.25*	*2.62*	*2.53*	*2.00*	*-0.53**	*2.40*
Para	*2.89*	*4.69*	*1.80****	*3.54*	*2.36*	*2.76*	*0.40*	*2.45*	*1.87*	*1.89*	*0.02*	*1.88*
Paraíba	*4.32*	*7.43*	*3.11****	*5.58*	*2.66*	*3.40*	*0.74****	*2.85*	*2.18*	*2.11*	*-0.07*	*2.17*
Paraná	*2.82*	*3.36*	*0.54****	*3.04*	*2.51*	*2.46*	*-0.05*	*2.50*	*2.06*	*1.85*	*-0.21*	*2.00*
Pernambuco	*3.27*	*5.02*	*1.75****	*3.94*	*2.63*	*2.97*	*0.35**	*2.72*	*2.43*	*3.16*	*0.72***	*2.61*
Piauí	*4.29*	*7.85*	*3.56****	*5.77*	*3.04*	*4.08*	*1.04****	*3.30*	*3.30*	*3.98*	*0.67**	*3.45*
Rio de Janeiro	*3.54*	*4.09*	*0.55****	*3.76*	*2.93*	*2.71*	*-0.21*	*2.87*	*2.66*	*2.26*	*-0.40*	*2.55*
Rio Grande do Norte	*3.78*	*6.66*	*2.88****	*4.92*	*2.46*	*3.39*	*0.94****	*2.68*	*1.95*	*2.35*	*0.40*	*2.03*
Rio Grande do Sul	*3.13*	*3.81*	*0.68****	*3.43*	*2.96*	*2.72*	*-0.24***	*2.90*	*2.83*	*2.26*	*-0.57***	*2.68*
Rondônia	*3.05*	*4.74*	*1.69****	*3.72*	*1.57*	*1.86*	*0.29**	*1.66*	*1.22*	*1.20*	*-0.03*	*1.22*
Roraima	*3.57*	*4.80*	*1.23****	*4.14*	*2.05*	*2.48*	*0.43**	*2.19*	*1.66*	*1.74*	*0.09*	*1.69*
Santa Catarina	*2.69*	*2.72*	*0.04*	*2.70*	*2.71*	*2.38*	*-0.33**	*2.61*	*2.66*	*2.03*	*-0.64**	*2.49*
São Paulo	*3.00*	*3.25*	*0.25****	*3.10*	*3.26*	*2.67*	*-0.59**	*3.09*	*3.16*	*2.50*	*-0.66****	*2.98*
Serpige	*3.93*	*6.44*	*2.52****	*4.92*	*2.62*	*3.39*	*0.77***	*2.82*	*2.49*	*2.95*	*0.46*	*2.60*
Tocantins	*2.44*	*4.49*	*2.05****	*3.28*	*1.25*	*1.57*	*0.33*	*1.32*	*0.96*	*1.16*	*0.19*	*1.00*

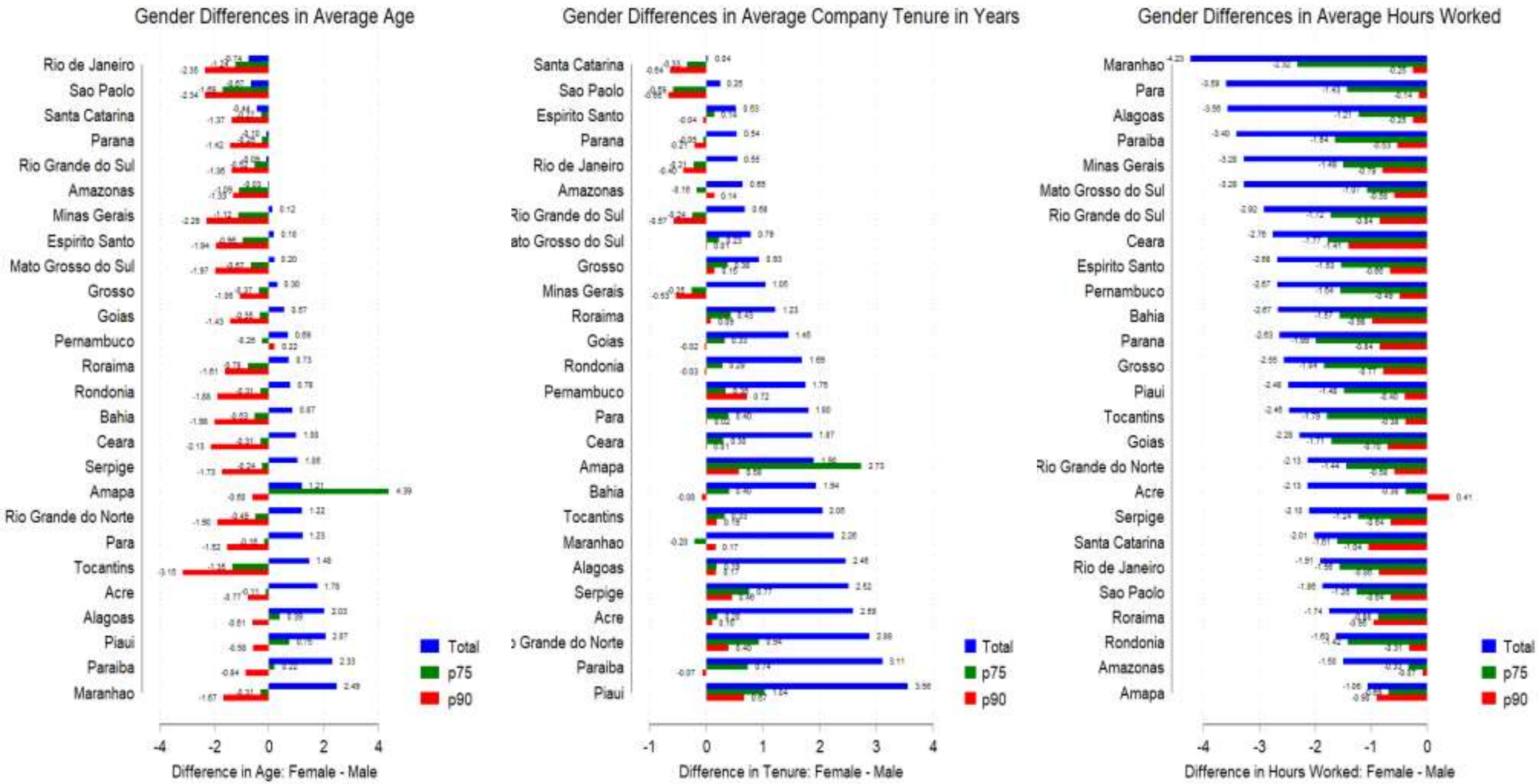
*Note: Company tenure is a variable that measures the continuous (in terms of employment) time a person spends with the same employer. Interruption periods are included here if at the end of the interruption reason (i.e. maternity leave) the employee comes back to work for the same employer. The variable a continuous variable. It is constructed as the total number of months of employment with the same employer divided by 12, so that we interpret it as year. K stands for the normalized index of the knowledge intensity of an economic sector. 75p represents the 75 percentiles of the index of knowledge intensity of an economic activity. 90p represents the 90 percentiles of the index of knowledge intensity of an economic activity. Significance level. * p <0.1; ** p<0.05; *** p<0.01 for the test of the null that the difference is not different from zero.*

