



The Business School  
for the World®

Working Paper

2023/66/EPS

## Exploring the Role of Firms in the Decline of the Gender Wage Gap over the Past Two Decades

Marco G. Palladino

Banque de France, [marcoguidopalladino@gmail.com](mailto:marcoguidopalladino@gmail.com)

Alexandra Roulet

INSEAD, [alexandra.roulet@insead.edu](mailto:alexandra.roulet@insead.edu)

Mark Stabile

INSEAD, [mark.stabile@insead.edu](mailto:mark.stabile@insead.edu)

November 22, 2023

While the gender gap in firm pay premiums is well documented, the evidence on how it has evolved over time and its contribution to the decline in the gender wage gap is mixed. We leverage new comprehensive employer-employee data and three distinct methodologies to answer these questions. We find that 20% of the decline in the hourly wage gender gap between 2002 and 2019 in France comes from a convergence in the sorting of men and women across firms. This is driven by women increasingly sorting into industries hosting higher-wage firms, rather than gaining access to better firms within industries.

Keywords: Gender Wage Gap; Firm Pay Premiums; Sorting.

JEL Classification: J16, J31, J71

We would like to thank Pierre Cahuc, Kerstin Holzheu, Barbara Petrongolo, Roland Rathelot, Jean-Marc Robin and in particular Stéphane Bonhomme, Francois Gerard, Thibaut Lamadon, Christian Moser and Raffaele Saggio for their helpful comments. We thank seminar participants at Queen Mary University, Stockholm University, Princeton, Banque de France and Zurich UZH as well as participants at the 4<sup>th</sup> CEPR-Bank of Italy Labor Workshop, the Junior Micro Macro Labor Conference at Columbia University, INSEAD Brown Bag and Sciences Po PhD Lunch. The technology to access the data remotely is supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'avenir' programme (ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD).

Working Paper is the author's intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from [publications.fb@insead.edu](mailto:publications.fb@insead.edu)

Find more INSEAD papers at <https://www.insead.edu/faculty-research/research>

Copyright © 2023 INSEAD

## 1. Introduction

Understanding what drives differences in wages between women and men has been the subject of considerable economic research for decades (Blau and Kahn 2017; Goldin 2014; Olivetti and Petrongolo 2016). Recently, research examining the role that firms play in contributing to the gender wage gap aims to separate two important drivers: the extent to which women sort into lower-paying firms and the extent to which women extract less surplus than men within a given firm. In particular, the seminal work of Card, Cardoso, and Kline (2016) has adapted the Abowd, Kramarz, and Margolis (1999) (AKM from now on) framework to assess the impact of firm-specific pay policies on the gender wage gap. This methodology estimates wage regressions with both worker and firm fixed effects separately by gender in order to compute gender-specific firm effects which allow the researcher to compute a gender gap in firm pay premiums and to decompose it into a sorting or *between-firm* channel (what women would be paid if they worked at the firms where men work) and a bargaining or *within-firm* one (what women would be paid at these firms if they were men).

The empirical application in Card, Cardoso, and Kline (2016) (CCK from now on) uses Portuguese data, but several papers have since then leveraged this methodology to study other contexts (Bruns 2019; Casarico and Lattanzio forthcoming; Coudin, Maillard, and Tô 2018; Gallen, Lesner, and Vejlin 2019; Morchio and Moser 2019; Sorkin 2017). Although magnitudes vary across countries, results consistently show that sorting between firms explains most of the firm component of the gender wage gap (see Table A1 for a summary of selected estimates). In other words, gender differences in pay premiums are largely driven by men and women working in different firms rather than similar men and women being paid differently at the same firm.

While the gender gap in firm pay premiums is therefore well documented across jurisdictions, there is less definitive evidence on how this gap, and its decomposition

into between and within-firm differences, have evolved over time. Casarico and Lattanzio (forthcoming) shows that in Italy, the gender gap in firm pay premiums remained roughly stable between 1995 and 2015 with a decrease in the between component compensated by an increase in the within component, while Bruns (2019) shows on the contrary that in Germany the gender gap in firm pay premiums widened between 1995 and 2008, driven by an increase in both components, and especially in the between part. To contribute to this research question, we use a newly built and almost exhaustive linked employer-employee dataset for France and rely on three approaches to document time-series development. First, we estimate AKM regressions in three intervals of six years each – 2002-2007, 2008-2013, and 2014-2019 – as similarly done in Bruns (2019) and Casarico and Lattanzio (forthcoming). This way, we obtain a set of three-point estimates about the development of the contribution of firms to the gender wage gap. In principle, a researcher could use narrower time intervals to obtain more data points. However, the AKM specification is not suited for short panels – and particularly so when looking at men and women separately – because the shorter the panel, the fewer the movers observed and the more restrictive the estimation sample. Indeed, in a specification with both firm and worker fixed effects, firm effects are only identified for firms that are connected by worker mobility and *gender specific* firm effects are only correctly identified within the “dual connected set”, the intersection of the male and female largest connected sets. Therefore, to provide more granular time series evidence, we implement the rolling AKM (R-AKM) method of Lachowska et al. (2023) by estimating AKM on six-year panels for every possible starting year in our data. We then recover the gender gap in firm pay premiums and the between and within-firm components separately for each year of each panel estimation. Finally, we leverage insights from Bonhomme, Lamadon, and Manresa (2019) (BLM from now on). Instead of estimating firm effects for each firm, they propose to first group firms that are sufficiently similar and then estimate one effect for each of the resulting clusters. Reducing the dimensionality of

the estimation has several nice properties. The one particularly useful in our context is that this method makes it easier to estimate very short panels. We exploit the clustering approach to estimate the development of the contribution of firms to the gender wage gap over time using 2-year panels as in the BLM framework. This procedure works in the absence of the dual-connected set restriction and allows us to examine year-by-year changes in each component.

