THE GENDER PAY GAP DURING LIFE CYCLE IN THE STEM LABOR MARKET

Cecilia Machado , Laísa Rachter , Mariana Stussi , Fábio Schanaider

Cecilia Machado , Laísa Rachter , Mariana Stussi , Fábio Schanaider

IDRC GRANT / SUBVENTION DU CRDI : - GENDER DISPARITIES, CAREER CHOICES, AND WAGE DYNAMICS IN SCIENCE, TECHNOLOGY, ENGINEERING AND MATHEMATICS OCCUPATIONS IN BRAZIL

The Gender Pay Gap during Life Cycle in the STEM Labor Market

Cecilia Machado [*] Laísa Rachter[†] Fábio Schanaider [‡]

Mariana Stussi [§]

April 5, 2023

Abstract

This paper documents the dynamics of the gender gap over time and over the life cycle in STEM occupations in Brazil. Using a matched employer-employee data on the Brazilian formal labor market and a novel classification of occupations, we compare the STEM and non-STEM gender gaps in the formal labor market. We establish four results. First, we document that, while the ratio of women to men among workers is stable in the formal labor market as a whole, it increases in STEM. This is consistent with an increase in the intensive margin of labor supply of women in STEM occupations. Second, we find that the likelihood of STEM workers continuing employed in the formal sector in the following years is higher than the likelihood of the typical formal worker. Nevertheless, the likelihood of holding a STEM job is smaller, in particular for women. Third, we estimate that that occupational controls and firms' characteristics explain each around 35% of the gender pay gap in STEM occupations, more than in the rest of the formal labor market. Fourth, we show that the gender pay gap in STEM is decreasing over generations, a pattern similar to the one found in the rest of the formal labor market.

[*]FGV EPGE. Email: cecilia.machado@fgv.br

[†]FGV IBRE. Email: laisarachter@gmail.com

[‡]FGV EPGE. Email: fabioschanaider@gmail.com

[§]FGV EPGE. Email: maristussi@gmail.com

1 Introduction

Over the last century, gender equality has experienced remarkable progress, characterized by a convergence in the roles of men and women. Among them, the convergence of human capital stands out, marked by a narrowing between men and women in the participation of the workforce, education and occupations.

However, there is still a sizable wage gap in most countries. In developed countries, the literature suggests that conventional human capital variables explain little of the gender wage gap, whereas gender differences in occupations and industries, as well as differences in the gender division of labor remain important. In particular, women's work force interruptions and hours worked remain significant an important determinant of the gender pay gap, especially in high-skilled occupations (Blau & Kahn, 2017). This occurs because technologies and management structures reduce the degree of substitution among workers, increasing the return for working long and specific hours and amplifying the gender pay gap (Goldin, 2014). One example is STEM occupations in which gender differentials in participation and earnings are typically higher than in other occupations. Due to the importance of these occupations for productivity growth and innovation, it is important to understand the gender differentials in such occupations, their evolution, and their magnitude.

In this paper, we use data on the Brazilian formal labor market from the *Relação Anual de Informações Sociais* (RAIS) to present a comprehensive picture of the STEM labor market in Brazil during the last couple of decades, emphasizing gender disparities over participation, hours worked and earnings. We examine the gender pay gap of STEM workers and examine the factors that shape the gender differentials across the life cycle. Throughout the paper, we compare our results to those of Machado et al. (2018), that documents the trajectory of the gender gap over time and over the life cycle of the Brazilian formal labor market.

First, we examine the trends in wages and participation. We show that there is a convergence trend in the gender earnings gap in the STEM labor market and there is no reduction in the gender participation gap in this market. On the contrary, the participation gap increases in the analyzed period. This had been previously documented in '*Women in the STEM Labor Market in Brazil*'. However, unlike the formal labor market, the STEM market experiences an increase in the ratio women to men of full time workers, suggesting a rise in the intensive margin of labor supply of women.

Second, we document the transition dynamics in STEM and non-STEM occupations. STEM occupations[1] are occupations from the formal labor market that belong to the STEM fields of science, technology, engineering, and mathematics, while non-STEM occupations consist of the remaining occupations.

The likelihood of a worker in a STEM occupation (henceforth called a 'STEM worker'[2]) being formally employed in the following two decades after high school is 20 percentage points higher than the average formal worker. Nevertheless, the likelihood of being in a STEM job in the same succeeding years is 40 percentage points lower for men and 50 percentage points lower for women in most of the periods. This suggests not only that there is little stability in holding a STEM job, but also that women have a harder time holding those jobs for multiple years.

Third, we investigate the role of occupations and firms in explaining the gender differentials. Occupational controls and firms' characteristics explain each around 35% of the gender pay gap. This is more than in the formal labor market as a whole. This could be due to sorting[3] of workers into high and low-paying firms and occupations within the STEM field.

To explore how much of the gender earnings gap is due to earnings inequality across establishments and the sorting of workers by firms in STEM occupations, we investigate whether women are disproportionately employed in lower-wage establishments and move with less frequency to higher-wage jobs. Particularly, we follow Goldin et al. (2017) and decompose how much of the gender earnings gap in the two decades following high school

[1]The classification of STEM Occupations was done manually following recomendations from the U.S. Bureau of Labor Statistics (USBLS, 2012a,b). CBO-2002 four-digit occupations were classified as STEM according to the USBLS definition of STEM occupations. The full classification can be found in the report 'STEM Classification in the Formal Labor Market in Brazil'.

[2]We do not observe the workers' degree field of study in our dataset, only their educational attainment level.

[3]The term sorting is often used by economists to refer to the way that market forces partition economic agents across segments of a market. Workers "sort" across jobs according to their qualifications and preferences for job attributes. In a similar way, households "sort" across neighborhoods according to their wealth and their preferences for public goods, social characteristics, and commuting opportunities.

is explained to shifting employment across establishments differing in mean earnings and how much is due to differential wage growth within establishments. We then compare our estimates for STEM jobs with other occupational groups to understand how unbalanced the sorting of workers in STEM occupations is when compared to the rest of the formal Brazilian labor market.

We find that mean establishment earnings (MEE)[4] explain 55% to 70% of the gender earnings gap in STEM occupations. When we control for mean establishment earnings, the STEM gender earnings gap shift up without changing the stability trend in the evolution of the gap. This suggests that the sorting of men into higher-wage establishments occurs when STEM workers enter the job market, and this sorting remains throughout all the analyzed years.

STEM jobs are the occupations in which MEE explains the largest fraction of the gender gap, along with Social Sciences. This means a higher proportion of men working in the high-wage establishment than women in these occupations. Management occupations have the highest baseline earnings gap, which nearly doubles in the period, and around half of it is explained by mean establishment earnings. For health and law occupations, however, mean establishment earnings explain little or nothing about the gender pay gap, meaning that the allocation of men and women in higher-wage establishments is more balanced.

Finally, we analyze the life cycle dynamics of gender differentials. We find suggestive evidence that the gender pay gap in STEM has been reducing over the generations as also found in non-STEM occupations. Throughout the life cycle, the gender pay gap expands until around the age of 45, and then reduces until the end of the career. We do not observe substantial differences in the gender gap over the life cycle between white and non-white workers.

The remaining of this paper is structured as follows. Section 2 presents the data. Section 3 describes the empirical strategy used to document the gender pay gap over the life cycle. Section 4 documents some empirical facts which motivate our work and justify some of our

[4] Mean establishment earnings are calculated by taking the average earnings for each establishment across the whole period for all STEM employees, and therefore is time invariant. This is further explained in Section 3.

modeling strategies. Section 5 presents the results, and section 6 concludes.

2 Data

We use data from the *Relação Anual de Informações Sociais* (RAIS) for the period 2003 to 2019. *RAIS* is a longitudinal administrative database with employer-employee information provided by the Ministry of Labor. In Brazil, firms are required to report all workers formally employed at some point in the previous year. Each worker is identified by a unique identifier (*PIS* or *CPF*) and the firms are identified by the *CNPJ*. *RAIS* provides information about workers (type of employment, length of employment, gender, color, age, education, monthly salary, contract hours, occupation, and information on leave), as well as characteristics of the firm (sector of activity, size average wages).

We restricted the analysis to a sub-sample of workers between 18 and 65 years old and who worked all months in a given year. Workers with inactive jobs in December 31 of the year were dropped. In the case of individuals with more than one active job, the main one is considered the one with the highest earnings. Real earnings were calculated by multiplying December earnings (in minimum wages) by the value of the minimum wage for each year. The *INPC* (*Índice Nacional de Preços ao Consumidor*) is used to deflate wages to the most recent period (2019).

To determine which workers work in STEM occupations, we use the classification proposed in the document '*STEM Classification in the Formal Labor Market in Brazil*' and match *CBO-2002*'s occupational codes in *RAIS* with our classification codes. We look at all workers who have been in STEM jobs in at least one year between 2003 and 2019 and searched for them in the remaining years in *RAIS*. This is done so that workers who were classified as STEM in one year can be observed in all periods they were in the formal labor market, regardless of whether they were in STEM jobs in all years. This prevents confusion between exiting the formal job market and exiting the STEM market. In the end, our database consists of a panel[5] of STEM workers, for all the years that we were able to observe these

[5] It is important to differentiate a panel of STEM workers from a stacked cross-section data set of STEM workers. In the first case, as it is the case of this report, we track all individuals that were once in a STEM occupation between 2003 and 2019 in all years they appear in the formal labor market, regardless of switching

workers in the formal labor market.

3 Empirical Strategy

Following Goldin et al. (2017), we divide the sample into three education by age groups: (1) individuals between 18-22 years in 2003 who did not complete high school; (2) individuals between 23–27 years in 2003 who completed high school but did not go to university; (3) individuals between 28-32 years in 2003 who completed university. We choose ages to ensure each group would likely have completed the highest grade and still be young enough to be early in their careers. Our dependent variable is the real income (in log) in the main job for each individual i in the year t.

We estimate the following baseline econometric model using a panel of all i individuals who were in a STEM occupation in at least one of the years t between 2003 and 2019:

$$y_{it} = \alpha_{et} + \delta_a + \sum_{j=2003}^{2019} \phi_j \times F_i + \epsilon_{it}, \qquad (1)$$

in which y_{it} is the real income (in log) for each individual i in the year t; α_{et} is an education-year fixed effect[6], δ_a is an exact age fixed effect; F_i is a sex indicator, and ϵ_{it} is an idiosyncratic error term.

Our coefficients of interest are the ϕ_j. These coefficients map the trajectory of the gender *gap* over the life cycle conditional on age and education. It is worth noting that the year j will also follow the age of the individuals.

occupations or not. In the second case, workers who are in STEM occupations are selected in each year, while the remaining workers are dropped. Therefore, in the stacked cross-section, STEM workers who remain employed in the formal labor market in non-STEM occupations can be confused with STEM workers who leave the formal labor market.

[6]In econometrics, fixed effects is a statistical regression technique in which the intercept of the regression model is allowed to vary freely across individuals or groups. It is often applied to panel data in order to control for any individual-specific attributes that do not vary across time, such as sex or ethnicity. Intuitively, including fixed effects allows the model to control for all variables that are common to a set of individuals or group, whether they're observed or not, as long as they stay constant within some larger category. In other words, when adding educational, occupational or firms fixed effects, we are controlling for all things that are common to that group of individuals that are constant over time.

To assess the role of occupational and firm's characteristics in the gender pay gap, we extend the baseline model. First, we augment equation (1) by including a set of 3-digit occupational fixed effects Θ_i for observed employment in 2003:

$$y_{it} = \alpha_{et} + \delta_a + \sum_{j=2003}^{2019} \phi_j \times F_i + \Theta_i + \epsilon_{it}, \tag{2}$$

Equation (2) is identical to equation (1) with the exception of occupational fixed effects. Thus, the comparison between the coefficients ϕ_j in these two equations enables us to understand the extent to which different occupational sectors explain the evolution of the gender wage gap in STEM throughout the life cycle.