Table A. 3. Average Age of Employees by State

State	Average Age													
	All Employment				75p K				90p K				Difference	
	Male	Female	Difference	Total	Male	Female	Difference	Total	Male	Female	Difference	Total	p90-p75	p90-T
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Acre	*34.21*	*35.99*	*1.78****	*35.00*	*32.20*	*32.09*	*-0.11****	*32.16*	*31.92*	*31.15*	*-0.77****	*31.68*	*-0.48****	*-3.33****
Alagoas	*34.60*	*36.63*	*2.03****	*35.32*	*33.22*	*33.62*	*0.39****	*33.35*	*32.93*	*32.32*	*-0.61****	*32.79*	*-0.56****	*-2.53****
Amapá	*34.54*	*35.74*	*1.21****	*35.04*	*32.88*	*37.27*	*4.39****	*34.41*	*32.58*	*31.98*	*-0.60****	*32.45*	*-1.96****	*-2.59****
Amazonas	*34.20*	*34.18*	*-0.03*	*34.19*	*31.85*	*30.76*	*-1.09****	*31.49*	*32.14*	*30.81*	*-1.33****	*31.78*	*0.29****	*-2.41****
Bahia	*34.66*	*35.54*	*0.87****	*35.01*	*33.87*	*33.33*	*-0.53****	*33.73*	*34.02*	*32.04*	*-1.98****	*33.60*	*-0.12****	*-1.41****
Ceará	*34.29*	*35.29*	*1.00****	*34.71*	*33.06*	*32.75*	*-0.31****	*32.97*	*33.05*	*30.92*	*-2.13****	*32.53*	*-0.44****	*-2.18****
Espírito Santo	*34.11*	*34.29*	*0.18****	*34.18*	*33.14*	*32.18*	*-0.96****	*32.93*	*32.94*	*31.00*	*-1.94****	*32.58*	*-0.34****	*-1.60****
Goiás	*33.42*	*33.99*	*0.57****	*33.64*	*31.73*	*31.38*	*-0.35****	*31.63*	*31.27*	*29.84*	*-1.43****	*30.92*	*-0.71****	*-2.72****
Grosso	*33.11*	*33.42*	*0.30****	*33.22*	*31.54*	*31.17*	*-0.37****	*31.44*	*31.00*	*29.94*	*-1.06****	*30.74*	*-0.70****	*-2.48****
Maranhão	*34.45*	*36.94*	*2.49****	*35.44*	*33.42*	*33.11*	*-0.31****	*33.35*	*33.10*	*31.43*	*-1.67****	*32.83*	*-0.53****	*-2.61****
Mato Grosso do Sul	*33.92*	*34.12*	*0.20****	*34.00*	*32.68*	*32.01*	*-0.67****	*32.47*	*32.62*	*30.65*	*-1.97****	*32.12*	*-0.35****	*-1.88****
Minas Gerais	*34.26*	*34.39*	*0.12****	*34.31*	*33.16*	*32.04*	*-1.12****	*32.87*	*33.25*	*30.96*	*-2.29****	*32.70*	*-0.17****	*-1.61****
Para	*34.20*	*35.43*	*1.23****	*34.64*	*33.25*	*33.09*	*-0.16****	*33.21*	*33.14*	*31.62*	*-1.52****	*32.88*	*-0.33****	*-1.76****
Paraíba	*35.52*	*37.85*	*2.33****	*36.47*	*33.20*	*33.42*	*0.22****	*33.26*	*32.12*	*31.28*	*-0.84****	*31.94*	*-1.32****	*-4.53****
Paraná	*33.83*	*33.73*	*-0.10***	*33.79*	*32.87*	*32.62*	*-0.25****	*32.80*	*32.43*	*31.01*	*-1.42****	*32.06*	*-0.74****	*-1.73****
Pernambuco	*34.87*	*35.56*	*0.69****	*35.13*	*33.87*	*33.62*	*-0.25****	*33.80*	*33.96*	*34.18*	*0.22****	*34.01*	*0.21****	*-1.12****
Piauí	*35.41*	*37.48*	*2.07****	*36.27*	*33.34*	*34.09*	*0.75****	*33.53*	*33.81*	*33.23*	*-0.58****	*33.68*	*0.15****	*-2.59****
Rio de Janeiro	*36.00*	*35.26*	*-0.74****	*35.29*	*33.75*	*33.25*	*-0.49****	*33.63*	*33.69*	*31.80*	*-1.90****	*33.31*	*-0.32****	*-1.98****
Rio Grande do Norte	*34.80*	*36.02*	*1.22****	*34.30*	*33.06*	*32.54*	*-0.52****	*32.92*	*32.85*	*31.49*	*-1.36****	*32.49*	*-0.43****	*-1.82****
Rio Grande do Sul	*34.34*	*34.25*	*-0.09****	*35.70*	*35.53*	*34.29*	*-1.24****	*35.16*	*35.31*	*32.96*	*-2.35****	*34.69*	*-0.47****	*-1.01****
Rondônia	*33.27*	*34.05*	*0.78****	*33.58*	*31.28*	*30.97*	*-0.31****	*31.18*	*30.90*	*29.02*	*-1.88****	*30.43*	*-0.76****	*-3.15****
Roraima	*34.14*	*34.88*	*0.73****	*34.49*	*32.09*	*31.32*	*-0.78****	*31.84*	*31.12*	*29.52*	*-1.61****	*30.57*	*-1.28****	*-3.92****
Santa Catarina	*33.01*	*32.57*	*-0.44****	*32.82*	*31.94*	*31.67*	*-0.27****	*31.86*	*31.75*	*30.38*	*-1.37****	*31.38*	*-0.48****	*-1.44****
São Paulo	*34.12*	*33.45*	*-0.67****	*35.40*	*33.41*	*33.17*	*-0.24****	*33.35*	*33.91*	*32.18*	*-1.73****	*33.51*	*0.16****	*-1.89****
Serpige	*34.98*	*36.04*	*1.06****	*34.16*	*31.31*	*29.96*	*-1.35****	*30.98*	*31.79*	*28.64*	*-3.15****	*31.20*	*0.22****	*-2.97****
Tocantins	*33.56*	*35.04*	*1.48****	*33.85*	*33.67*	*31.98*	*-1.69****	*33.17*	*33.68*	*31.34*	*-2.34****	*33.06*	*-0.12****	*-0.79****

***Note**: We refer only to the active labor force with legal age for working, i.e. 16-65 y.o. Age is measured in years. K stands for the normalized index of the knowledge intensity of an economic activity. 75p represents the 75 percentiles of the index of knowledge intensity of an economic activity. 90p represents the 90 percentiles of the index of knowledge intensity of an economic activity. Significance level. * p <0.1; ** p<0.05; *** p<0.01 for the test of the null that the difference is not different from zero.*

Figure A. 3. Gender Differences in Age, Company Tenure and Average Hours Worked by State

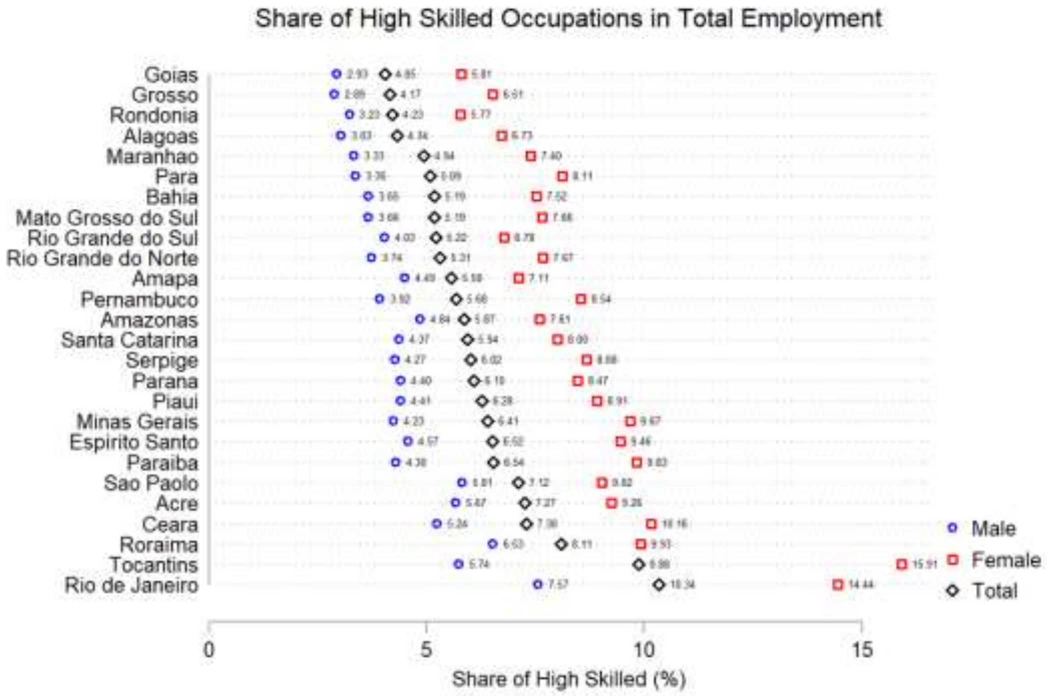


Note: 75p represents the 75 percentiles of the index of knowledge intensity of an economic activity, while 90p represents the 90 percentiles of the same index.