We show that there is a significant decline in the between-firm component of the gender gap in firm pay premiums since 2000. It accelerates around 2014 and explains roughly 20% of the decline in the unconditional gender wage gap over the period of analysis. Our results are remarkably consistent whether we use six-year panels in the AKM framework or the firm clustering approach with two-year panels. We then turn to understanding the mechanisms. 75% of the decrease observed in the between-firm component of the gender wage gap can be attributed to women increasingly sorting into industries that host higher-wage firms and only 25% to women gaining access to better firms within specific industries over time.

We also observe an increase in the within-firm component of the gap, which begins early in our period and flattens out by 2010. Overall, therefore, we don't observe a decline in the gap in firm pay premiums between the early 2000s and the late 2010s. Our estimation results are thus consistent with Casarico and Lattanzio (forthcoming). However, we argue that the observed rise in the within-firm component of the gender gap in pay premiums should be interpreted with caution. Indeed our analyses suggest that it is likely to be the result of the normalization assumption rather than due to an inherent economic mechanism. In order to make meaningful comparisons across genders within each period, firm effects need to be normalized. Following the literature, we do this relative to the restaurant sector. This normalisation only affects the within-firm component of the gender gap in pay premiums and, as shown by Gerard et al. (2021), provides a conservative, lower bound estimate. Further ensuring the comparability

of within-firm estimates over time requires the additional assumption that the pay-setting behaviour for men and women within this industry remains consistent across the different estimation periods. The results are sensitive to this assumption and in calculating an upper bound we can no longer be statistically sure that the within-firm component has increased over our time period.

The rest of the paper proceeds as follows: Section 2 describes the data. Section 3 presents the estimation framework. Section 4 shows the results, and section 5 concludes.

## 2. Data

Our dataset is derived from the matched employer-employee registers in France known as DADS data. This comprehensive dataset provides valuable information on workers' employment, including their earnings, their hours of work, their firm and other administrative data for each of their jobs. The data is pseudonymous, with individuals being assigned unique codes that change annually, enabling cross-sectional analysis. However, it does not allow for long-term panel analysis for workers. Traditionally, panel analysis of workers in France has been conducted using the *DADS Panel*. This panel consists of a sample of individuals who are followed over time, with a sampling frequency of 1/24 before 2002 and 1/12 after.

To enhance our analysis, we utilize a recently constructed and nearly exhaustive workers' panel based on the original dataset described in detail by Babet, Godechot, and Palladino (2022). The DADS files for each year provide job variables at the individual level for the current and the previous year. This overlap allows for matching between yearly files at the worker level based on common information such as establishment ID, gender, number of hours worked, job duration, dates of employment, municipality of work and residence, earnings, and age. Using these matching procedures, Babet, Godechot, and Palladino (2022) achieved a high matching success rate of 98% for individuals between

2002 and 2019.

The importance of having access to exhaustive datasets should not be underestimated. When the sample size is reduced, as is the case in narrow panels, uncertainty in the estimation increases. Further, the identification of AKM models relies on mobile workers transitioning between firms, and this is only possible within the group of interconnected firms. Sampling methods significantly reduce the proportion of firms belonging to the main connected component, leading to a further selection bias favoring larger and more interconnected firms in the estimation sample. The challenge becomes even more pronounced when considering gender-specific AKM models, as the identification requires limiting data to the “dual-connected” set, i.e., the intersection of the female- and male-connected sets of firms.

Given that the procedure making the dataset almost universal for workers is not applicable before 2002, our analysis focuses on the period from 2002 to 2019. We exclude public sector employees, temporary agency workers, self-employed individuals, interns, and apprenticeships from our sample. Additionally, we restrict our analysis to workers aged between 20 and 65 years. For each year, we link each worker to the firm where they earned the highest income during the year (or, in cases of equal earnings, the firm they worked for the most). We drop workers with less than 90 days worked in their main job and firms with less than two employees in a given year.

Our main wage measure is the hourly wage, as we have information on hours worked. We compute it by dividing gross annual wage by annual hours worked. We exclude jobs with hourly wages below 80% of the legal minimum hourly wage for the corresponding year or exceeding 1000 times the minimum hourly wage.