One important issue for estimating equation (2) is the fact we are analyzing the role of occupations conditional on being STEM occupations. Because the choice of working in STEM is endogenous, the coefficients will typically understate wage differentials across genders.

Second, we expand the model to include firm characteristics ($\Omega_{f(i)t}$):

$$y_{it} = \alpha_{et} + \delta_a + \sum_{j=2003}^{2019} \phi_j \times F_i + \Theta_i + \Omega_{f(i)t} + \epsilon_{it} \tag{3}$$

The term $\Omega_{f(i)t}$ is a firm-specific effect. It is composed by the size of the firm (total number of employees), average salary (in log) of the establishment per year, percentage of women working in the firm in that year and fixed effects of firm's activity sector according to CNAE (3 digits). It also includes a time-invariant firm fixed effect that does not change over time.

Our analysis is carried out by comparing the different ϕ_j estimated in (1), (2) and (3). Comparing the coefficients estimated in different models it is possible to examine the extent to which the selection of workers between and within companies affects the wage differentials between men and women throughout the life cycle (conditional on being in the STEM market). Intuitively, this comparison enables us to pin down the relative contribution of each component (*i.e.*, the role of occupations and firms) in explaining gender wage gaps.

Heterogeneous effects by Education

We complement the analysis presented before by modifying the econometric model to capture potential heterogeneous effects based on educational level:

$$y_{it} = \alpha_{et} + \delta_a + \sum_{g=1}^{3} \sum_{j=2003}^{2019} \phi_{jg} \times F_i + \Theta_i + \Omega_{f(i)t} + \epsilon_{it}, \tag{4}$$

in which g is a educational group.

3.1 Heterogeneous effects by Establishment Earnings

We extend the Equation 1 (baseline model) by adding $ln(MEE)_{it}$, mean establishment earnings (in log), which is the mean for each establishment across the whole period for all STEM employees, and therefore is time invariant and is not calculated from only those in the three education groups sample. In addition to $ln(MEE)_{it}$, we include controls for firms' characteristics $\Omega_{f(i)t}$, composed of the size of the firm (total number of employees) and the percentage of women working in the firm in that year. We also add Θ_i, the three-digit CBO occupation, and ζ_f, the three-digit firm activity sector according to CNAE.

$$y_{it} = \alpha_{et} + \delta_a + \sum_{j=2003}^{2019} \phi_j \times F_i + \gamma_{it} ln(MEE)_{it} + \Omega_{f(i)t} + \Theta_i + \zeta_f + \epsilon_{it} \tag{5}$$

We include controls progressively in stages and compare the different ϕ_j estimated in the equations to assess the role of each additional component in the gender pay gap. We seek to know if women are sorting into lower-wage establishments and to what extent this sorting contributes to the gender earnings gap compared to its other components, such as occupation and other firm characteristics. In other words, we aim to document how much of the gender earnings gap in the two decades after school is due to shifting employment by gender across establishments with different average earnings and how much is due to differential wage growth within establishments. Finally, we compare equations (1) and (2) between STEM jobs and other areas, namely management, health, law, and social sciences.

4 Stylized Facts

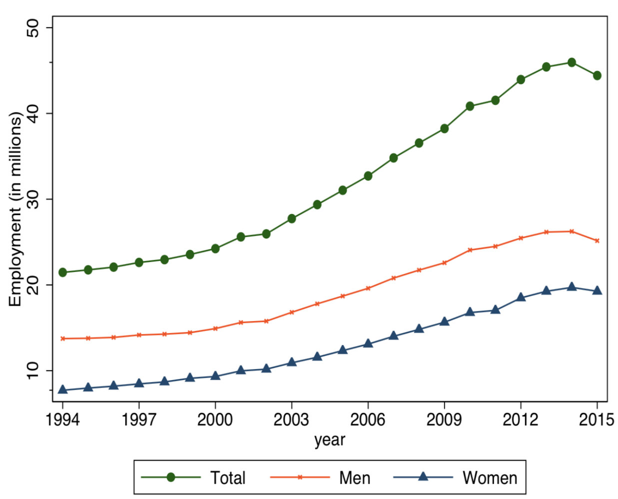
4.1 Participation by Gender: Extensive and Intensive Margins

Before presenting our main results, we present some stylized facts about the STEM labor market in Brazil. Figures 1a and 1b plot the employment by gender in the formal labor market and in the STEM labor market, respectively. While formal employment grew in parallel between men and women, in the STEM fields employment increased more among men, although the widening of the participation gap stabilized from 2011 onwards.

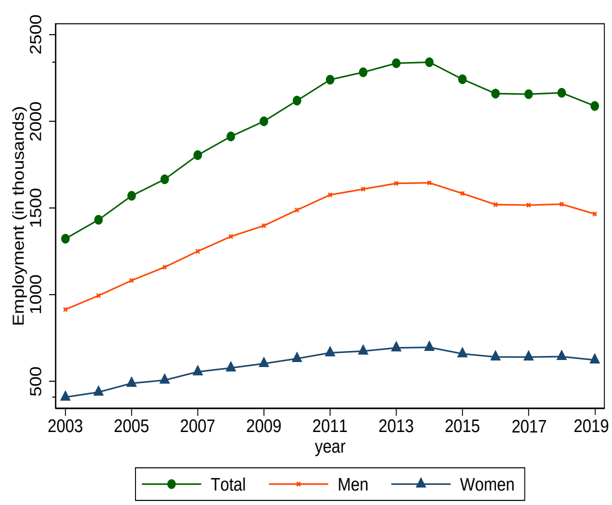
However, just looking at the number of male and female workers in the two markets does not tell the whole story. When we look at the intensive margin of the labor supply, *i.e.*, considering the number of hours worked and not only the number of people employed, we see that the fraction of women working full-time has increased more in STEM than in the rest of the formal labor market. This is shown in figures 2a and 2b.

Figure 1: Employment in the Brazilian Formal Labor Market by Gender

(a) Formal Labor Market, period 1994-2015



(b) STEM Labor Market, period 2003-2019



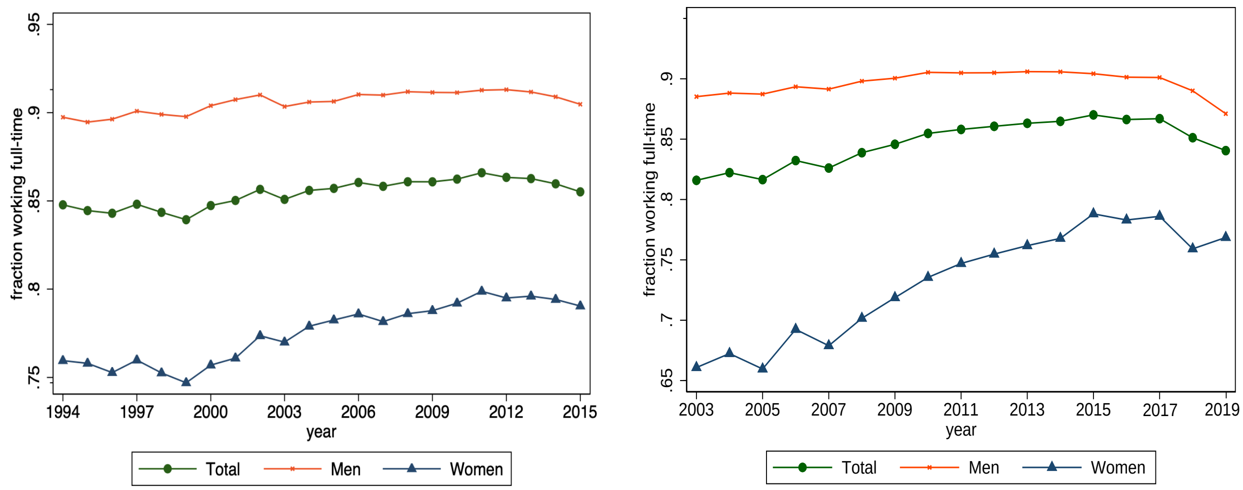
Source: Machado et al. (2018)

In the past, the fraction of full-time female workers was smaller in STEM than in the formal job market (66% in 2003 against 77%), and in recent years this fraction has increased to near 80% in both markets. The fraction of full-time male workers in the formal labor market remained stable at 85%, and experienced a modest jump from 81% to 84% in the STEM market. It is worth noticing that in the last couple years both men and women in STEM have seen a drop in the fraction of full-time workers, despite a slight recovery in 2019.[7]

Figure 2: Full-time Workers by Gender

(a) Formal Labor Market, period 1994-2015

(b) STEM Labor Market, period 2003-2019



Source: Machado et al. (2018)

To see the whole picture, figures 3a and 3b plot the evolution of the ratio women to men in employment, earnings and full-time participation for both markets. In the formal labor market, there is an accelerated convergence in participation and a stable earnings ratio, while in the STEM market there is a convergence in wages and a slight drop in participation. However, when comparing the intensive and extensive margins of participation[8], we see that

[7] A possibility to explain this stylized fact is that, in 2017, Brazil passed the first meaningful reform of labor laws since the 1940s. The changes diluted collective bargaining, reduced the scope for legal action in labor disputes, regulated remote work and gave companies more control over workers' hours and vacation time, allowing for more flexible schemes of work.

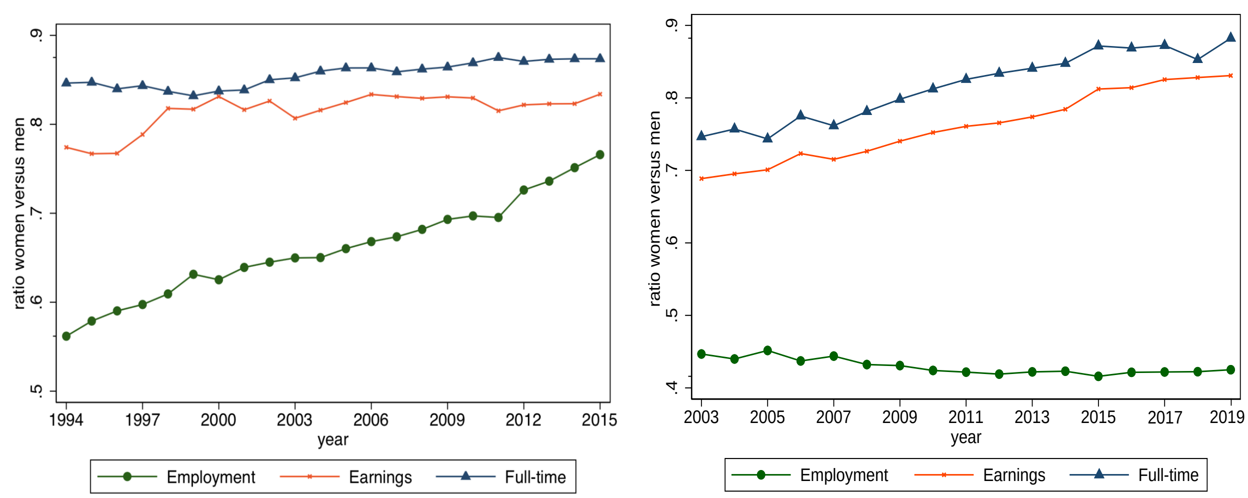
[8] In the labor economics literature, the terms intensive and extensive margins of the labor supply are used to define two different sides of the workers' participation. The extensive margin refers to weather an individual decides to work or not, and the intensive margin refers to how many hours of work the individual was working in a certain period.

opposite phenomena occurs in the labor market, and on STEM occupations in particular. While in the formal market there is a rapid convergence in the ratio women to men of employment (extensive margin) and constancy in the ratio women to men of full time workers (intensive margin), in the STEM market it is the intensive margin that is converging, while the extensive margin shows a relative stability with a slight downward trend in the period. This means that, although the number of women working in the STEM market is stable over the period, women are working more hours.