Figure A. 4. Labor Force Gender Decomposition by State

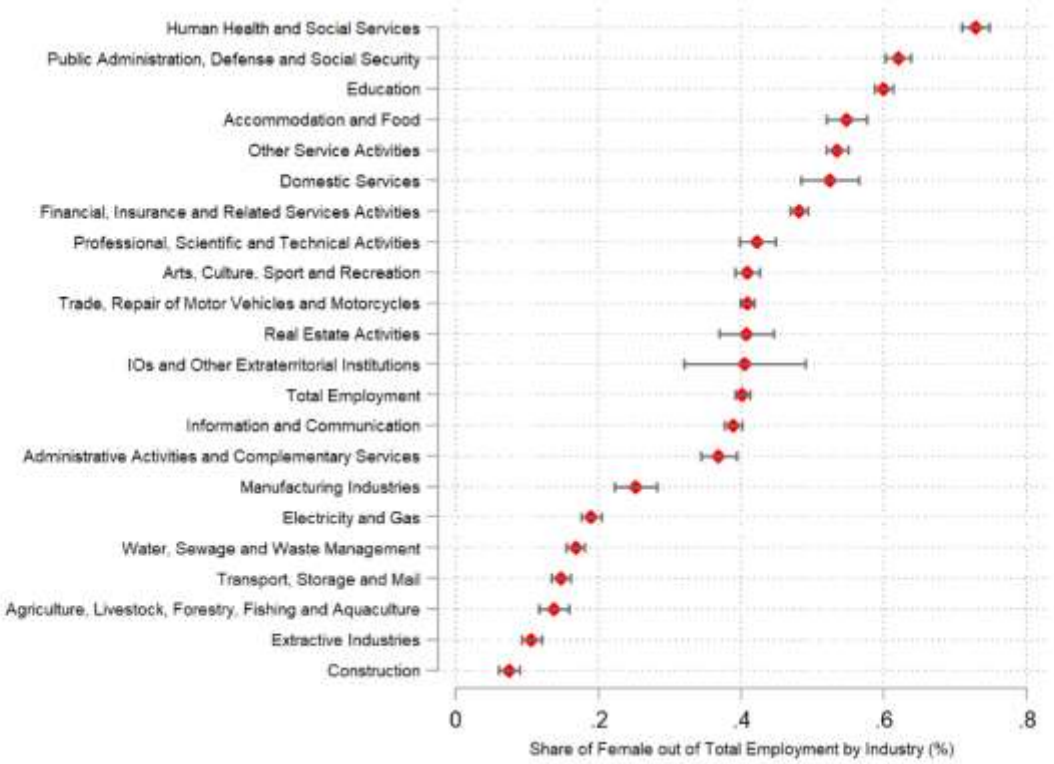
Note: 75p represents the 75 percentiles of the index of knowledge intensity of an economic activity, while 90p represents the 90 percentiles of the same index.

Figure A. 5. Share of High-skilled occupations in Total Employment by State



D. Gender Segregation by Sector of the Economy Segregation

Figure A. 6. Gender Occupational Segregation



E. Counterfactual analysis for all states – Methodology

We use the methodology proposed by DiNardo, Fortin and Lemieux, 1996 to estimate the counterfactual distributions of wages for female and male if they had opposite gender's characteristics. We start from what we have in the data:

$$f(w|female = 1) = \int f(w|x, female = 1) \cdot f(x|female = 1)dx \qquad (16)$$

A counterfactual to this distribution, for individuals that are treated as males in terms of wages but have the female characteristics is defined by:

$$\int f(w|x, female = 0) \cdot f(x|female = 1)dx$$

$$= \int f(w|x, female = 0 \cdot \frac{P[female = 1] \bullet f(x)}{P[z = 1]} dx$$

$$= f(w|female = 0) \cdot$$

$$\int \frac{P[female = 1|x]}{P[female = 0|x]} \cdot \frac{P[female = 0]}{P[female = 1]} \cdot \frac{f(w|x, female = 0) \cdot P[female = 0|x] \cdot f(x)}{f(w|female = 0) \cdot P[female = 0]} dx$$

$$= f(w|female = 0) \cdot \int \frac{P[female = 1|x] \cdot P[female = 0]}{P[female = 0|x] \cdot P[female = 1]} \cdot \frac{f(x, w, female = 0)}{f(w, 0)} dx$$

$$= f(w|female = 0) \cdot \int r(x) \cdot f(x|w, female = 0) dx$$

$$= f(w|female = 0) \cdot E[r(x)|w, female = 0]$$

And we also have that (Kernel estimates of the wage distribution):

$$\hat{f}(w|female = 1) = \frac{1}{n_0 h_0} \sum K\left(\frac{w - w_i}{h_n}\right)$$

$$\hat{E}[r(x)|w, female = 0] = \frac{\sum K\left(\frac{w - w_i}{h_n}\right) r(x_i)}{\sum K\left(\frac{w - w_i}{h_n}\right)}$$

$$\Rightarrow \int f(w|x, female = 0) \cdot f(x|female = 1) dx = \frac{1}{n_0 h_0} \sum K\left(\frac{w - w_i}{h_n}\right) \cdot r(x_i)$$

Which is nothing else than the Kernel distribution weighted by:

$$r(x_i) = \frac{P[female = 1|x]}{P[female = 0|x]} \cdot \frac{P[female = 0]}{P[female = 1]} \tag{17}$$

Where:

- w = log of wages
- f(•) = probability distribution function
- female = a dummy that is equal to 1 when the individual is female, and equal to zero when the individual is male
- x = a set of covariates that includes:
    - labor market participation such as company tenure, age (as a proxy for experience), number of missing days from work
    - education dummies

- occupation and economic sector dummies
- ethnicity, year dummies

F. Counterfactual analysis for all states – Results

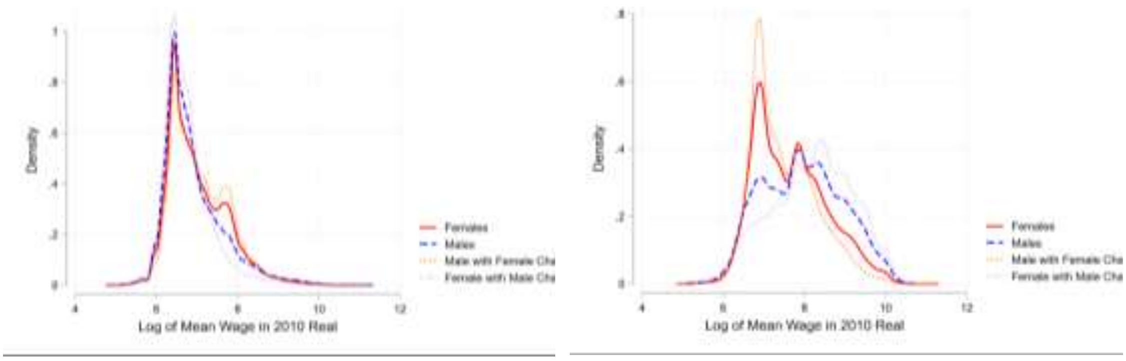
The granularity and size of our data allows us to study not just differences in average wage but also differences in wage distribution in a non-parametric way. Here we estimate the counterfactual wage distributions for female and male if they had the same observable characteristics as the opposite gender. The theoretical foundation behind this estimate is presented in Appendix E. Constructing these two counterfactual distributions allows us look at the drivers of gender inequality in wages beyond the mean. If gender inequalities were driven only by the observed individual characteristics, the real and counterfactual distributions should perfectly overlay. In other words, this would mean that the explanatory variables explain 100% of the differences observed.

We conduct this analysis on two samples for each state. First, we report the counterfactual distribution for total employment. Second, we do the same for high-skilled only. Specifically, for high-skilled occupations we are interested to see if the changing patterns of the counterfactual distribution are like those identified in the analysis on the total employment sample.

Since the interpretation of this analysis is similar for all states, in what follows we focus on Acre, alphabetically the first state. We report the entire counterfactual analysis in Figure A.8.