### 3. Estimation framework

#### 3.1. A two-way fixed effect model

One important aspect of network data - like a matched employer-employee dataset - is the existence of two-sided unobserved heterogeneity. To account for this, in the context of wage regressions, the first and most widely used approach has been two-way fixed-effects log wage models pioneered by AKM, where wages of individual  $i$  in year  $t$  working for firm  $J$  are expressed as the sum of a worker fixed effect  $\alpha_i$  and a firm fixed effect  $\Psi_j$ :

$$(1) \quad \text{Log}(w_{it}) = \alpha_i + \Psi_{J(i,t)}^{G(i)} + \beta^{G(i)} X'_{it} + r_{it}$$

As we are interested in gender-specific effects, we estimate equation (1) separately for each gender  $G(i)$ . Controls  $X_{i,t}$  include a polynomial in age and 1-digit occupation<sup>1</sup>. The wage premium  $\Psi_j$  for any specific firm  $j$  can only be determined in comparison to a reference firm or group of firms. To compare these estimates across gender, we need to normalize them consistently. We describe our normalization procedure in section 3.2 below.

This econometric model is based on three key assumptions. First, it assumes that there is no interaction effect between firm type and worker type, meaning that the firm-specific wage premium will be the same for all workers regardless of their characteristics. Second, it assumes exogenous mobility, meaning that the time-varying residual component of wages  $r_{it}$  is uncorrelated with the choice of moving. Third, the model is

---

<sup>1</sup>The dataset has certain limitations regarding the measurement of experience and tenure. Years of entry are left censored in 1993 and inconsistent before 2002. Therefore, we only control for wage returns to seniority using age as a proxy for experience.

static and hence rules out the presence of dynamic factors in the realization of wages. Although these assumptions may seem implausible, they have withstood scrutiny from recent research. Bonhomme, Lamadon, and Manresa (2019) found only slight deviations from the additive linear model<sup>2</sup>. Di Addario et al. (2023) developed an extension of the model to include, in addition to the usual firm fixed effects, a set of “origin” effects, essentially firm effects for the firm where the worker previously worked. They find that this does not significantly increase the explanatory power of the model. Furthermore, when they separate their analysis by gender, they document that differences in “origin” effects make essentially no contribution to the gender wage gap, while differences in “destination” effects, the usual fixed effects for the firm where the worker works, do matter.

Our first estimation approach consists in running AKM regressions in three intervals of six years each – 2002-2007, 2008-2013, and 2014-2019. For each of the three periods, we build a dual-connected sample and estimate the model. As can be seen in Table 1 the main difference between the dual connected set and the full sample in each period is firm size: in the dual-connected set firms are on average bigger. Besides that, worker and firm characteristics are fairly similar.

### **3.2. Normalization**

In AKM models, firm effects are only identified relative to a constant. We thus cannot compare male and female effects without first normalizing them consistently. CCK’s original paper and subsequent literature using this method have de-measured all firm effects relative to average effects in either low-value-added firms or in the hotel and

---

<sup>2</sup>They do not have information on hours worked so they choose not to include female workers. Palladino, Stabile, and Roulet (2022) finds little evidence of complementarities in wages for both men and women, which suggests that an additive specification without interaction terms between worker and firm types sufficiently captures potential differences in wage setting between genders.



restaurant sector with the hypothesis that this method identifies firms with no surplus to share. These two normalization choices usually generate consistent results.

We follow the approach of Gerard et al. (2021) and normalize the firm effects only with respect to the restaurant sector<sup>3</sup>. This assumes that any gender wage gap in this industry is due either to differences in productivity (captured by the worker fixed effects) or to differences in the occupational composition (as we control for one-digit occupation in all regressions). Gerard et al. (2021) shows that the resulting firm pay premiums should be considered as lower bound estimates, while upper bound estimates would be obtained if we assumed that any gender wage gap in the restaurant sector is driven by differential pay setting behaviors for men and women.

### 3.3. Between firm versus within firm decomposition

Once we have estimated firm effects for men and women separately and normalized them, we can decompose them into a between component (*sorting* channel) and a within component (*bargaining* channel) as in CCK.

$$(2) \quad E[\Psi_j^M | Male] - E[\Psi_j^F | Female] = E[\Psi_j^M - \Psi_j^F | Male] + E[\Psi_j^F | Male] - E[\Psi_j^F | Female]$$

The left-hand side of equation (2) is the difference between the average firm fixed effect for men and the average firm fixed effect for women. We can then decompose it into two components. The first term on the right-hand side represents the bargaining, or within-firm, channel. It is the average difference between the firm fixed effects for men and women, assuming that women are represented in the different firms in the

---

<sup>3</sup>In hotels, economies of scale are more important. This effect is reflected in the average firm fixed effects for both men and women, which are significantly lower in restaurants than in hotels.

same proportion that men are, i.e. holding sorting fixed. The other two terms represent the sorting, or between-firm, channel. It is the difference between the average firm effect for women and what it would be if women were represented in the different firms in the same proportion that men are. If men and women were equally represented in each firm, this sorting term would be equal to zero<sup>4</sup>.

### 3.4. Rolling panels and clustering approach

To get more time-series granularity, we complement the more standard approach described in Section 3.1 with two alternative methods.