Figure 3: Gender Differences (ratio women to men) in Employment, Average Pay and Full-time Workers

(a) Formal Labor Market, period 1994-2015

(b) STEM Labor Market, period 2003-2019



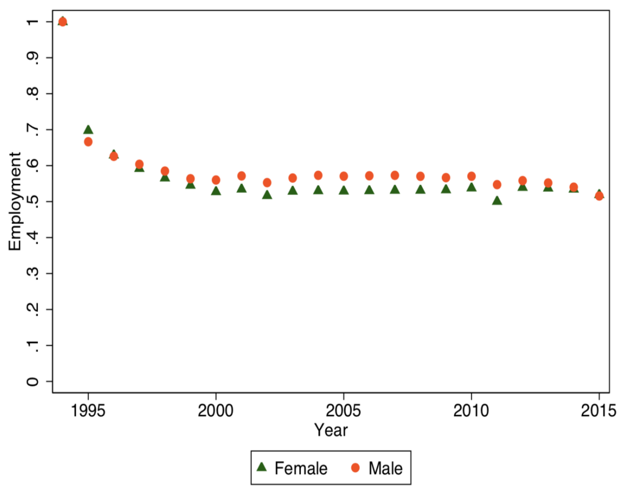
Source: Machado et al. (2018)

4.2 Probability of being formally employed in the following years

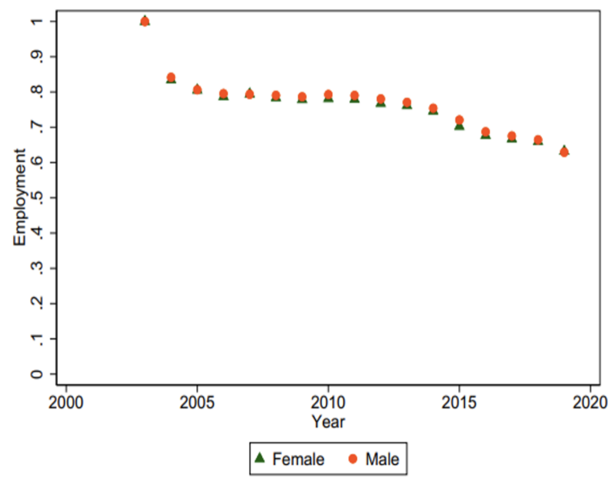
To assess the stability of formal and STEM employment, we plotted the likelihood of being formally employed in the following years for a given cohort and employed in the first year of the sample. As we have a different sample period from Machado et al. (2018) for the STEM market, we considered a different cohort, but with the same age in the first year of our sample (2003). Figure 4a shows the original graph by Machado et al. (2018), which shows the likelihood of being formally employed in the succeeding years for the cohort born in 1967-1974 and employed in 1994. Figure 4b shows the likelihood of being formally employed in the succeeding years for the cohort born in 1976-1983 and employed in 2003, considering only STEM workers.

Figure 4: Likelihood of being formally employed in the succeeding years

(a) Formal Labor Market, cohort born in 1967-1974 and employed in 1994



(b) STEM Labor Market, cohort born in 1976-1983 and employed in 2003



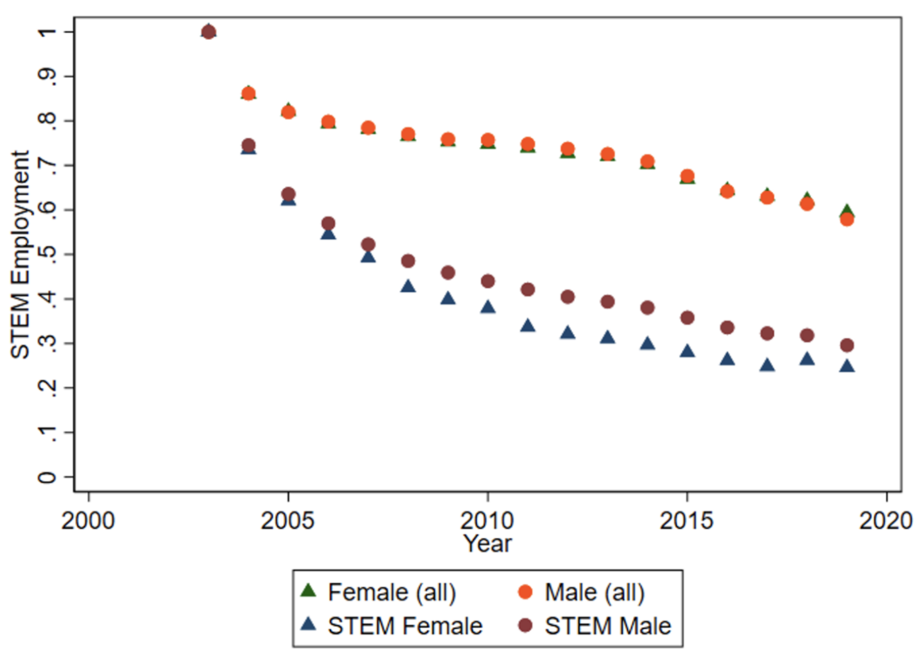
Source: Machado et al. (2018)

In the formal labor market the probability of being employed drops sharply in the first year, declines slowly until 2002 and then stabilizes. There is a small difference in the likelihood faced by men and women, with the likelihood of women being formally employed being slightly lower than that of men between 2000 and 2010. For the STEM market, the scenario is different. First, the likelihoods faced by men and women in this market are virtually identical throughout the entire period. In second place, the likelihood of being employed faced by STEM workers also suffers a greater drop in the first year, but this drop is smaller than that observed in the formal labor market. After the fall, there is a period of stabilization, and in recent years a new downward trend appears.

We also check the probability of remaining employed in the STEM field as a measure of job stability. Figure 5 plots likelihood of being employed in a STEM job against the likelihood of being formally employed in any formal job in the succeeding years, for male and female STEM workers. That is, we want to assess the probability of STEM workers remaining in a STEM occupation versus the probability of a STEM worker remaining employed in the following years, regardless of field, to understand if STEM workers remain in STEM fields.

We found that the probability of being employed in a STEM job drops sharply until 2007, after which it starts to have a smoother downward trend, reaching a mere 30% in 2019. While the probability of a person who was employed in a STEM job being formally employed in 2019 is around 70%, the probability that this person is employed in a STEM occupation is 40 percentage points lower, or 30%. There is yet another striking fact: as of 2008, there is a growing gap between the likelihood of a woman being in a STEM job versus that of a man, that is not observed in the probability of being formally employed. This gap reaches its peak between 2011-2017, when it is around 10 percentage points.

Figure 5: Likelihood of being formally employed and of being in a STEM job in the succeeding years: cohort born in 1976-1983 and employed in STEM occupations in 2003



5 Results

5.1 Relative Importance of Occupational and Firm Controls in the Earnings Gap

Moving on to our main results, we focus on the specific cohort born between 1976-1983 and employed in 2003. Figure 6 plots the estimated coefficients, ϕ_j for equations (1), (2) and (3). For comparison purposes, we show the results for both the formal labor market (6a) and for the STEM market (6b).

Considering the baseline model (Equation 1), there is a gender pay gap above 20 percent over the entire life cycle for the formal market, and over 30 percent for the STEM market. Adding occupational controls[9] reduces the gap in both markets to around 20 percent over the entire life cycle, equalizing the gaps between the two markets. This makes sense since one is a subset of the other, selected based on occupations of the formal labor market.

Finally, adding firm controls[10] narrows the gap even further in both markets in a similar way. Estimating equation (3) show us that the gender gap doesn't seem to differ much when we compare the formal labor market with the STEM subset, and both remain steady around 10-15% over the life cycle.

5.2 Heterogeneous Results by Education

Next, we restricted workers to those belonging to the following educational-age groups: did not complete High school at ages 18-22, High school graduates and with incomplete College at ages 23-27 and College graduates and above at ages 28-32.

Figures 7, 8 and 9 show the earnings gender gap admitting heterogeneity by educational group for the STEM Market. Equivalent graphs for the formal sector as a whole are available in the Appendix for comparison.

Figure 7 plots the estimated coefficients for the three educational groups considering the

[9] Occupational controls mean including a set of 3-digit occupational fixed effects Θ_i. See Section 3.

[10] Firm controls are represented by the term $\Omega_{f(i)t}$, which is a firm-specific effect. It is composed by the size of the firm (total number of employees), average salary (in log) of the establishment per year, percentage of women working in the firm in that year and fixed effects of firm's activity sector according to CNAE (3 digits). It also includes a time-invariant firm fixed effect that does not change over time. See Section 3.

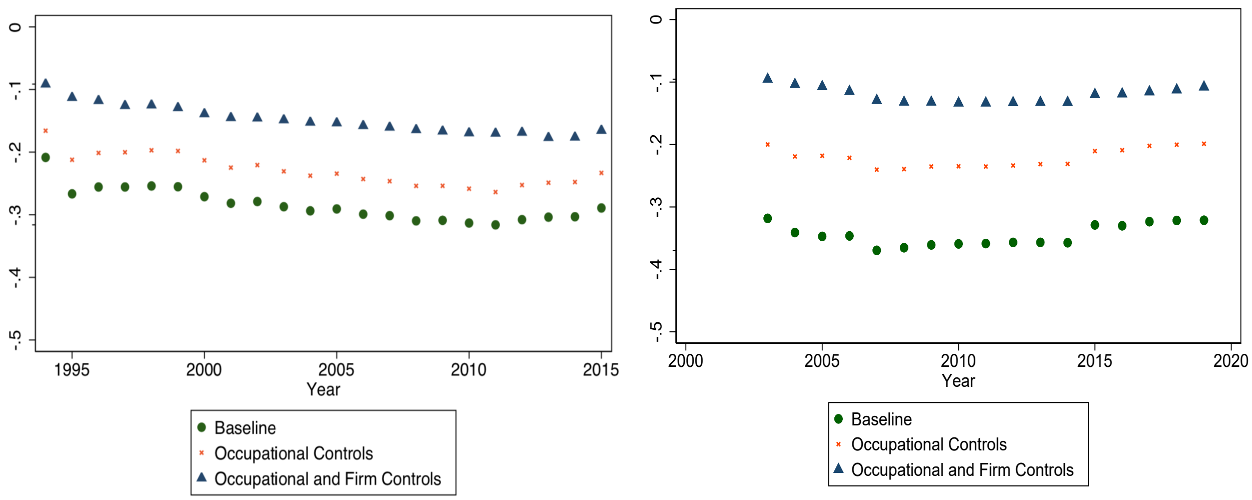
Figure 6: Relative Importance of the Controls in the Earnings Gender Gap

(a) Formal Labor Market,

period 1994-2015

(b) STEM Labor Market,

period 2003-2019



Source: Machado et al. (2018)

Figure 7: Earnings Gender Gap (in log differences) Throughout the Life Cycle by Educational Level: Baseline Model, STEM Labor Market

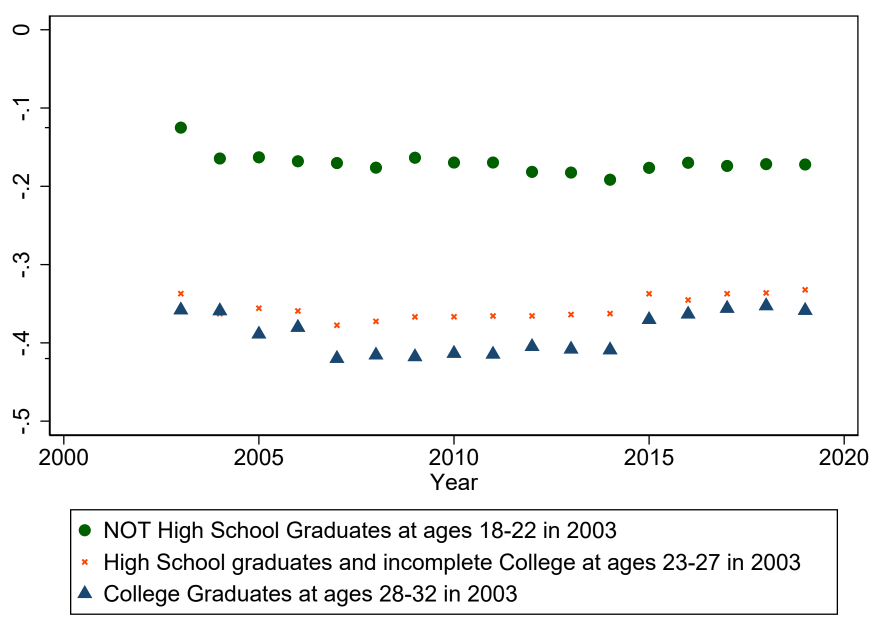


Figure 8: Earnings Gender Gap (in log differences) Throughout the Life Cycle by Educational Level: Occupational Controls, STEM Labor Market