Figure A. 7. Counterfactual Analysis

A. All Employment **B. High-skilled occupations**



Note: The estimates are using the following covariates: knowledge intensity measure, growth rate of the share of high-skilled labor at micro-region level, age, tenure, hours worked per week, and dummies for interruptions, literacy, ethnicity, economic sector, occupations, micro-region, and year.

.In panel A of Figure A.7. we exploit the distributional changes for the entire employment. The main take away is that economic sector and occupations explain part of the gender gap, but part of the gender gaps in wages is left to be explained by other mechanisms, especially for the upper part of the wage distribution. Occupational and sectoral segregation have more explanatory power for wages above the mean. Panel B of Figure A.7. shows that for wages higher than 8 log points (approximately 2,980 Brazilian Real adjusted to 2010, which is 1,695 USD in 2010[13]) sectoral and occupational segregation generates an important part of the inequality.

[13] For an average exchange rate of 0.568458 USD for 1BRL, according to x-rates.com

Figure A. 8. Counterfactual Analysis for All States

Legend

Females

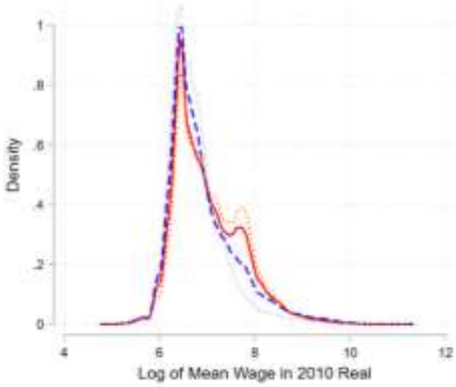
Males

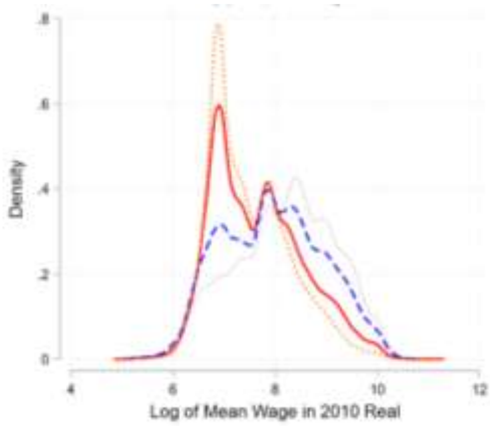
Male with Female Charact.

Female with Male Charact

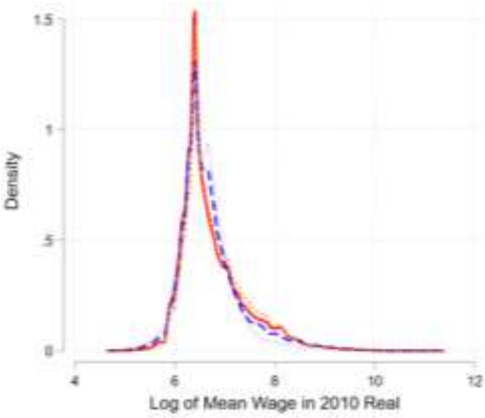
State | **All Employed Population** | **Employment in High-skilled occupations Only**