First, we implement the rolling AKM (R-AKM) method of Lachowska et al. (2023) by estimating AKM on six-year panels for every possible starting year in our data<sup>5</sup>. For each of the resulting thirteen periods, we build a dual-connected sample and estimate the model. We then apply decomposition (2) separately for each year of each panel estimation. Within each six-year panel, firm effects are fixed. Year-to-year variations in

---

<sup>4</sup>Note that we can rewrite the formula above holding the distribution of women across firms as fixed instead of that of men for the within-firm channel and using the male premium for the between-firm channel.

$$(3) \quad \begin{aligned} E[\Psi_k^M | Male] - E[\Psi_k^F | Female] &= E[\Psi_k^M - \Psi_k^F | Female] + \\ &E[\Psi_k^M | Male] - E[\Psi_k^M | Female] \end{aligned}$$

Like Bruns (2019), we prefer decomposition (2) as it reports sorting as what women would be paid if they worked at firms where men worked and the within component as what they would be paid at these firms if they were men (instead, under the alternative decomposition, sorting gives what men would be paid at firms where women work, which seems a less natural way to define our objects of interest.)

<sup>5</sup>Contrary to the original paper, we don't correct for the known bias currently referred to as the limited mobility bias (Andrews et al. 2008) since our evidence is entirely built on levels and not second-order moments. Indeed, the fixed effect estimates of the level of worker and firm effects are unbiased under the usual assumptions of AKM models (Bonhomme et al. 2023).

the firms' component of the gender wage gap can be due either to year-to-year changes in the populations of firms and workers and/or to changes in the composition of workers across existing firms. Across the different panels, within-year differences are instead driven by changes in the estimated fixed effects and/or differences in the composition of the dual-connected sets.

Second, we build sixteen 2-year subperiods following BLM data structuring in their static model. For every year  $t$ , starting in 2002, we focus on workers fully employed in years  $t$  and  $t+2$  and firms present in both years. A move is defined when the firm identifier for a worker changes between  $t$  and  $t + 2$ , hence implicitly happening in the course of  $t + 1$ . This conservative sample selection allows capturing as accurately as possible individual job moves between existing firms. For each of the resulting subperiods, we partition firms into clusters based on the similarity of their earnings distributions. Indeed, the AKM specification is ill-suited for short panels, especially in the case of analyzing genders separately. Cluster fixed effects then replace individual firm fixed effects. Specifically, the clusters are obtained by solving the following weighted k-means problem:

$$(4) \quad \min_{k(1), \dots, k(J), H_1, \dots, H_K} \sum_{j=1}^J n_j \int (\widehat{F}_j(y) - H_{k_j}(y))^2 d\mu(y)$$

where  $n_j$  is the average number of workers of firm  $j$ ,  $\mu$  corresponds to the discrete support of the cdf (in our case deciles),  $k(1), \dots, k(J)$  denotes a partition of firms into  $K$  known classes and  $H_1, \dots, H_K$  are generic cdfs. The minimization is for all possible partitions and class-specific cdfs. In other words, for each possible partition of firms into  $K$  groups, we can compute for each firm the distance between its log earnings cdf  $\widehat{F}$  and the centroid cdf  $H$  of the group to which the firm is assigned. Then for each partition, we can sum across all firms these distances (squared). The algorithm picks

the partition which minimizes this sum. We use the empirical distribution of earnings of stayers only, and we re-cluster firms every period. The baseline value for  $K$  is 1000, explaining more than 95% of the between-firm dispersion in earnings for men and women in each subperiod while preserving the dual connectivity of the entire sample (Table A2).

## 4. Results

### 4.1. Time series evidence on the contribution of firms to the gender wage gap

We begin by presenting descriptive evidence on the evolution of the gender wage gap in France over the last 20 years in Figure 1. Over this time period, the unconditional gender gap in daily wage (labeled the “Raw” gap in the left panel) fell from over 30% to less than 25% while the unconditional gender gap in hourly wage (the “Raw” gap in the right panel) went from 18% to 11.5%. Figure 1 then shows the gender gap in both daily and hourly wages controlling for occupation, occupation plus industry and occupation plus firm fixed effects. 1-digit occupations, 2-digit industries and firms all play a role in explaining the gender gap in daily pay and firms fixed effects explain more than just industry dummies. For the hourly wage, our main object of interest, industries and firms seem to be playing a very similar role, a point we will come back to later. We also see that occupations explained one third of the gender gap in hourly wage at the beginning of the period and now explain only a fifth, consistent with Goldin (2014).

Of course Figure 1 is only descriptive: controlling for firm effects, for instance, does not address the issue that different types of workers sort into different firms and that average differences in pay across firms may not be driven only by pure « firm effects ». We therefore implement the methodology outlined in Section 3 to better understand the role of firms. Still controlling for 1-digit occupation, we compute gender-specific

firm effects and find that the sorting of men and women across different types of firms has decreased significantly over the period of analysis. The results for three periods between 2002 and 2019 are reported in Table 2. Column 2 shows the overall gender gap in hourly wages. Column 3 reports the gender gap in firm pay premiums. Columns 4 and 7 break down this overall firm effect into a between and a within component. In 2002-2007 the between component accounted for 2 log points, while in 2014-2019 it was less than half that, i.e. 0.9 log points. Thus, of the 5 log points decrease in the hourly wage gap between the beginning and the end of our sample period (column 2), 1 log point can be attributed to a convergence in the sorting of men and women across firms (column 4).

Figure 2 presents the same kind of evidence but using rolling panels (Figure 2A) and 2-year panels with firm clusters effects instead of firm fixed effects (Figure 2B). In both cases the between-firm component declines, starting in around 2014 and continues through to the end of the period. Note that the Bruns (2019) paper which found an increase in the between-firm estimates stops its analysis in 2008 and compares two periods: 1995-2001, which is out of our sample, and 2002-2008. Thus our results are not necessarily inconsistent with these findings<sup>6</sup>.