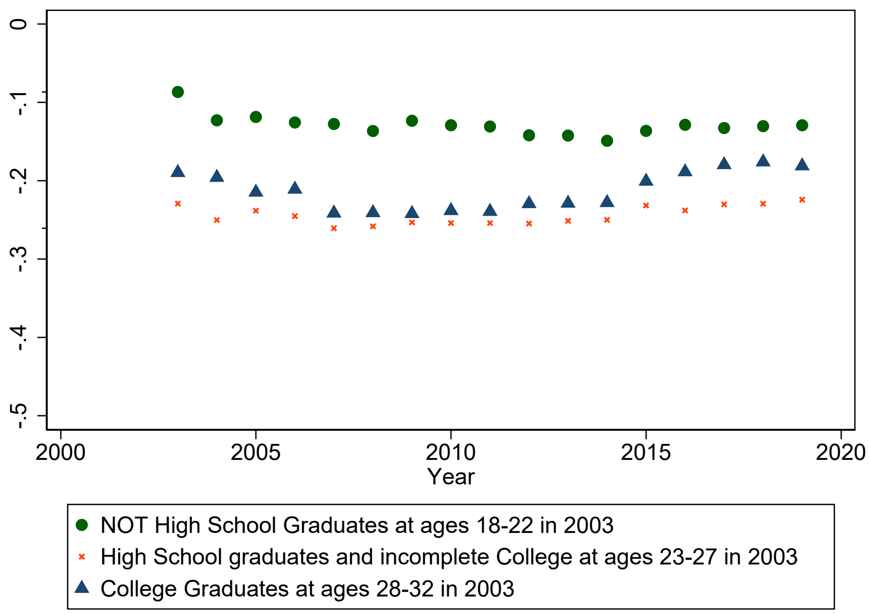
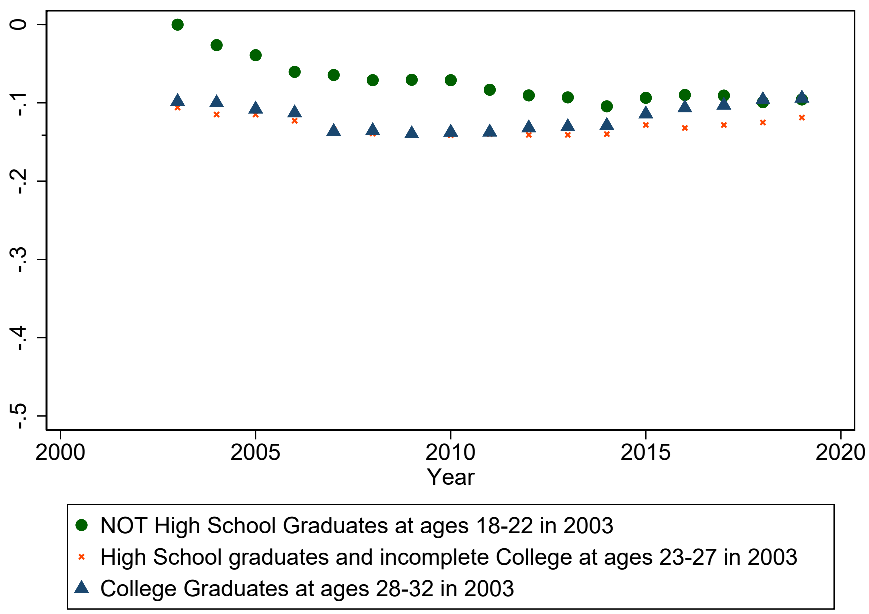


Figure 9: Earnings Gender Gap (in log differences) Throughout the Life Cycle by Educational Level: Occupational and Firm Controls, STEM Labor Market



baseline model. The gender wage gap is higher for the more educated workers (around 40% for high school graduates and up) and lower for the less educated workers (10-20% for not high school graduates). When adding occupational controls, in Figure 8, the gender earnings gap remains close to the baseline estimates for the less educated workers, but shifts up considerably (that is, the gap decreases) for workers with a high school degree or more. Indeed, the gap for the two higher education categories stays around 20% for all years estimates. Finally, as we control for firms characteristics (and occupational sector), there is another upward shift in the gender pay gap, this time for all educational groups (Figure 9). In fact, the difference of the gender pay gap between the three educational groups decreases, bringing together the gap of all the three groups. The gender pay gap of all educational groups become very similar by the end of life cycle, in particular, from 2015 onwards. In this period, the gender earnings gap is around 10% for all educational groups.

Although the results are not directly comparable to those of the formal job market because they treat different birth cohorts, we can see some similarities and differences in the relative importance of the controls for the gender pay gap of the three educational groups, though they must be interpreted with caution. First, the results are very similar for the baseline model and for the occupational controls specification, as we observe a greater gap (and of magnitude alike) for the higher educational groups and an upward shift for the same group when adding occupational controls. However, when controlling for firm characteristics, although the gap continues to decrease and approximate the coefficients of the three occupational groups in the STEM market, this movement is less strong for the formal labor market, so that the gender pay gap of 20 percent remains (especially for the group of college graduates). In the STEM market, adding firm controls equals the gap of the groups of the high school graduates and college graduates, while the gap of not high school graduates remains smaller. Furthermore, while the gap of the formal job market has almost parallel trajectories for all educational levels, there is a convergence in the trajectories of all the educational groups in the STEM market that is not observed in the formal market.

5.3 Sorting: Mean Establishment Earnings

Figure 10 shows the evolution of the estimated ϕ_j in Equations 1 and 5, along with three other estimations in which we add different sets of control variables: (i) $ln(MEE)$; (ii) firm size and fraction of women with establishment fixed effects; and (iii) the combination of (i) and (ii) without establishment fixed effects.[11]

The baseline model estimates show a relatively stable gender earnings gap from 2003 to 2019, ranging between 32-37% in the period. Equation (2), whose coefficients are represented by the squares in the graph, shows a stable gap of around 7-15% throughout the entire period. This means that occupation, industry, and firm controls account for about two-thirds to 80% of the earnings gap in the period.

To understand the role of the different components of equation (2), we include each control separately to the baseline model, or Equation (1). The graph in Figure 10 shows that estimations including only $ln(MEE)$, other firm controls, and these two combined generate very similar estimators to those in equation (2). Because the results for equation (2) are almost identical to those including occupation, industry, and firm controls, we will hence refer to equation (2) as adding ln(MEE).

This suggests that the addition of mean establishment earnings is the main factor driving the upward shift of ϕ_j between equations (1) and (2). With the addition of ln(MEE), all gender earnings gap estimates shift. This indicates that men in STEM occupations are disproportionately employed in higher-wage establishments.

Comparison with Other Sectors.

In Figures 11 to 14, we compare the results of equations (1) and (5) with analogous estimations for occupations in other sectors. Figure 11 compares the STEM gender earnings gap with management occupations, figure 12 compares with health occupations, 13 with law occupations and figure 14 with social science occupations.

Management occupations present an expanding gender gap, with the largest baseline gap among the analyzed occupations, reaching almost 78% in 2014. Even when we include MEE controls, the gap is still larger than the baseline estimates of STEM occupations. The gap

[11]Establishment fixed effects had to be omitted in this specification due to perfect multicollinearity with mean establishment earnings.

Figure 10: Relative Importance of *MEE* in the Earnings Gender Gap in STEM Occupations

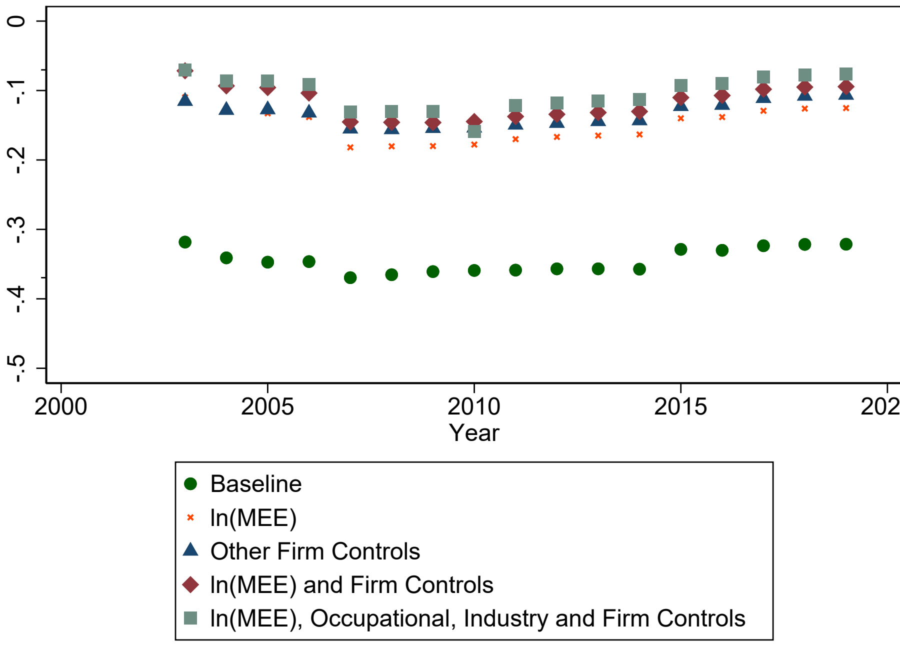
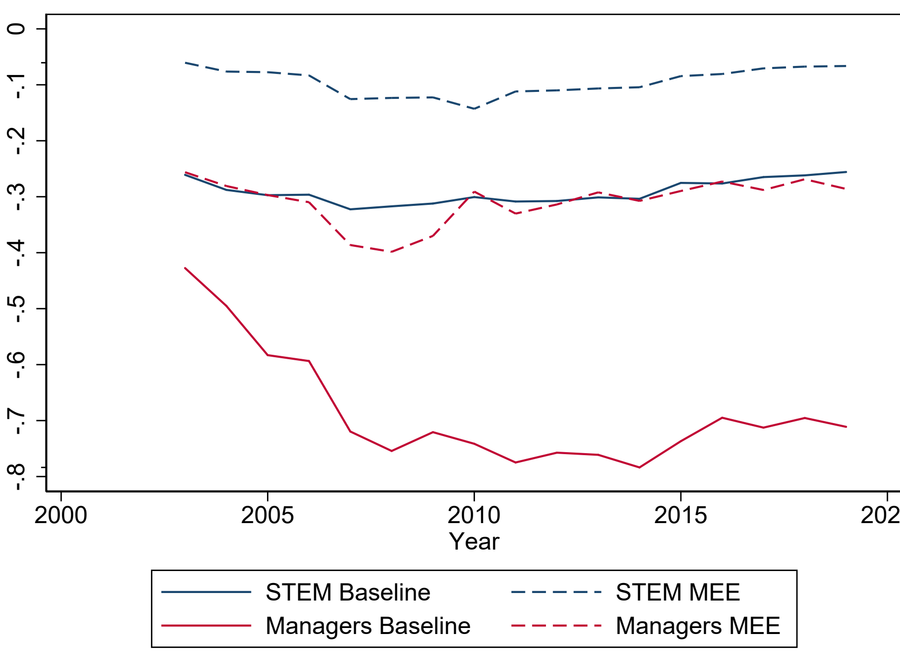