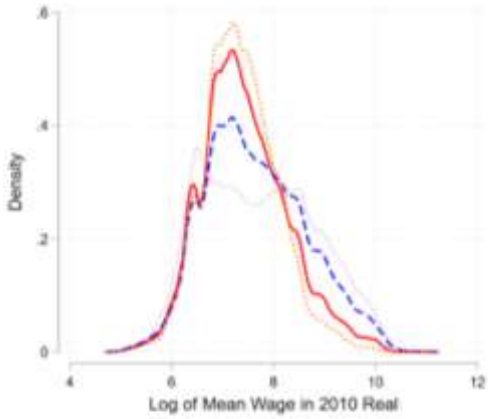
Acre



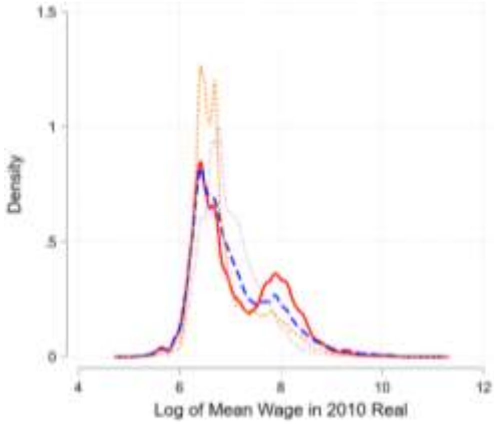


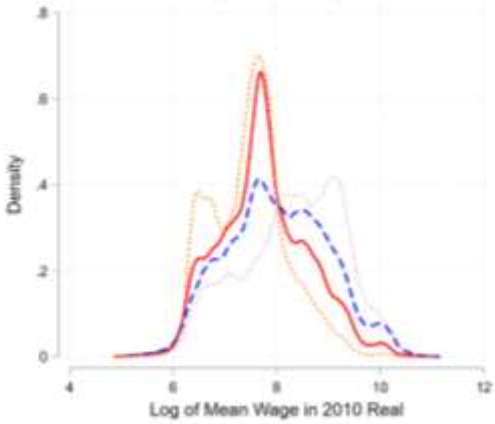
Alagoas



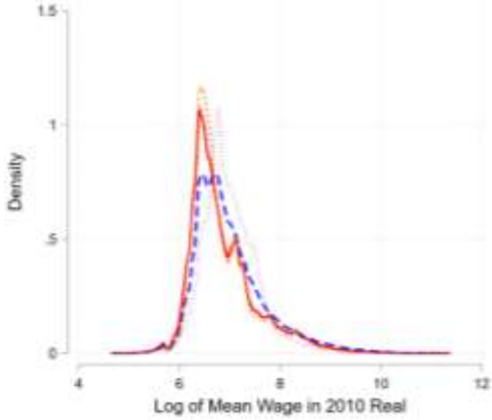


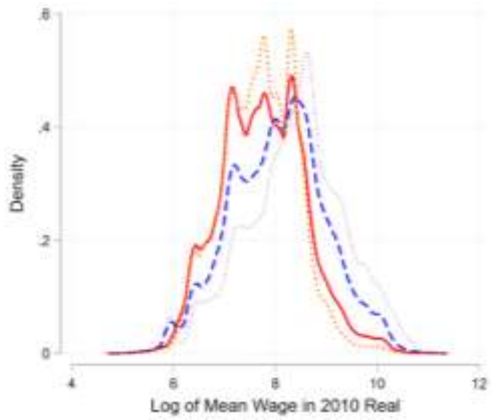
Amapá



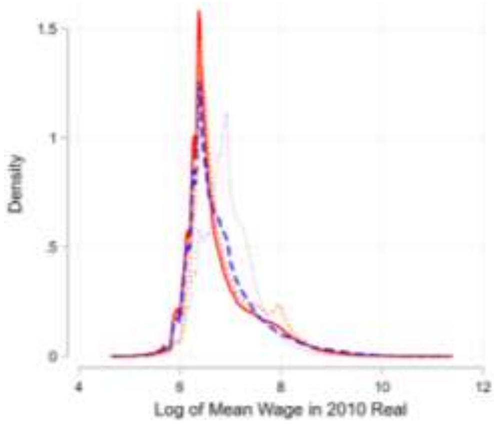


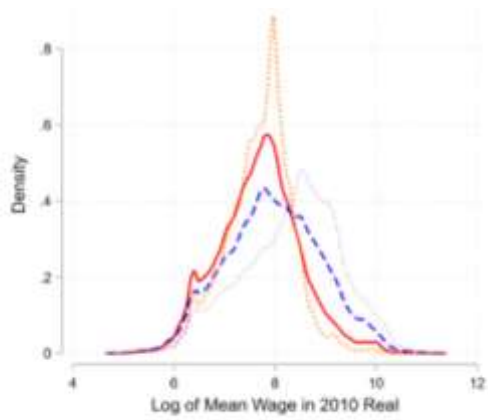
Amazonas



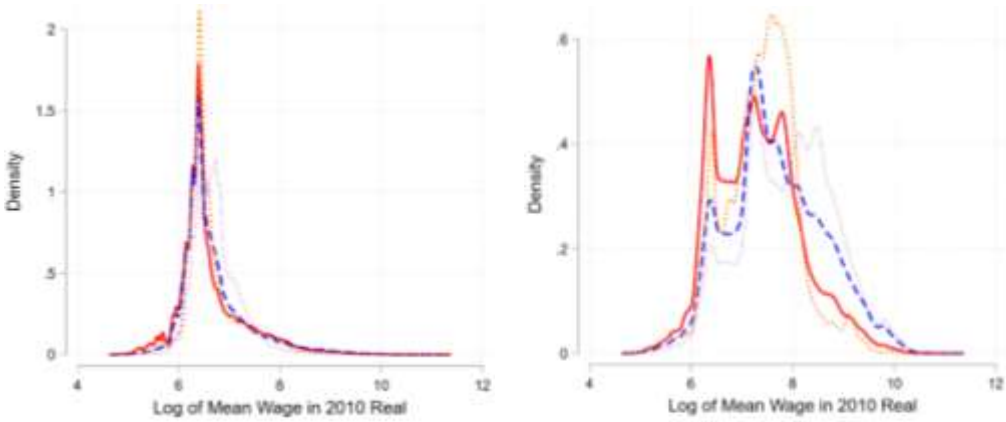


Bahia

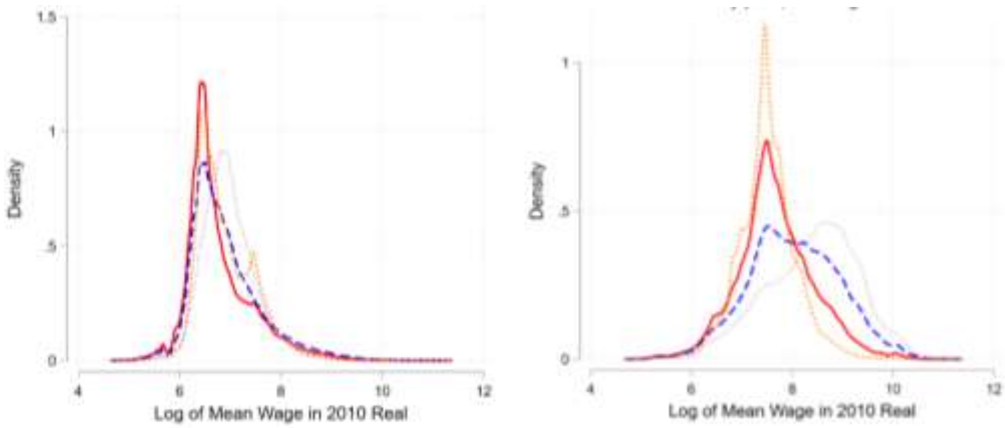




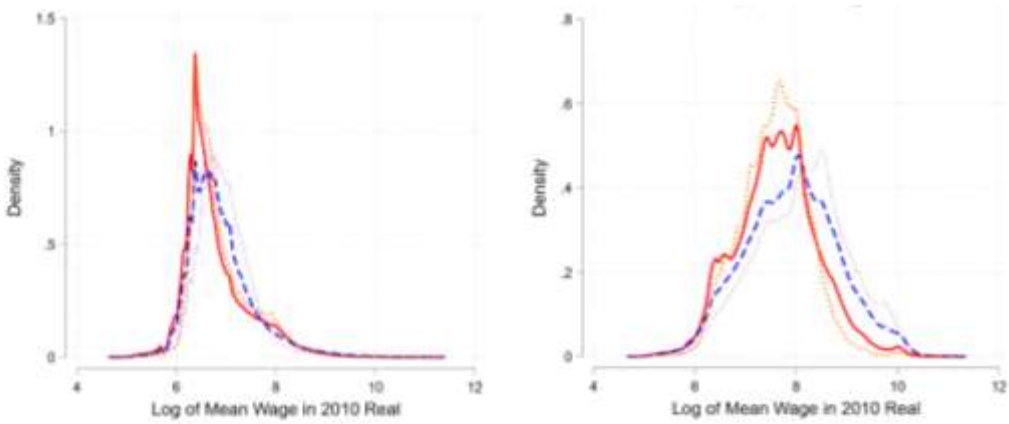
Ceará



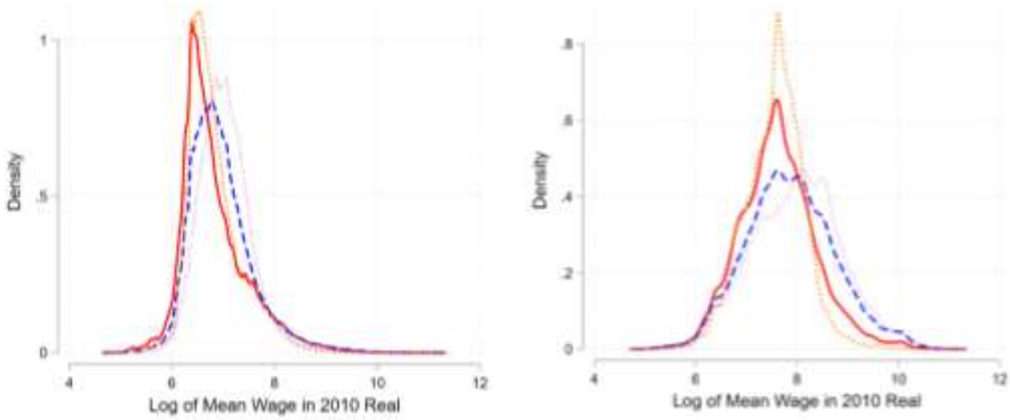
Espírito Santo



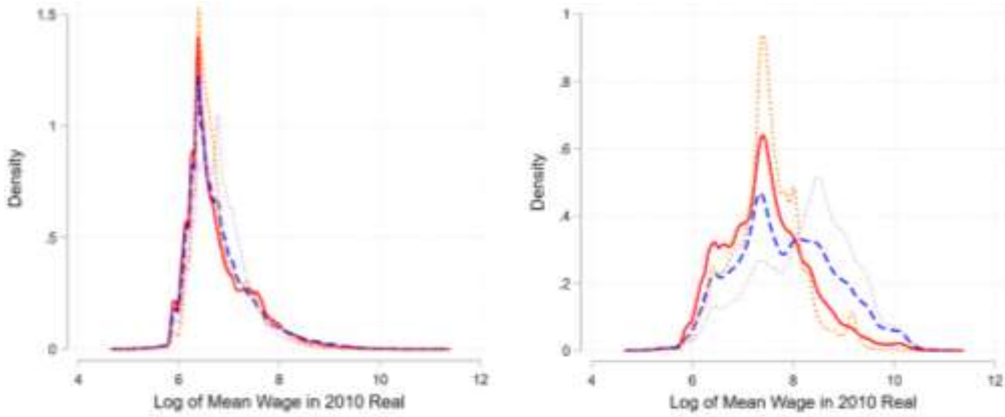
Goiás



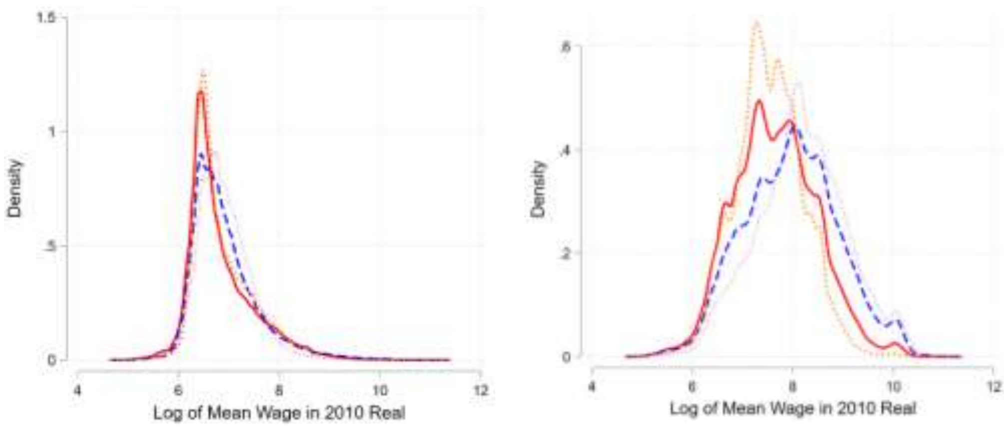
Grosso



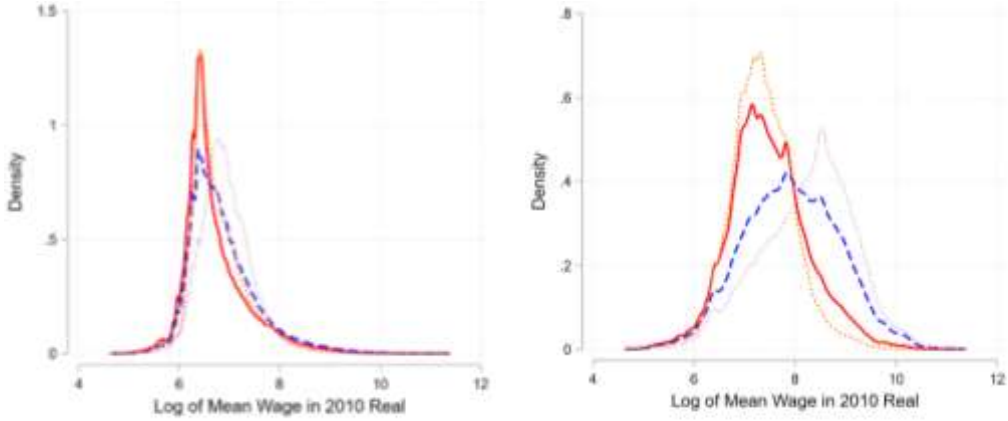
Maranhão



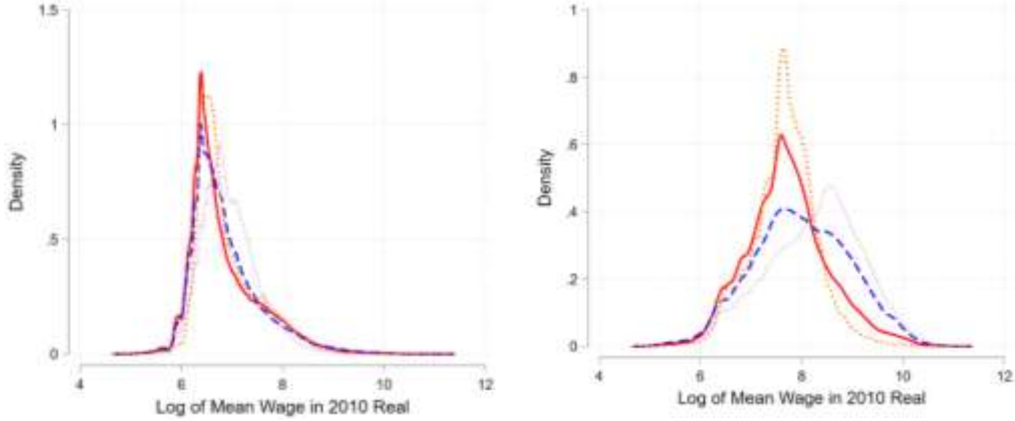
Mato Grosso do Sul



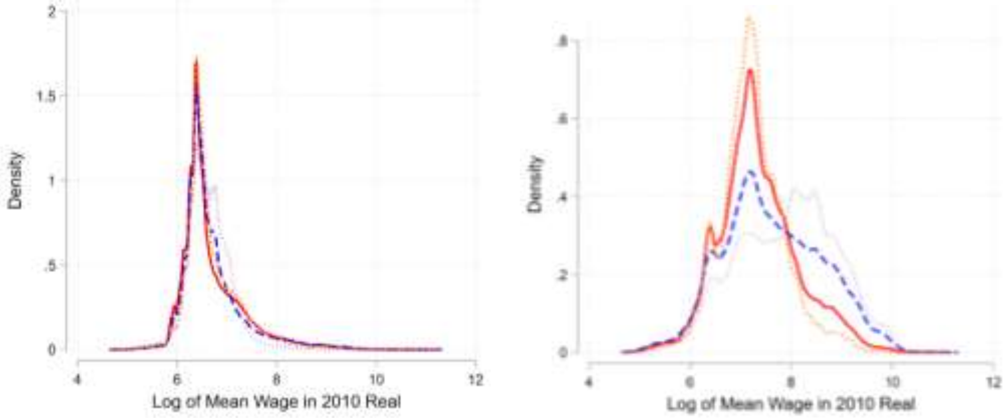
Minas Gerais



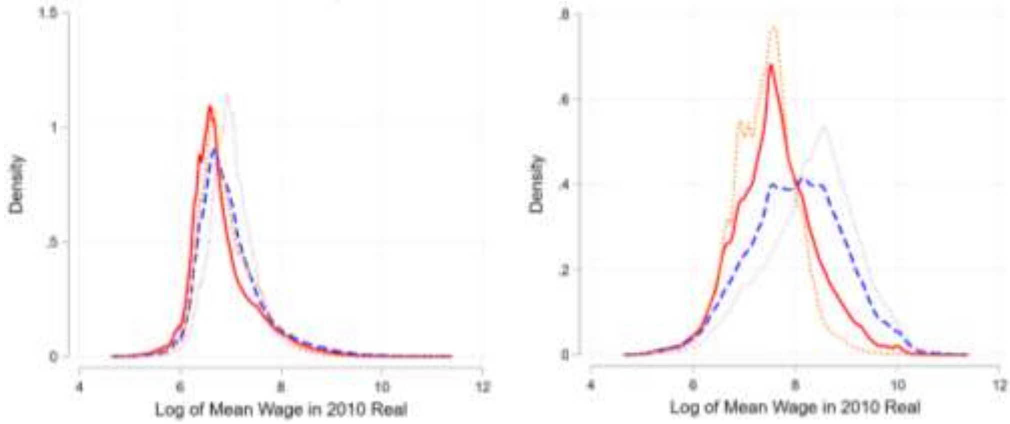
Para



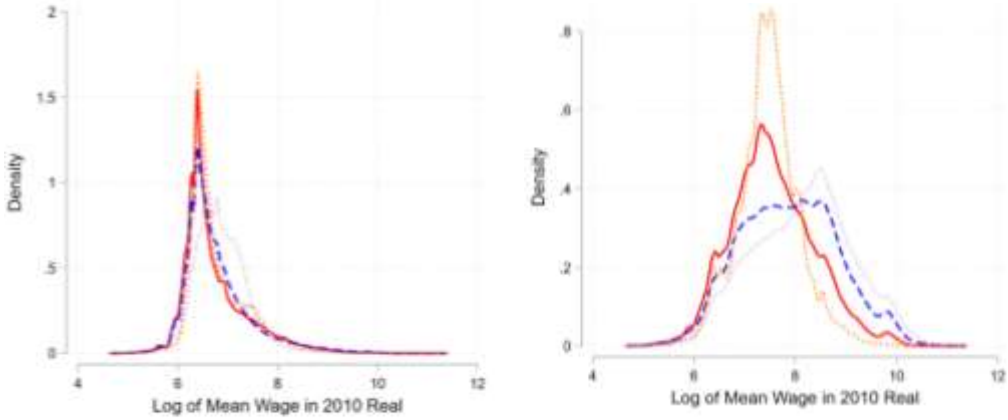
Paraíba



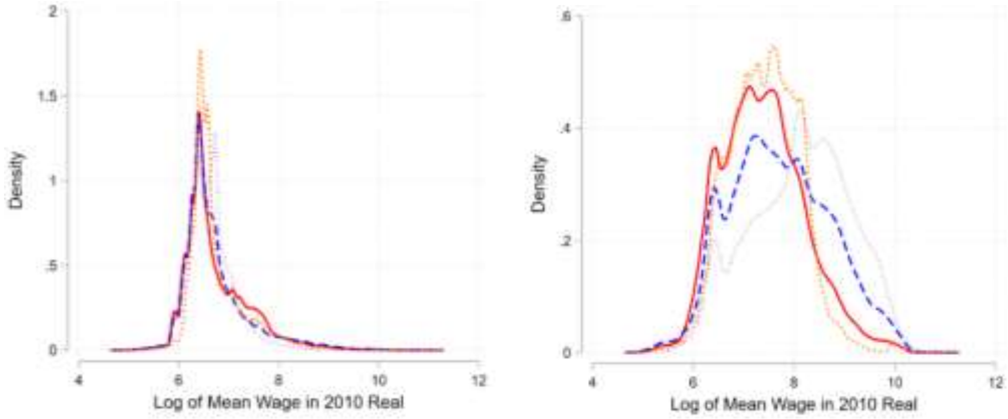
Paraná



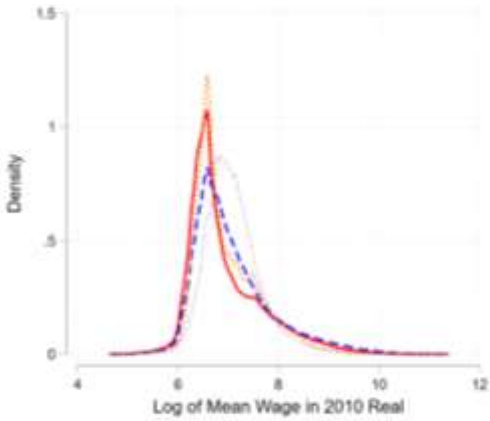
Pernambuco

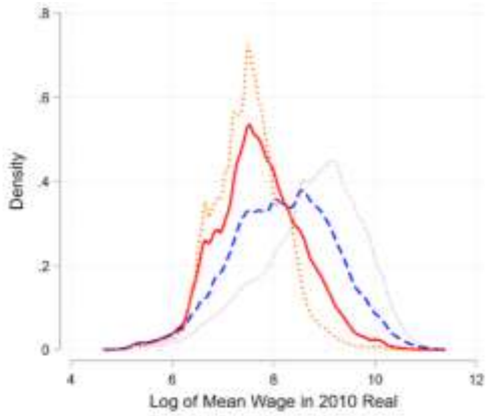


Piauí

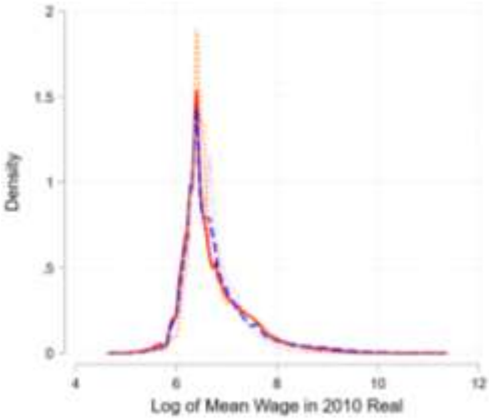


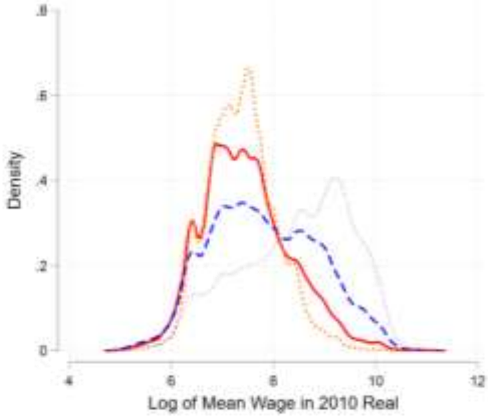
Rio de Janeiro



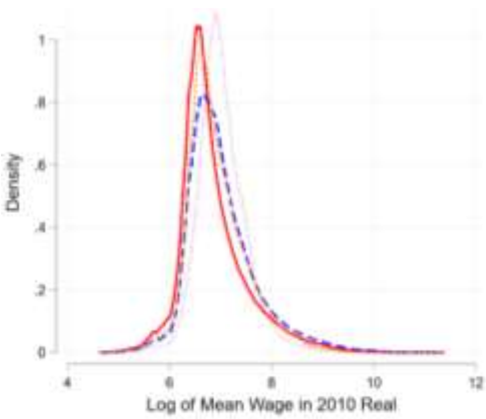


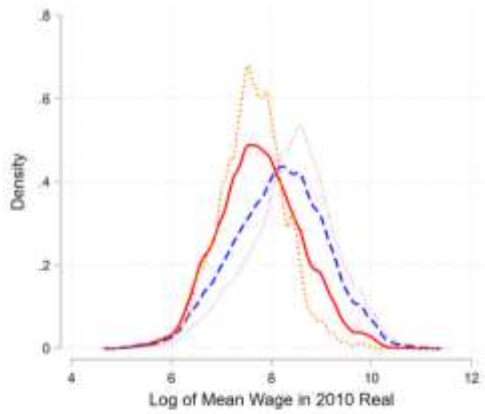
Rio Grande do Norte



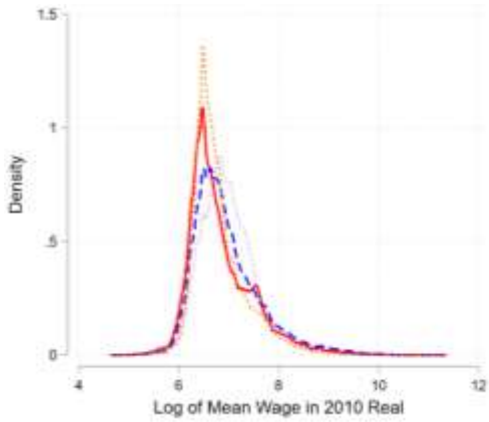


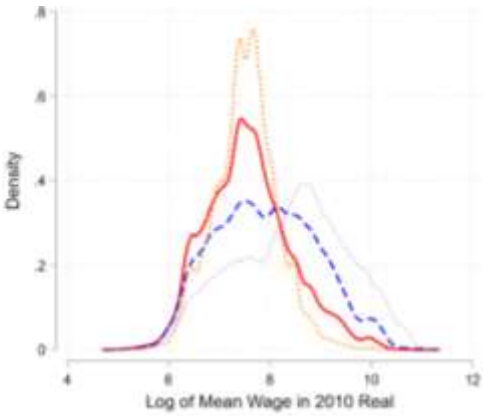
Rio Grande do Sul



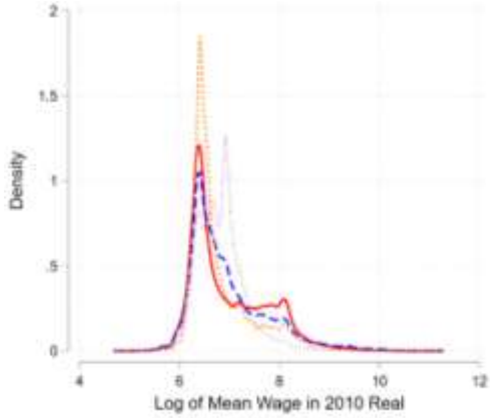


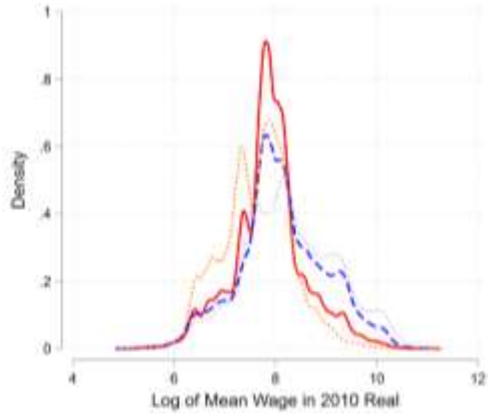
Rondônia



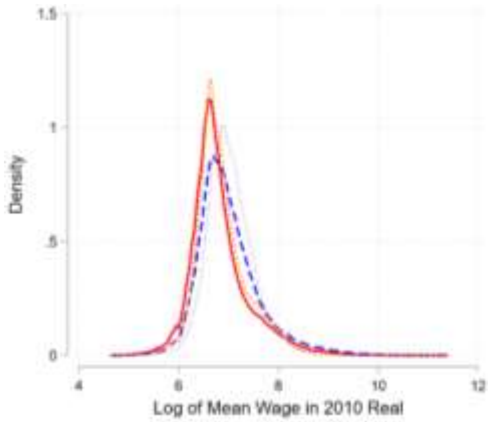


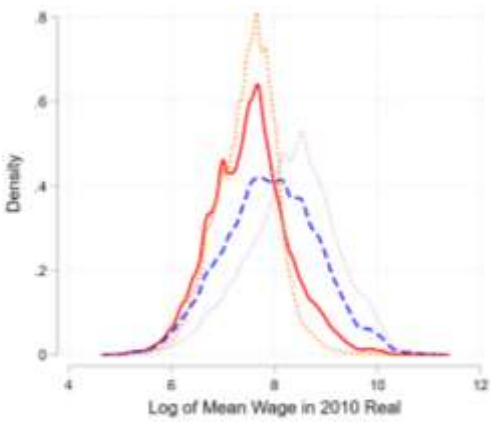
Roraima



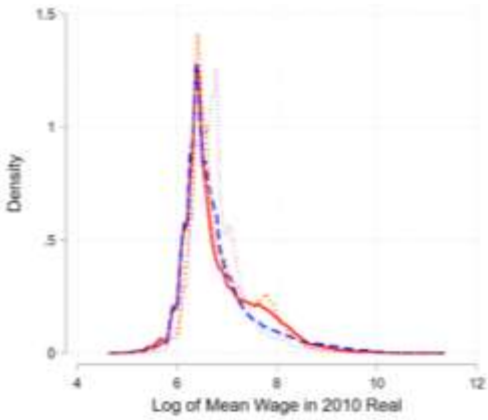


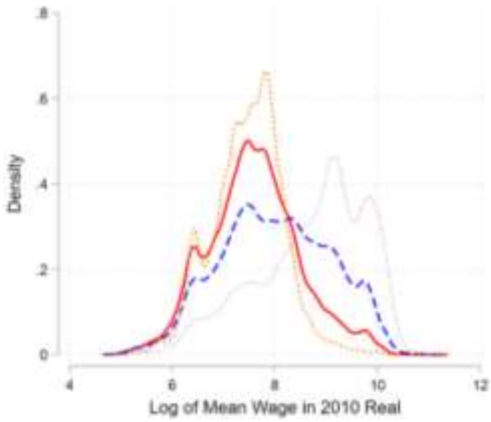
Santa Catarina



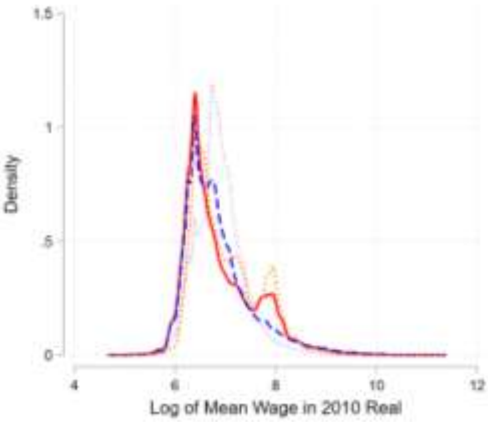


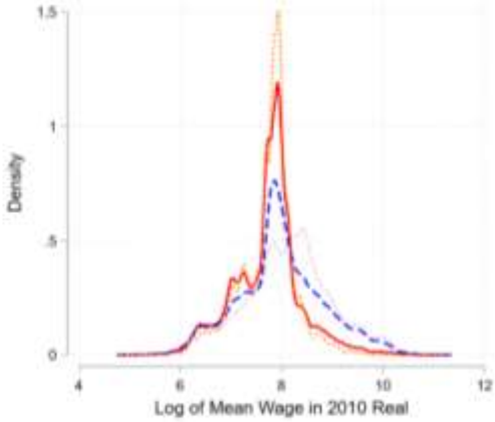
Serpige