These figures, and Table 2, also show that the within component has increased since the beginning of the period, in particular until around 2010. This explains why the overall gender gap in pay premiums did not decrease over our period of analysis, and even increased from 1.7 log points in our first period of analysis to 2.4 in the last one. We now turn to further investigate the mechanisms behind these developments in the between and within components.

---

<sup>6</sup>Note also that this paper differs from ours in its selection of workers - it restricts to full-time workers, we don't - and wage measurement - it uses daily wage where we focus on the hourly wage.

## 4.2. Mechanisms

### 4.2.1. Interpreting the decline in the between component

The decline in the between component of the gender gap in firm pay premiums can be driven both by increased convergence of men and women across firms within industries or by increased convergence across industries. We investigate this by further decomposing the between-firm component introduced in Equation 2 into a between-industry channel and a within-industry one:

$$(5) \quad E[\Psi_{j \in S}^F | Male] - E[\Psi_{j \in S}^F | Female] = \overline{E[\Psi_{j \in S}^F | Male]} - \overline{E[\Psi_{j \in S}^F | Female]} + \\ E[\Psi_{j \in S}^F - \overline{\Psi_{j \in S}^F} | Male] + E[\Psi_{j \in S}^F - \overline{\Psi_{j \in S}^F} | Female]$$

where  $\overline{\Psi_{j \in S}^F}$  is the average female firm fixed effect for all firms  $j$  belonging to two-digit industry  $S$ .

The results are shown in Table 2, columns 4, 5 and 6. We see that in all periods the between firm component (column 4) is mostly explained by a between industry channel (column 5): 1.6 log points out of 2 in the first period, 1.6 out of 1.8 in the middle period and 0.8 log point out of 0.9 in the last period. Most importantly for our purposes, we also see that most, i.e. 75%, of the decline in the between component over our period of analysis is explained by a decline in this between-industry channel. In other words, the convergence in the sorting of men and women between firms is primarily a convergence in the sorting of men and women between industries. The sorting across firms within industries is second order both in explaining the between-firm component of the gender gap in firm pay premiums at any point in time and in explaining the decline in this component over time. These results are robust to the use of rolling AKM, as shown in Figure A1, which also shows that the decline in the between-industry gap began in

the early 2010s and has continued since then. These findings are consistent with the fact that in Figure 1 above, when trying to explain the gender gap in hourly wage (right panel), controlling for firms rather than just industries does not add much.

#### **4.2.2. Interpreting the increase in the within-firm component**

We next turn to probing the robustness of the increase in the within-firm component of the gender gap in pay premiums between 2002 and 2019. One feature that could play a role in explaining changes over time in the within-firm estimates is the normalization procedure. As explained earlier, we normalize all firm effects relative to the restaurant sector in order to make meaningful comparisons between genders within each period. Further ensuring the comparability of within-firm estimates over time requires the additional assumption that the pay-setting behaviour for men and women within this industry remains consistent across the different estimation periods. Following Gerard et al. (2021), we consider our main estimates to be lower bounds, assuming that any gender pay gap in the restaurant industry is due to differences in worker productivity. To test the sensitivity of our results to this assumption, we compute upper bounds that capture the opposite assumption, i.e. that any gender pay gap in the restaurant industry is entirely due to differential wage-setting behaviour. To do this, we regress log wages on a dummy variable for men in the restaurant industry, controlling for 1-digit occupation and a polynomial in age (our usual controls). The resulting coefficient for men is then added to the lower bound estimate and we plot the range between the lower and upper bounds of the within-firm component. These bounds should plausibly encompass the true within-firm estimate. We do this using the rolling panels estimation and the results can be seen in Figure 3. Visually, it is inconclusive from this exercise whether the within-firm component has increased over time and suggests to us some caution.

A second possible driver for the observed increase in the within-firm component may be changes in the distribution of the within-firm gender gap in pay premiums over time. In the appendix, we further explore these distributions. We plot women's average firm effects relative to that of men, separately for our first and last periods (figure A2, holding sorting fixed, i.e. using male employment weighting for both men and women). The within-firm estimates for each period (reported in Table 2) are an average of all these points. The parts of the distribution where women's firm effects are furthest from the 45-degree line are parts where the within-firm gap in pay premiums is largest. We see that the slopes are not statistically different, which means that any change over time in the estimate of the within-firm component must be driven by differences in the constant term. The constant is precisely the term subject to the normalisation assumption explored above. This reinforces for us the fact that the estimated increase in the within-firm component should be interpreted with caution, in contrast to the decrease in the between-firm component, which is not sensitive to normalisation assumptions.

## **5. Conclusion**

We provide new evidence on the evolution of the gender gap in firm pay premiums using newly linked employer-employee data. We show that there is a significant decline in the between-firm component of the gender gap in pay premiums since 2000. Our results are remarkably consistent whether we use six-year panels in the AKM framework, rolling AKM panels, or the firm clustering approach with two-year panels. We then turn to understanding the mechanisms. 75% of the decrease observed in the between-firm component of the gender wage gap can be attributed to women increasingly sorting into industries that host higher-wage firms and only 25% to women gaining access to better firms within specific industries over time.