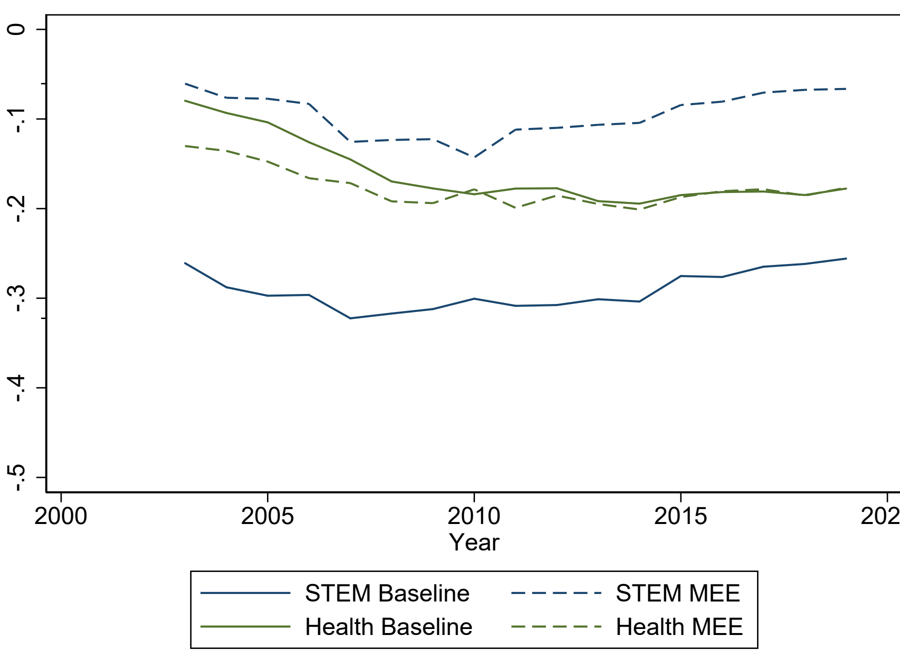
Figure 11: Gender Earnings Gap in STEM and Managers Occupations, 2003-2019



with $ln(MEE)$ controls increases from 25% in 2003 to 28% in 2019, reaching its maximum value in 2008 (39%). Mean establishment earnings contribute to about half of the gender earnings gap throughout the period.

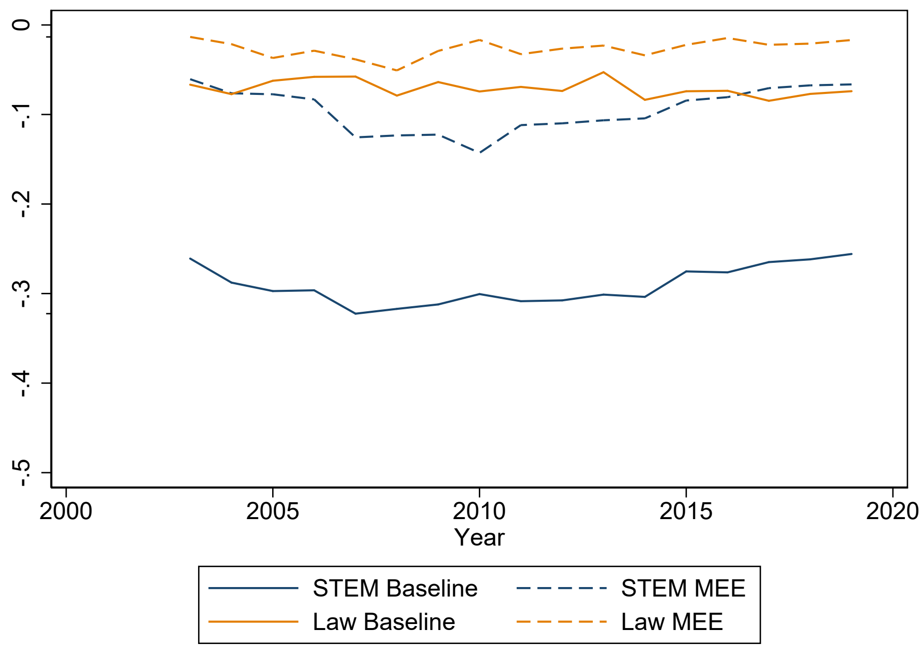
What draws the attention of the gender earnings gap in management occupations is that the gap widens a lot in the baseline estimation, which is not accompanied by the gap given mean establishment earnings. This means that a small part of the widening gap in these occupations is due to men shifting disproportionately into higher-paying establishments. However, most of the gender earnings gap can be attributed to women's lower ability to increase their salaries within firms. Finally, most widening occurs in the first five years of our age range, when most families are being formed. On the other hand, this is not observed for STEM occupations. This suggests that men sort disproportionately into higher-wage firms in STEM occupations from the moment they enter the job market, and this sorting remains throughout the entire period.

Figure 12: Gender Earnings Gap in STEM and Health Occupations, 2003-2019



Health and Law occupations show a small baseline gender earnings gap. In fact, Health occupations had an expanding gap of 8% in 2003, increasing to 18% in 2019. Furthermore, the estimates of equations (1) and (2) for health occupations are very similar, with a slight

Figure 13: Gender Earnings Gap in STEM and Law Occupations, 2003-2019

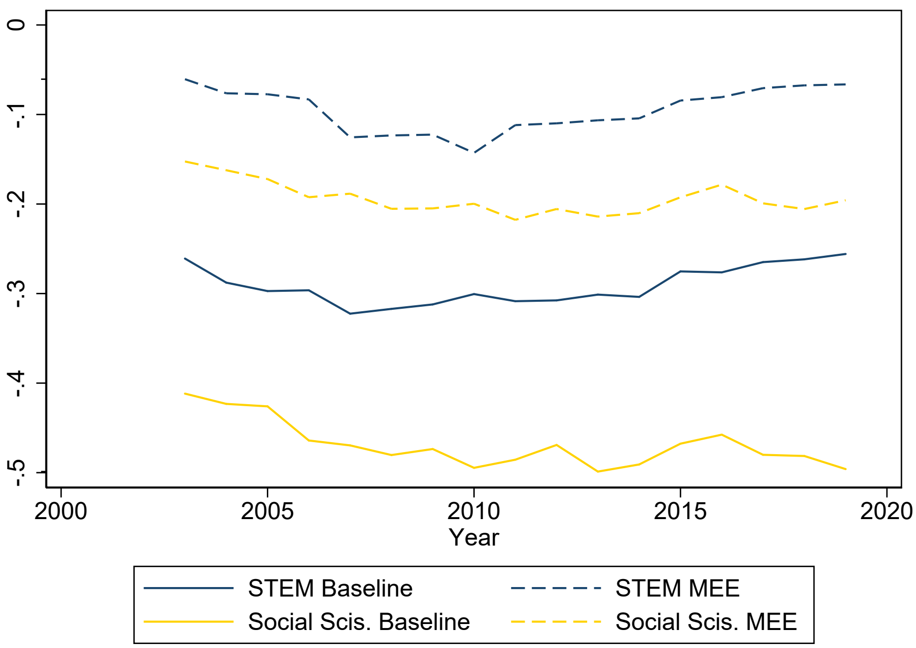


inversion of the gaps at the beginning of the period. It is the only time and occupation for which the gap of the $ln(MEE)$ estimation is larger than the baseline gap. This inversion happens until 2012, when the gaps of the two estimates equalize. This suggests that more women are employed in higher-wage establishments early in their careers, but the proportion of men and women allocated in higher-wage establishments balances over time. The widening in the first years of employment in health occupations is likely due to men shifting towards higher wage establishments or women shifting towards lower-wage establishments.

For Law occupations, we also have that the estimates of equations (1) and (2) are very similar, with a slightly smaller gap for the $ln(MEE)$ estimates. These occupations have the smallest gap of the period, between 4 and 9% for both estimates. Although there is a small difference in percentage points between the two estimations, mean establishment earnings explain about a third to half of the gender pay gap. However, the proportion of the wage gap explained by $ln(MEE)$ is still lower than that of other occupations, indicating that the proportion of men and women allocated in high/low wage establishments in Law occupations is more egalitarian in relation to the other occupations analyzed.

Social Sciences occupations are, along with STEM occupations, the occupations in which

Figure 14: Gender Earnings Gap in STEM and Social Sciences Occupations, 2003-2019



the $ln(MEE)$ contributes the most proportionately to the earnings gender gap in the period. While the baseline model shows an expanding earnings gap of 41 to 50% in the period, equation (2) shows the coefficient increased from 13 to 18%. This means that mean establishment earnings account for about 60% of the gender earnings gap. Even when controlling for mean establishment earnings, the gap is more significant than STEM occupations.

Figure 15a and 15b summarize what was presented in the other figures. All together, STEM occupations have a higher gender gap in the baseline model than Health and Law occupations and a smaller gender gap than Social Sciences and Management Occupations. When controlling for mean establishment earnings, Law occupations are the only group with a smaller gender gap than STEM occupations.

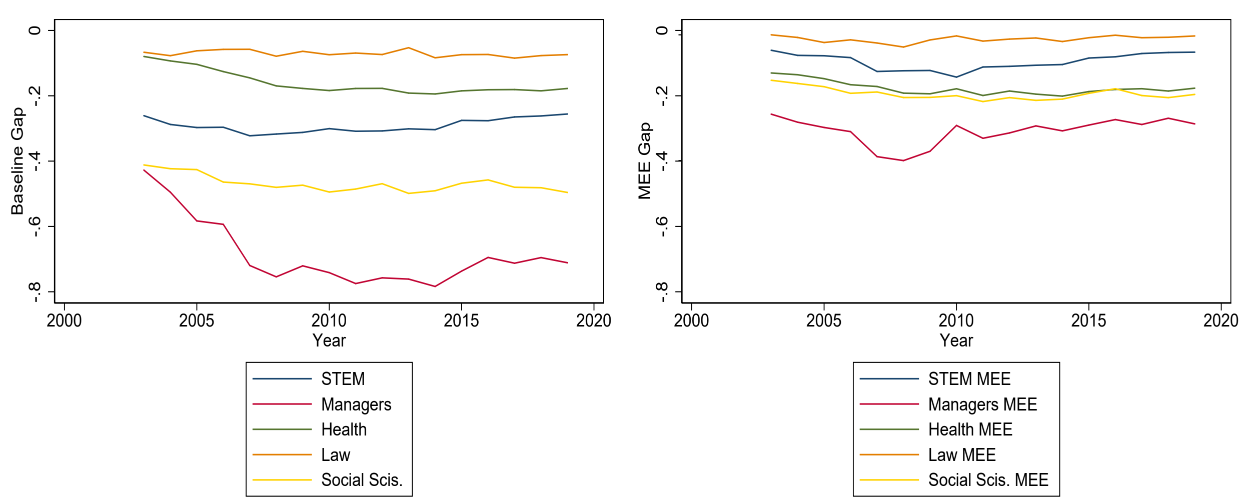
Evaluating the importance of mean establishment earnings in explaining the gender pay gap in various occupations, we observe that the relative importance of $ln(MEE)$ is greater for STEM, management, and social sciences occupations. In these occupations, mean establishment earnings explain from 50% to 70% of the gender pay gap, being higher in STEM and Social Sciences. This means that men in STEM, management, and social sciences occupations are disproportionately more employed in higher-wage establishments. On the other hand, for health and law occupations, mean establishment earnings explain little or nothing about the gender pay gap, meaning that men and women are more equally allocated in higher-wage establishments.

More revealing is that when we control for mean establishment earnings, the STEM gender earnings gap shifts up without changing the slope so that the gap remains stable in both specifications. This means that the sorting of men into higher-wage establishments occurs when they enter the job market and remains throughout all the analyzed years.