Tocantins





Note: Please refer to notes of Figure A.7.

G. Determinants of Wages in Brazil for High-skilled occupations

Table A. 4. Determinants of Wages for High-skilled occupations

	Dependent variable: Log of Average Monthly Wage in 2010 Real				
	(1)	(2)	(3)	(4)	(5)
Acre	-0.195***	-0.182***	-0.035***	-0.035***	-0.023***
	(0.008)	(0.007)	(0.006)	(0.005)	(0.005)
Alagoas	-0.247***	-0.178***	-0.068***	-0.052***	-0.042***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Amapá	-0.238***	-0.198***	-0.073***	-0.060***	-0.047***
	(0.011)	(0.011)	(0.009)	(0.009)	(0.009)
Amazonas	-0.305***	-0.256***	-0.128***	-0.117***	-0.091***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Bahia	-0.295***	-0.228***	-0.090***	-0.080***	-0.067***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ceará	-0.430***	-0.399***	-0.226***	-0.215***	-0.192***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Espírito Santo	-0.327***	-0.273***	-0.102***	-0.086***	-0.053***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Goiás	-0.328***	-0.269***	-0.109***	-0.104***	-0.084***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Grosso	-0.295***	-0.268***	-0.110***	-0.100***	-0.091***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Maranhão	-0.401***	-0.317***	-0.090***	-0.069***	-0.042***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Mato Grosso do Sul	-0.338***	-0.314***	-0.081***	-0.078***	-0.067***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Minas Gerais	-0.394***	-0.351***	-0.117***	-0.102***	-0.077***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Para	-0.306***	-0.245***	-0.090***	-0.078***	-0.060***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Paraíba	-0.344***	-0.283***	-0.110***	-0.096***	-0.093***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Paraná	