We also observe an increase in the within-firm component of the gender gap in pay premiums, which begins early in our period and flattens out by 2010. Overall, therefore, we observe a slight increase in the gender gap in firm pay premiums between the early 2000s and the late 2010s. However, we contend that the observed rise in the within-firm component of the gender wage gap is a result of the normalization assumption rather than an inherent economic mechanism.

Overall, the consistency of our results across methodologies provides some comfort in understanding roughly 20 percent of the decline in the overall gender wage gap over our period of analysis through increased convergence of the sorting of men and women across firms and points to the importance of cross-industry sorting as the primary mechanism.

## References

- Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms." *Econometrica* 67 (2): 251–333.
- Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (3): 673–697.
- Babet, Damien, Olivier Godechot, and Marco G. Palladino. 2022. "In the Land of AKM: Explaining the Dynamics of Wage Inequality in France." *Working Paper*.
- Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55 (3): 789–865.
- Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2023. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics* 41 (2): 291–322.
- Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A distributional framework for matched employer employee data." *Econometrica* 87 (3): 699–739.
- Bruns, Benjamin. 2019. "Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap." *American Economic Journal: Applied Economics* 11 (2): 74–113.
- Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *The Quarterly Journal of Economics* 131 (2): 633–686.
- Casarico, Alessandra, and Salvatore Lattanzio. forthcoming. "What Firms Do: Gender Inequality in Linked Employer-Employee Data." *Journal of Labor Economics*.
- Coudin, Elise, Sophie Maillard, and Maxime Tô. 2018. "Family, firms and the gender wage gap in France." *IFS Working Papers W18/01*.
- Di Addario, Sabrina, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. 2023. "It ain't where you're from, it's where you're at: Hiring origins, firm heterogeneity, and wages." *Journal of Econometrics* 233 (2): 340–374.

- Gallen, Yana, Rune V. Lesner, and Rune Vejlin. 2019. "The labor market gender gap in Denmark: Sorting out the past 30 years." *Labour Economics* 56 (C): 58–67.
- Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card. 2021. "Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil." *American Economic Review* 111 (10): 3418–57.
- Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4): 1091–1119.
- Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury. 2023. "Do firm effects drift? Evidence from Washington administrative data." *Journal of Econometrics* 233 (2): 375–395.
- Morchio, Iacopo, and Christian Moser. 2019. "The Gender Gap: Micro Sources and Macro Consequences." *Society for Economic Dynamics Meeting Papers* (143).
- Olivetti, Claudia, and Barbara Petrongolo. 2016. "The Evolution of Gender Gaps in Industrialized Countries." *Annual Review of Economics* 8 (1): 405–434.
- Palladino, Marco G., Mark Stabile, and Alexandra Roulet. 2022. "Understanding the role of firms in the gender wage gap over time, over the life cycle, and across worker types." *CEPR Discussion Papers* (16671).
- Sorkin, Isaac. 2017. "The Role of Firms in Gender Earnings Inequality: Evidence from the United States." *American Economic Review* 107 (5): 384–87.

# Main Tables and Figures

TABLE 1. Summary statistics

	2002-2007				2008-2013				2014-2019			
	Full sample		Dual Connected Set		Full sample		Dual Connected Set		Full sample		Dual Connected Set	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
Age	39.05	38.28	38.94	38.01	39.73	39.05	39.66	38.77	40.65	40.08	40.62	39.85
Gross Annual Wage (€)	28,128.80	20,079.05	29,041.40	20,571.97	32,213.34	23,629.97	33,247.54	24,234.34	34,858.37	26,729.68	36,170.48	27,517.63
Annual Hours Worked	1,582.64	1,398.01	1,581.87	1,406.80	1,591.85	1,405.28	1,592.51	1,414.02	1,569.68	1,405.83	1,575.75	1,416.11
Executives (%)	1.21	0.31	0.92	0.22	1.43	0.51	0.99	0.35	1.51	0.58	0.90	0.34
Managers, Professionals (%)	17.61	11.15	19.26	11.87	18.29	13.64	19.98	14.45	20.37	16.85	22.32	17.93
Intermediate Occupations (%)	21.08	23.31	22.09	23.91	18.43	18.37	19.52	18.84	17.36	18.40	18.48	18.87
Service and sales workers (%)	11.94	46.63	12.38	44.50	14.45	50.04	14.98	47.87	15.46	47.36	15.79	45.06
Manual Laborers (%)	48.16	18.60	45.36	19.50	47.40	17.43	44.53	18.50	45.30	16.81	42.51	17.79
Restaurant Industry (%)	3.25	4.14	3.13	4.07	3.81	4.87	3.74	4.89	4.42	5.14	4.24	5.07
Firm Size (# employees)	7,196.87	6,478.87	8,519.98	7,418.17	7,429.79	6,990.18	8,764.22	8,044.53	7,722.50	7,881.43	9,050.63	9,017.00
Person/Yr Obs	48,345,240	31,162,588	40,823,556	27,211,476	49,014,206	31,715,658	41,537,751	27,552,936	51,299,216	33,742,933	43,759,766	29,489,045
# of Individuals	15,054,754	10,604,903	12,779,691	9,239,207	12,644,707	9,095,722	10,762,804	7,924,807	15,086,288	10,481,791	12,788,320	9,138,844
# of Firms	817,731	763,717	335,722	335,727	842,439	787,676	356,264	356,264	1,005,776	914,297	393,970	393,970