Another thing that stands out in our findings is that they are substantially different from the reported by Goldin et al. (2017) for the United States. While the author documents a widening gap during the first decade and a half after schooling ends, particularly for college graduates, we find a stable gap in STEM occupations over the analyzed period. Even when the author looks exclusively at the Technology sector, widening the gap remains. Another differential of our results concerning the United States is how mean establishment earnings explain the gender earnings gap. While in technology occupations in the US the MEE explains between 24 to 30 percent of the initial widening, in STEM occupations in Brazil, this value is much higher, between 60 and 80% of the stable gender earnings gap.

Figure 15: Gender earnings gap in various occupations, 2003-2019

(a) Baseline (b) ln(MEE)



5.4 Earnings Gender Gap Throughout the Life Cycle

Another way to see changes in the earnings gender gap by age is to construct synthetic birth cohorts (Goldin, 2014). This was documented previously for the Brazilian formal labor market by Machado et al. (2018). As our time interval is different from their work[12], we select different cohorts for tracking the evolution of the earnings gap over the life cycle for STEM occupations, and therefore our results cannot be directly compared to the rest of the formal market. The evolution of the earnings gender gap throughout the life cycle in the formal labor market can be consulted in Figure A.1, in the Appendix B.

We track the evolution of the earnings gap over the life cycle for selected cohorts. We selected cohorts born in 1959-1963, 1964-1968, 1969-1973, 1974-1978, 1979-1983, 1984-1988, 1976-1983, following workers employed in STEM occupations at some point during the period 2003-2019 (Figure 16).

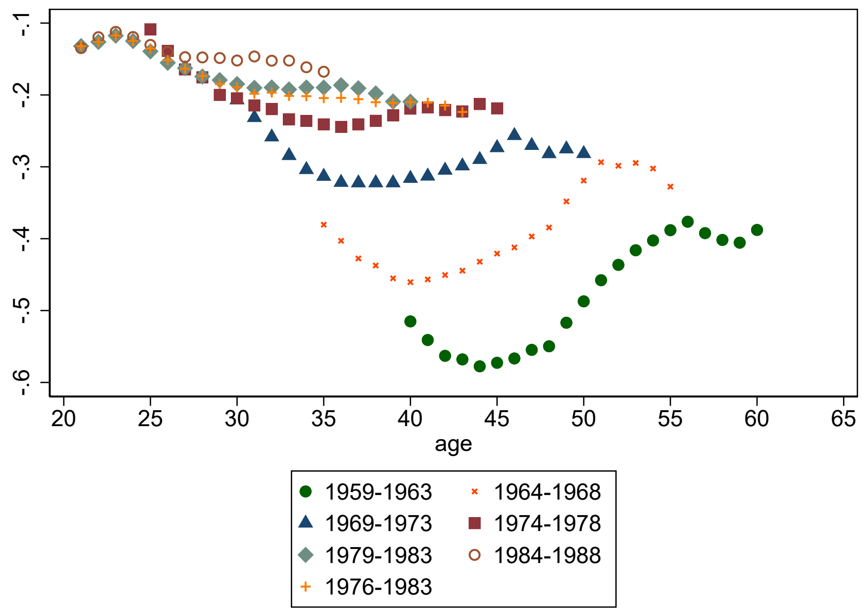
The figure shows that the gap expands throughout the life cycle until age 45, when it starts to narrow. This decrease is more drastic towards the end of the career for older cohorts. In the older cohorts, there is also a small increase in the gap in the last years of the career, which is not observed in the formal labor market. The difference in the gap between older cohorts is more pronounced than that of more recent cohorts, so the gap has been narrowing and stabilizing around 10-20% among younger cohorts in the STEM market.

Heterogeneous Results by Race.

To find out whether the results differ by race, we divided the cohorts into two groups: whites and non-whites. It is important to note that race-color information is noisy in *RAIS*, with a high number of missing observations (7.5% of the cohort sample), which were dropped. In selected cohorts of STEM workers, 62.9% of observations were from white workers, and the remaining 37.1% were from non-white workers. Figures 17a and 17b show the life cycle of

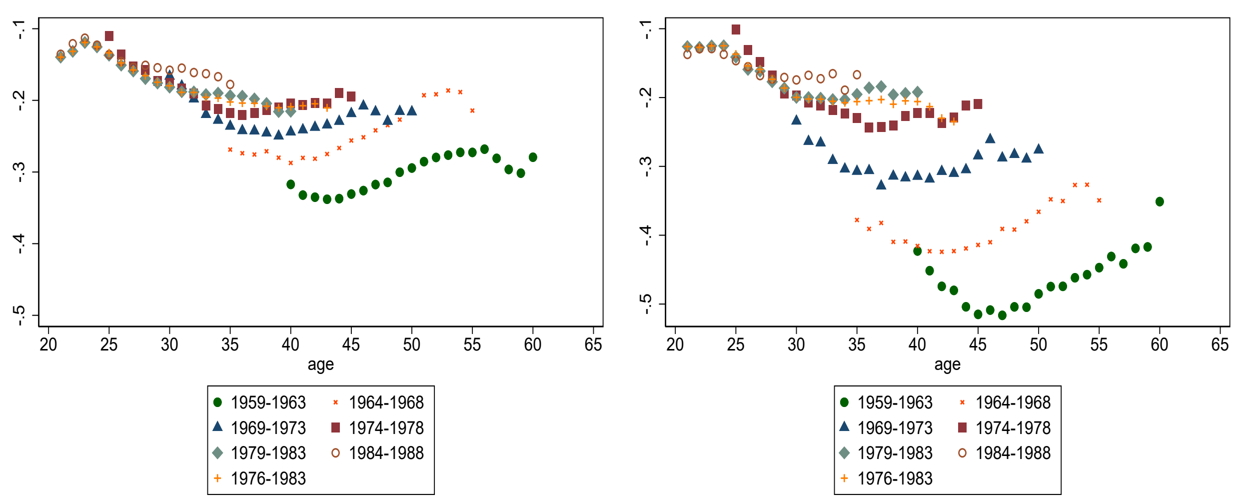
[12] Machado et al. (2018) use selected cohorts from 1950-1954, 1955-1959, 1960-1964, 1965-1969, 1970-1974, 1975-1979 and 1967-1974, to track workers formally employed at some point during the period 1994-2015 (see Figure A.1 in the Appendix). To maintain consistency with the previous analysis, we kept age differences constant in relation to the first year of the sample. That is, workers born in 1950-1954 were between 40-44 years old in 1994. Likewise, workers born in 1959-1963 were also between 40-44 years old in 2003. The age differences for the other cohorts are analogous.

Figure 16: Evolution of the Earnings Gender Gap (in log differences) throughout the Life Cycle by Selected Birth Cohorts, STEM Labor Market



the earnings gender gap in log differences for cohorts in the STEM market, for white workers and non-white workers, respectively.

Figure 17: Evolution of the Earnings Gender Gap (in log differences) throughout the Life Cycle by Selected Birth Cohorts by Race, STEM Labor Market



(a) White Workers

(b) Non-white Workers

The figures show that the evolution of the earnings gender gap over the life cycle between white and non-white workers is very similar. For both groups, the gap widens throughout the life cycle until about age 45, when it starts to narrow. As in the STEM market in general, for white and non-white workers the gap of the most recent cohorts is very similar and smaller, and the differences between the gaps are mainly verified among the older cohorts. The main difference between the two groups is that the increase in the gap at the end of the career verified in the older STEM cohorts is only seen among white workers.

6 Conclusion

In this paper, we use data on the Brazilian formal labor market and a classification of STEM occupations to document gender disparities over participation, hours worked and earnings in the Brazilian STEM labor market.

First, we document that there is convergence trend in earnings in the STEM labor market that is not accompanied by a reduction the participation gap. However, unlike the formal labor market, the STEM market experiences an increase in the ratio women to men of full time workers, consistent with an increase in the intensive margin of labor supply of women in STEM occupations.

Second, we document the transition dynamics in STEM and non-STEM occupations. The likelihood of a STEM worker being formally employed in the succeeding years is 20 percentage points higher than the average formal worker. Nevertheless, the likelihood of being in a STEM job in the succeeding years is 40 percentage points lower for men and 50 percentage points lower for women in most of the periods. This suggests not only that there is little stability in holding a STEM job, but also that women have a harder time holding those jobs for multiple years.

Third, we investigate the role of occupations and firms in explaining the gender differentials. Occupational controls and firms' characteristics explain each around 1/3 of the gender pay gap. We argue this could be due to sorting of workers into high and low-paying firms and occupations within the STEM field. We then decompose how much of the gender earnings gap in STEM occupations is explained due to shifting employment across establishments differing in mean earnings and how much is due to differential wage growth within establishments.

We find that mean establishment earnings (MEE) explain from half to 2/3 of the gender earnings gap in STEM occupations. This upward shift in the gender earnings gap occurs without changing its stability trend, suggesting that men sort into higher paying establishments when STEM workers enter the job market, and that this sorting remains throughout all the analyzed years.

Future research is still needed to investigate the reasons that lead this assortment of men to go to higher paying establishments already in the first years of their careers, that could not be studied in this work due to the restrictions of our database. Among the possibilities, women could be choosing establishments that offer advantages other than salary, such as flexible hours, diversity and inclusion in the work environment and other non-cash benefits.

These factors increase with women's greater family responsibilities. Barriers to women's participation in the labor force include the lack of policies supporting work-life balance, such as care services (Todd, 2013; Blau and Kahn, 2016), the incidence of cultural factors (Morton et al., 2014), employer discrimination (Arceo-Gomez and Campos-Vázquez, 2014; Galarza and Yamada, 2014; Bohnet, 2016), and the role of soft skills, such as women's lower propensity to negotiate compared with their male peers (Bohnet, 2016). Addressing the myriad obstacles to employment that women face requires multiple public policy actions; however, there is still only limited robust empirical evidence regarding what is the most effective intervention. Among the possibilities highlitghed in the literature, policies of shared responsibility for care seek to relax limitations on women's time, given the potential conflict that exists between time demanded for work and care activities. Maternity/paternity leave is one example (Baum and Ruhm, 2016; Berger and Waldfogel, 2004; Espinola-Arredondo and Mondal, 2008; Han, Ruhm, and Waldfogel, 2009).

References

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, *55*(3), 789–865.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Goldin, C., Kerr, S. P., Olivetti, C., & Barth, E. (2017). The expanding gender earnings gap: Evidence from the lehd-2000 census. *American Economic Review*, *107*(5), 110–14.

Machado, C., Neri, M., & Neto, V. P. (2018). The gender pay gap and the life cycle profile in the brazilian formal labor market. *Working Paper*.

Machado, C., Rachter, L., Stussi, M., & Schanaider, F. (2021). *Stem classification in the formal labor market in brazil.*

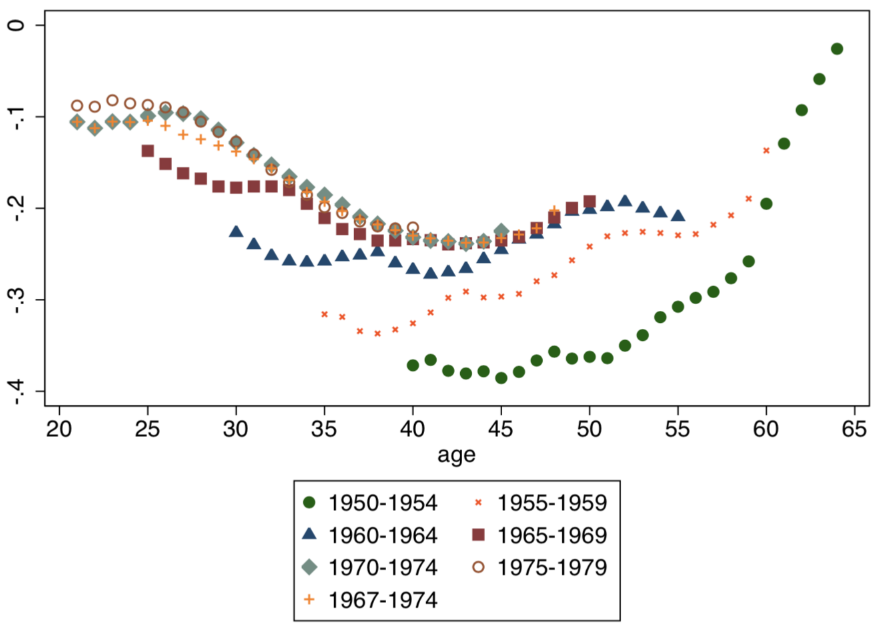
Menezes-Filho, N. A., & Muendler, M.-A. (2011). *Labor reallocation in response to trade reform* (Tech. Rep.). National Bureau of Economic Research.