-0.374***	-0.300***	-0.091***	-0.085***	-0.069***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Pernambuco	-0.364***	-0.300***	-0.098***	-0.086***	-0.069***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Piauí	-0.359***	-0.302***	-0.083***	-0.070***	-0.057***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Rio de Janeiro	-0.390***	-0.252***	-0.091***	-0.078***	-0.073***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rio Grande do Norte	-0.399***	-0.317***	-0.087***	-0.067***	-0.046***
	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Rio Grande do Sul	-0.313***	-0.313***	-0.103***	-0.099***	-0.073***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Rondônia	-0.358***	-0.329***	-0.131***	-0.113***	-0.079***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)
Roraima	-0.208***	-0.201***	-0.049***	-0.041***	-0.024***
	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)
Santa Catarina	-0.334***	-0.313***	-0.116***	-0.110***	-0.083***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
São Paulo	-0.318***	-0.266***	-0.122***	-0.116***	-0.106***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Serpige	-0.412***	-0.336***	-0.106***	-0.105***	-0.078***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Tocantins	-0.280***	-0.273***	-0.073***	-0.063***	-0.045***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)
Individual Level Covariates	✓	✓	✓	✓	✓
Economic Sector FE	-	✓	-	✓	✓
Occupation FE	-	-	✓	✓	✓
Economic Sector- Occupation FE	-	-	-	-	✓
Year FE	✓	✓	✓	✓	✓