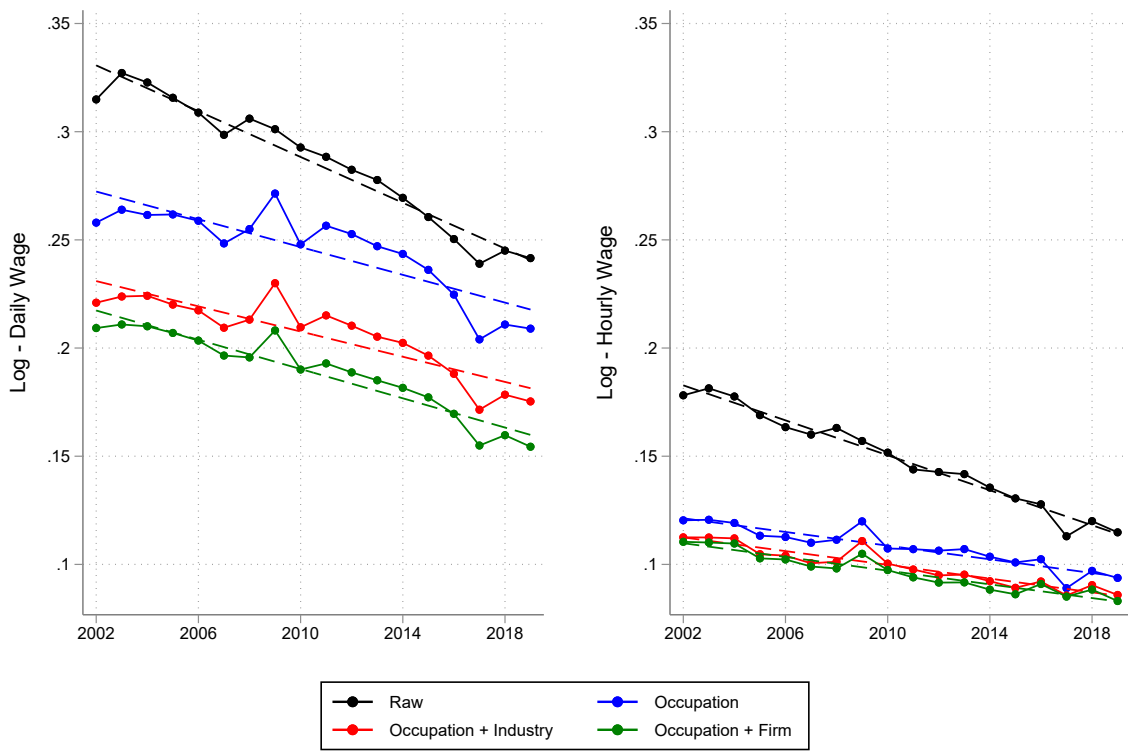
*Notes:* We exclude public sector employees, temporary agency workers, self-employed individuals, interns, and apprenticeships from our sample. Additionally, we restrict our analysis to workers aged between 20 and 65 years. We exclude jobs with hourly wages below 80% of the legal minimum hourly wage for the corresponding year or exceeding 1000 times the minimum hourly wage. We drop firms with less than 2 employees in a given year. The dual connected set refers to the intersection between the male and the female largest connected sets. All statistics are calculated across person-year observations.

TABLE 2. Firms' Contribution to the Gender Gap in Hourly Wages – Blocks

Period	(2)	(3)	(4)	(5)	(6)	(7)
	Unconditional gender gap	Gap in Firm pay premiums	Between-Firm component	Between-Firm Between-Sector	Between-Firm Within-Sector	Within-Firm component
2002-2007	0.1831	0.0172	0.0198	0.0159	0.0039	-0.0026
		9.42%	10.83%	8.70%	2.13%	-1.40%
2008-2013	0.1607	0.0298	0.0181	0.0157	0.0024	0.0117
		18.55%	11.24%	9.75%	1.49%	7.31%
2014-2019	0.1350	0.0238	0.0086	0.0076	0.0010	0.0152
		17.61%	6.35%	5.62%	0.73%	11.26%

*Notes:* For each sub-interval, we build a dual-connected sample, estimate a model as in Equation 1, normalize the firm effects with respect to the restaurant sector (see Section 3.2) and decompose the gender gap in pay premiums (Column 3) into a between-firm component (Column 4) and a within-firm component (Column 7) as described in Section 3.3. The second row in each cell corresponds to the respective quantity expressed as a percentage of the unconditional gender gap in hourly wage (Column 2). The between-firm component of the gender wage gap is further decomposed into a between-sector channel (Column 5) and a within-sector channel (Column 6) following Equation 5.