USBLS. (2012a). *Attachment a: Soc policy committee recommendation to omb. options for defining stem (science, technology, engineering, and mathematics) occupations under the 2010 standard occupational classification (soc) system.* https://www.bls.gov/soc/Attachment_A_STEM.pdf.

USBLS. (2012b). *Attachment c: Detailed 2010 soc occupations included in stem. options for defining stem (science, technology, engineering, and mathematics) occupations under the 2010 standard occupational classification (soc) system.* https://www.bls.gov/soc/Attachment_C_STEM.xls.

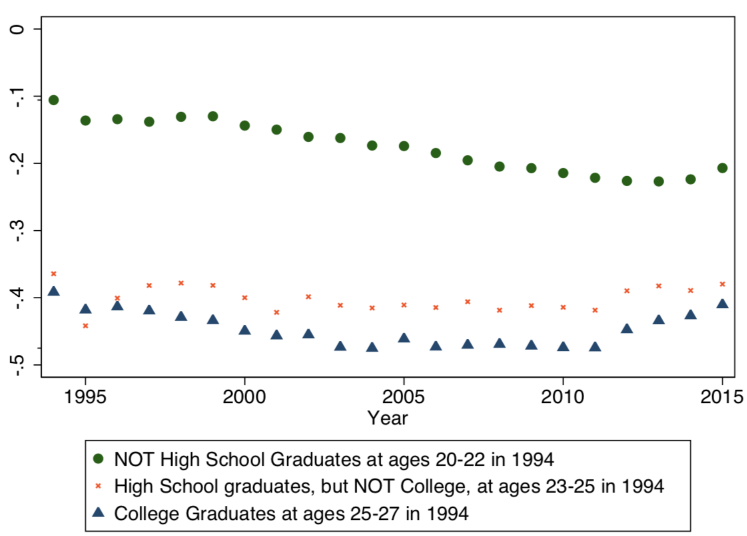
A Appendix

Figure A.1: Evolution of the Earnings Gender Gap (in log differences) throughout the Life Cycle by Selected Birth Cohorts, Formal Labor Market



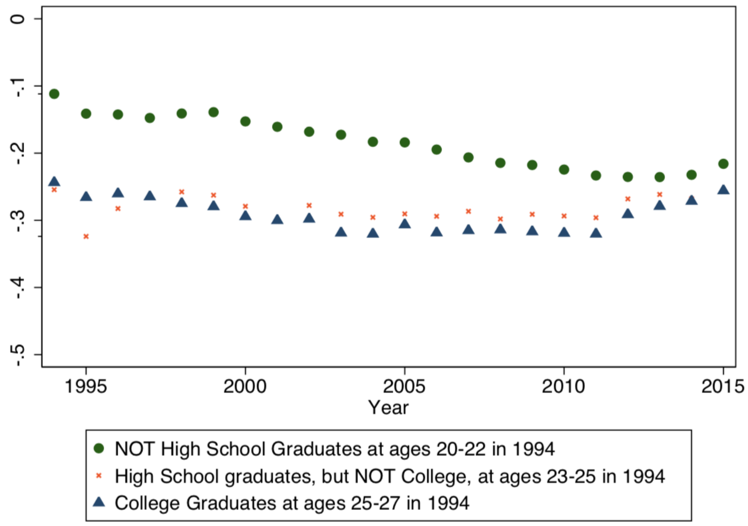
Source: Machado et al. (2018)

Figure A.2: Earnings Gender Gap (in log differences) Throughout the Life Cycle by Educational Level: Baseline Model, Formal Labor Market



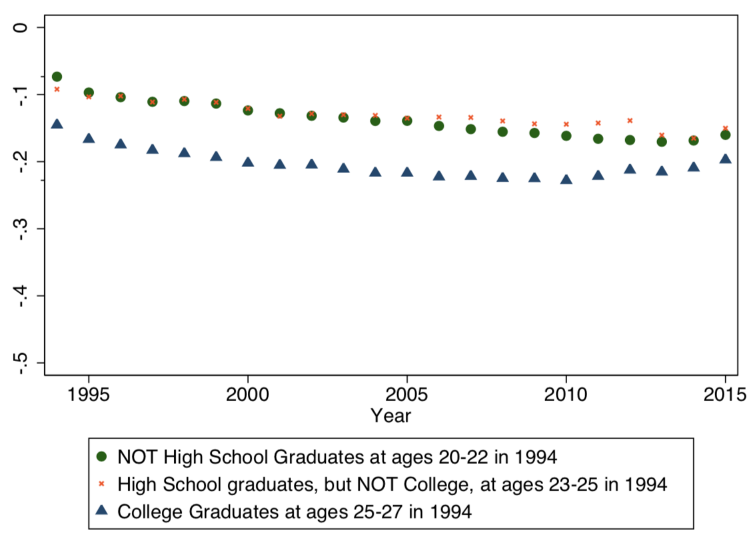
Source: Machado et al. (2018)

Figure A.3: Earnings Gender Gap (in log differences) Throughout the Life Cycle by Educational Level: Occupational Controls, Formal Labor Market



Source: Machado et al. (2018)

Figure A.4: Earnings Gender Gap (in log differences) Throughout the Life Cycle by Educational Level: Occupational and Firm Controls, Formal Labor Market



Source: Machado et al. (2018)

B Tables

Table 1: Gender pay gap (in log differences) throughout the life cycle controlling for occupation and firm characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Equation 1	Equation 2						Equation 3
Variables	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)
2003 $\times F$	-0.327***	-0.208***	-0.213***	-0.084***	-0.082***	-0.162***	-0.112***	-0.104***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
2004 $\times F$	-0.348***	-0.225***	-0.231***	-0.098***	-0.115***	-0.175***	-0.123***	-0.111***
	(0.003)	(0.003)	(0.003)	-0.002	(0.003)	(0.003)	(0.002)	(0.002)
2005 $\times F$	-0.353***	-0.223***	-0.243***	-0.110***	-0.120***	-0.173***	-0.122***	-0.115***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2006 $\times F$	-0.352***	-0.227***	-0.241***	-0.120***	-0.129***	-0.172***	-0.127***	-0.123***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2007 $\times F$	-0.376***	-0.246***	-0.276***	-0.134***	-0.152***	-0.192***	-0.149***	-0.135***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2008 $\times F$	-0.372***	-0.245***	-0.270***	-0.139***	-0.155***	-0.192***	-0.151***	-0.138***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2009 $\times F$	-0.367***	-0.241***	-0.267***	-0.141***	-0.154***	-0.189***	-0.148***	-0.138***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2010 $\times F$	-0.366***	-0.241***	-0.264***	-0.146***	-0.155***	-0.292***	-0.148***	-0.139***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2011 $\times F$	-0.365***	-0.241***	-0.264***	-0.149***	-0.158***	-0.189***	-0.144***	-0.139***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2012 $\times F$	-0.363***	-0.239***	-0.262***	-0.150***	-0.156***	-0.187***	-0.142***	-0.139***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2013 $\times F$	-0.362***	-0.236***	-0.262***	-0.150***	-0.154***	-0.183***	-0.139***	-0.138***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2014 $\times F$	-0.362***	-0.235***	-0.261***	-0.151***	-0.153***	-0.183***	-0.137***	-0.138***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2015 $\times F$	-0.334***	-0.214***	-0.227***	-0.139***	-0.133***	-0.162***	-0.116***	-0.125***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2016 $\times F$	-0.334***	-0.213***	-0.229***	-0.138***	-0.130***	-0.157***	-0.114***	-0.124***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2017 $\times F$	-0.328***	-0.206***	-0.221***	-0.136***	-0.123***	-0.151***	-0.105***	-0.120***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2018 $\times F$	-0.326***	-0.204***	-0.219***	-0.133***	-0.121***	-0.147***	-0.102***	-0.117***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
2019 $\times F$	-0.325***	-0.203***	-0.218***	-0.130***	-0.121***	-0.147***	-0.100***	-0.112***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
ln(size)			0.084***					0.026***
			(0.000)					(0.001)
ln(mp)				0.769***				0.713***
				(0.000)				(0.001)
% Female					-0.383***			0.061***
					(0.002)			(0.003)
Observations	4,942,383	4,942,383	4,942,383	4,942,383	4,942,383	4,942,383	4,811,803	4,811,803
R-squared	0.330	0.410	0.454	0.722	0.417	0.482	0.748	0.767
Education*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm's Sector	No	No	No	No	No	Yes	No	Yes
Firms FE	No	No	No	No	No	No	Yes	Yes

Standard errors in parenthesis

***p< 0.01, **p< 0.05, *p< 0.1

Table 2: Educational heterogeneity on gender pay gap (in log differences) throughout the life cycle controlling for occupation and firm characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Specification	Baseline Model (BM)			BM plus Occupation			BM plus Occupation and Firms		
Variables	<HS	HS	>HS	<HS	HS	>HS	<HS	HS	>HS
Educ $\times 2003 \times F$	-0.130***	-0.351***	-0.359***	-0.088***	-0.241***	-0.192***	0.002	-0.116***	-0.107***
	(0.008)	(0.004)	(0.005)	(0.007)	(0.003)	(0.005)	(0.006)	(0.002)	(0.003)
Educ $\times 2004 \times F$	-0.168***	-0.373***	-0.360***	-0.123***	-0.260***	-0.198***	-0.024***	-0.123***	-0.107***
	(0.009)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2005 \times F$	-0.165***	-0.364***	-0.389***	-0.117***	-0.246***	-0.217***	-0.037***	-0.124***	-0.116***
	(0.010)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2006 \times F$	-0.170***	-0.369***	-0.381***	-0.124***	-0.254***	-0.213***	-0.060***	-0.132***	-0.120***
	(0.010)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2007 \times F$	-0.172***	-0.388***	-0.420***	-0.126***	-0.270***	-0.244***	-0.064***	-0.142***	-0.141***
	(0.009)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2008 \times F$	-0.179***	-0.383***	-0.416***	-0.136***	-0.268***	-0.243***	-0.070***	-0.147***	-0.140***
	(0.009)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2009 \times F$	-0.166***	-0.378***	-0.418***	-0.122***	-0.262***	-0.244***	-0.070***	-0.145***	-0.144***
	(0.009)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2010 \times F$	-0.173***	-0.377***	-0.414***	-0.128***	-0.263***	-0.240***	-0.071***	-0.149***	-0.141***
	(0.009)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2011 \times F$	-0.173***	-0.376***	-0.415***	-0.130***	-0.263***	-0.242***	-0.083***	-0.148***	-0.141***
	(0.009)	(0.004)	(0.005)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2012 \times F$	-0.184***	-0.376***	-0.405***	-0.141***	-0.263***	-0.232***	-0.090***	-0.149***	-0.136***
	(0.009)	(0.004)	(0.006)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2013 \times F$	-0.185***	-0.373***	-0.408***	-0.141***	-0.259***	-0.232***	-0.093***	-0.148***	-0.135***
	(0.009)	(0.004)	(0.006)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2014 \times F$	-0.194***	-0.371***	-0.409***	-0.148***	-0.256***	-0.231***	-0.104***	-0.147***	-0.133***
	(0.009)	(0.004)	(0.006)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.003)
Educ $\times 2015 \times F$	-0.178***	-0.345***	-0.371***	-0.135***	-0.237***	-0.203***	-0.093***	-0.135***	-0.118***
	(0.009)	(0.004)	(0.006)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.004)
Educ $\times 2016 \times F$	-0.172***	-0.352***	-0.364***	-0.127***	-0.243***	-0.191***	-0.090***	-0.138***	-0.111***
	(0.010)	(0.004)	(0.006)	(0.009)	(0.004)	(0.005)	(0.006)	(0.003)	(0.004)
Educ $\times 2017 \times F$	-0.175***	-0.344***	-0.356***	-0.130***	-0.235***	-0.182***	-0.091***	-0.134***	-0.107***
	(0.010)	(0.004)	(0.006)	(0.009)	(0.004)	(0.006)	(0.006)	(0.003)	(0.004)
Educ $\times 2018 \times F$	-0.173***	-0.343***	-0.353***	-0.128***	-0.234***	-0.179***	-0.099***	-0.131***	-0.100***
	(0.010)	(0.004)	(0.006)	(0.009)	(0.004)	(0.006)	(0.006)	(0.003)	(0.004)
Educ $\times 2019 \times F$	-0.173***	-0.339***	-0.359***	-0.126***	-0.230***	-0.184***	-0.096***	-0.125***	-0.098***
	(0.010)	(0.005)	(0.006)	(0.009)	(0.004)	(0.006)	(0.007)	(0.003)	(0.004)
			(0.006)			(0.006)			(0.004)
ln(size)							0.026***	0.026***	0.026***
							(0.001)	(0.001)	(0.001)
ln(mp)							0.713***	0.713***	0.713***
							(0.001)	(0.001)	(0.001)
% Women							0.059***	0.059***	0.059***
							(0.003)	(0.003)	(0.003)
Observations		4,942,383			4,942,383			4,811,803	
R-squared		0.331			0.411			0.767	
Education*Year FE		Yes			Yes			Yes	
Occupation		No			Yes			Yes	
Firm's Sector		No			No			Yes	
Firms FE		No			No			Yes	