***Note**: Results are reported for high-skilled occupations only. The results reported in this Table and in the subsequent Tables use the full sample dataset for each reported state between 2003 and 2015 for the legal working age population, i.e. 16 – 65 y.o. All results are obtained with linear regression models using fixed effects for the specified variables.*

We employ the Correia (2016) proposed methodology for the multi-way fixed effects. Each regression from left to right differ in the levels of fixed effects as reported in the bottom of the Table. All regressions include the following covariates: age, age-squared, level of educational attainment, ethnic background, hours worked per week, leaves of absence, company tenure. Age is measured in years. All regressions include the age-squared variable. Days away from work is the number of days missed from work in each year between 2003 and 2015. Company tenure is a variable that represents the years spent by an individual within the same company (defined as continuous, based on the number of weeks worked within the same company divided by 12). Literacy variable is factor variable for different levels of literacy. Ethnic groups are also factor variables and includes the following seven groups: White, Indian, Black, Asian, Multiracial, Unidentified, and Unreported. Economic sectors refer to the 21 sectors of the economy as defined by the Ministry of Economy in Brazil. Occupations fixed effects uses the four-digit occupations codes as defined by the Brazilian Occupational Codes from the Ministry of Labor in Brazil. Standard errors are in parenthesis, clustered at individual and company level.
*Significance level. * p <0.1; ** p<0.05; *** p<0.01.*

H. Residual Gender gaps in wages by State for All Employment and for High-skilled occupations Only.

We report in Figure A.9. the residual gender gaps in wages by state for all employment and for high-skilled occupations only

Figure A. 9. Residual Gender gaps in wages by State