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Gender pay gap (in log differences) throughout the life cycle controlling for mean establishment earnings and firm characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification	Equation 1							Equation 2
Variables	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)	ln(w)
2003 $\times F$	-0.318***	-0.110***	-0.336***	-0.083***	-0.154***	-0.103***	-0.101***	-0.070***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
2004 $\times F$	-0.341***	-0.129***	-0.359***	-0.131***	-0.169***	-0.115***	-0.114***	-0.086***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2005 $\times F$	-0.347***	-0.133***	-0.378***	-0.146***	-0.167***	-0.114***	-0.113***	-0.086***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2006 $\times F$	-0.346***	-0.138***	-0.372***	-0.154***	-0.166***	-0.119***	-0.118***	-0.091***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2007 $\times F$	-0.370***	-0.182***	-0.409***	-0.183***	-0.185***	-0.143***	-0.142***	-0.131***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2008 $\times F$	-0.365***	-0.180***	-0.400***	-0.186***	-0.185***	-0.144***	-0.144***	-0.130***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2009 $\times F$	-0.361***	-0.180***	-0.396***	-0.186***	-0.182***	-0.142***	-0.142***	-0.130***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2010 $\times F$	-0.359***	-0.178***	-0.393***	-0.187***	-0.287***	-0.142***	-0.140***	-0.159***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2011 $\times F$	-0.359***	-0.170***	-0.391***	-0.190***	-0.182***	-0.138***	-0.138***	-0.121***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2012 $\times F$	-0.357***	-0.167***	-0.389***	-0.189***	-0.181***	-0.136***	-0.135***	-0.118***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2013 $\times F$	-0.357***	-0.165***	-0.392***	-0.189***	-0.177***	-0.133***	-0.133***	-0.115***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2014 $\times F$	-0.357***	-0.163***	-0.392***	-0.190***	-0.178***	-0.132***	-0.131***	-0.113***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2015 $\times F$	-0.329***	-0.140***	-0.351***	-0.164***	-0.156***	-0.112***	-0.111***	-0.093***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2016 $\times F$	-0.330***	-0.138***	-0.356***	-0.162***	-0.153***	-0.110***	-0.109***	-0.090***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2017 $\times F$	-0.324***	-0.129***	-0.349***	-0.156***	-0.146***	-0.100***	-0.100***	-0.081***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2018 $\times F$	-0.322***	-0.126***	-0.346***	-0.154***	-0.142***	-0.097***	-0.097***	-0.078***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
2019 $\times F$	-0.321***	-0.125***	-0.347***	-0.155***	-0.142***	-0.096***	-0.095***	-0.076***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
ln(MEE)		0.803***						0.741***
		(0.000)						(0.000)
ln(size)			0.087***				0.006***	-0.004***
			(0.000)				(0.001)	(0.000)
% Women				-0.553***			-0.033***	-0.032***
				(0.001)			(0.002)	(0.001)
	(0.000)	(0.003)	(0.001)	(0.001)	(0.000)	(0.000)	(0.003)	(0.003)
Observations	6,256,977	6,011,888	6,256,977	6,256,977	6,256,977	6,112,995	6,112,995	6,011,888
R-squared	0.298	0.674	0.350	0.314	0.470	0.753	0.753	0.705
Education*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	No	No	No	No	Yes	Yes	Yes	Yes
Firm's Sector	No	No	No	No	Yes	No	Yes	Yes
Firms FE	No	No	No	No	No	Yes	Yes	No

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Occupational Heterogeneity on the gender pay gap (in log differences) throughout the life cycle controlling for mean establishment earnings and firm characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Specification			Baseline Model (BM)					BM plus ln(MEE)		
Variables	STEM	Management	Health	Law	Social Scis.	STEM	Management	Health	Law	Social Scis.
Occup $\times 2003 \times F$	-0.261***	-0.427***	-0.080***	-0.067***	-0.412***	-0.061***	-0.256***	-0.130***	-0.013	-0.152***
	(0.003)	(0.011)	(0.007)	(0.015)	(0.005)	(0.002)	(0.009)	(0.005)	(0.010)	(0.004)
Occup $\times 2004 \times F$	-0.288***	-0.495***	-0.093***	-0.077***	-0.423***	-0.076***	-0.281***	-0.136***	-0.021*	-0.162***
	(0.003)	(0.015)	(0.008)	(0.016)	(0.006)	(0.002)	(0.011)	(0.005)	(0.011)	(0.005)
Occup $\times 2005 \times F$	-0.297***	-0.583***	-0.104***	-0.062***	-0.426***	-0.077***	-0.297***	-0.148***	-0.037***	-0.172***
	(0.003)	(0.017)	(0.008)	(0.017)	(0.007)	(0.002)	(0.013)	(0.006)	(0.012)	(0.005)
Occup $\times 2006 \times F$	-0.296***	-0.594***	-0.126***	-0.058***	-0.464***	-0.083***	-0.310***	-0.166***	-0.029**	-0.192***
	(0.003)	(0.018)	(0.008)	(0.018)	(0.007)	(0.002)	(0.013)	(0.006)	(0.012)	(0.005)
Occup $\times 2007 \times F$	-0.322***	-0.719***	-0.145***	-0.058***	-0.470***	-0.126***	-0.386***	-0.172***	-0.038***	-0.188***
	(0.003)	(0.019)	(0.009)	(0.018)	(0.008)	(0.002)	(0.014)	(0.006)	(0.012)	(0.005)
Occup $\times 2008 \times F$	-0.317***	-0.754***	-0.170***	-0.079***	-0.480***	-0.123***	-0.398***	-0.192***	-0.051***	-0.205***
	(0.003)	(0.020)	(0.009)	(0.018)	(0.008)	(0.002)	(0.015)	(0.006)	(0.013)	(0.006)
Occup $\times 2009 \times F$	-0.312***	-0.721***	-0.177***	-0.064***	-0.474***	-0.122***	-0.370***	-0.194***	-0.029**	-0.205***
	(0.003)	(0.021)	(0.009)	(0.019)	(0.008)	(0.002)	(0.015)	(0.006)	(0.013)	(0.006)
Occup $\times 2010 \times F$	-0.301***	-0.741***	-0.184***	-0.074***	-0.495***	-0.143***	-0.291***	-0.178***	-0.017	-0.200***
	(0.003)	(0.020)	(0.010)	(0.021)	(0.010)	(0.002)	(0.014)	(0.007)	(0.014)	(0.007)
Occup $\times 2011 \times F$	-0.309***	-0.775***	-0.178***	-0.069***	-0.486***	-0.112***	-0.330***	-0.199***	-0.033**	-0.218***
	(0.003)	(0.021)	(0.009)	(0.019)	(0.008)	(0.002)	(0.015)	(0.006)	(0.013)	(0.006)
Occup $\times 2012 \times F$	-0.308***	-0.757***	-0.177***	-0.074***	-0.469***	-0.110***	-0.314***	-0.185***	-0.026**	-0.206***
	(0.003)	(0.021)	(0.009)	(0.019)	(0.008)	(0.002)	(0.015)	(0.006)	(0.013)	(0.006)
Occup $\times 2013 \times F$	-0.301***	-0.761***	-0.192***	-0.053***	-0.499***	-0.106***	-0.292***	-0.195***	-0.023*	-0.214***
	(0.003)	(0.021)	(0.009)	(0.020)	(0.009)	(0.002)	(0.014)	(0.006)	(0.013)	(0.006)
Occup $\times 2014 \times F$	-0.304***	-0.784***	-0.194***	-0.084***	-0.491***	-0.104***	-0.307***	-0.201***	-0.034**	-0.210***
	(0.003)	(0.021)	(0.009)	(0.020)	(0.009)	(0.002)	(0.014)	(0.006)	(0.013)	(0.006)
Occup $\times 2015 \times F$	-0.275***	-0.737***	-0.185***	-0.074***	-0.468***	-0.084***	-0.290***	-0.187***	-0.022*	-0.192***
	(0.003)	(0.021)	(0.009)	(0.020)	(0.009)	(0.002)	(0.015)	(0.006)	(0.013)	(0.007)
Occup $\times 2016 \times F$	-0.276***	-0.695***	-0.181***	-0.074***	-0.458***	-0.081***	-0.273***	-0.181***	-0.015	-0.178***
	(0.003)	(0.022)	(0.009)	(0.020)	(0.009)	(0.002)	(0.015)	(0.006)	(0.013)	(0.007)
Occup $\times 2017 \times F$	-0.265***	-0.713***	-0.181***	-0.085***	-0.480***	-0.071***	-0.288***	-0.178***	-0.022	-0.199***
	(0.004)	(0.022)	(0.009)	(0.020)	(0.010)	(0.002)	(0.015)	(0.007)	(0.014)	(0.007)
Occup $\times 2018 \times F$	-0.262***	-0.695***	-0.185***	-0.077***	-0.481***	-0.067***	-0.269***	-0.185***	-0.021	-0.206***
	(0.004)	(0.021)	(0.010)	(0.021)	(0.010)	(0.002)	(0.015)	(0.007)	(0.014)	(0.007)
Occup $\times 2019 \times F$	-0.256***	-0.711***	-0.178***	-0.074***	-0.496***	-0.066***	-0.286***	-0.177***	-0.017	-0.196***
	(0.004)	(0.022)	(0.010)	(0.021)	(0.010)	(0.002)	(0.015)	(0.007)	(0.014)	(0.007)
ln(MEE)								0.741***		
								(0.000)		
Size								-0.004***		
								(0.000)		
% Women								-0.039***		
								(0.001)		
Observations			6,256,977					6,011,888		
R-squared			0.300					0.705		
Occ. Group*Year FE			Yes					Yes		
Occupation			No					Yes		
Firm's Sector			No					Yes		

Standard errors in parenthesis
***p< 0.01, **p< 0.05, *p< 0.1