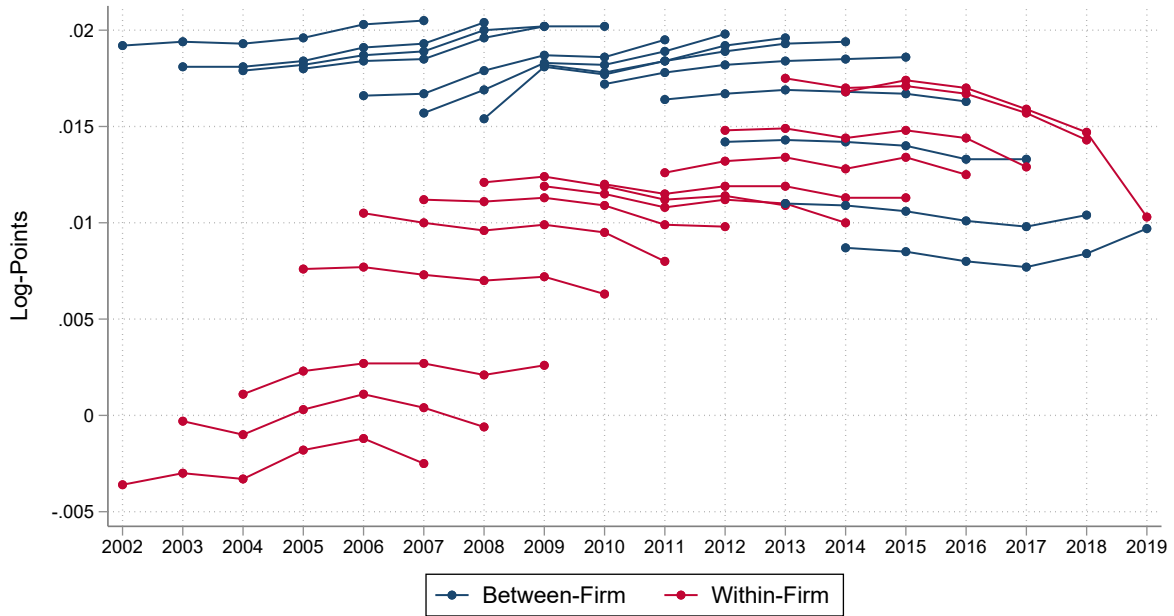
FIGURE 1. Gender Wage Gap in France over the period 2002-2019 - Daily vs Hourly



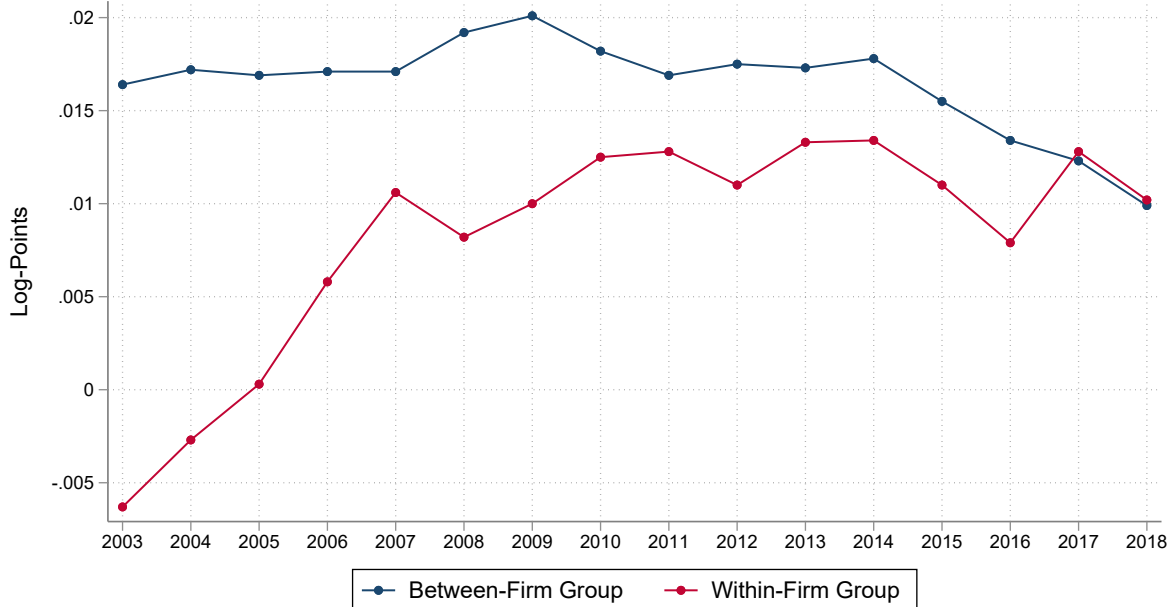
Notes: We plot the coefficients of a dummy for male workers from log wage regressions for each year in four different specifications. The “Raw” specification captures the gender wage gap once accounting only for age effects. The “Occupation” specification controls for 1-digit occupations (see Table 1 for the list), while “Occupation + Industry” further incorporates 2-digit sectors from Nace Rev. 2. Finally, “Occupation + Firm” controls for occupation and firm effects. In the left panel, we examine a daily measure of wage obtained by dividing the gross annual wage by the number of annual days worked. On the other hand, the right panel focuses on an hourly measure, computed as the gross annual wage divided by the annual hours worked. The break in 2017 is probably due to the fact that DADS, the main data source until then, was gradually replaced by a new administrative source, the DSN (“déclarations sociales nominatives”), starting in 2016.

FIGURE 2. Firms' Contribution to the Gender Wage Gap Over Time

A. Rolling Panels

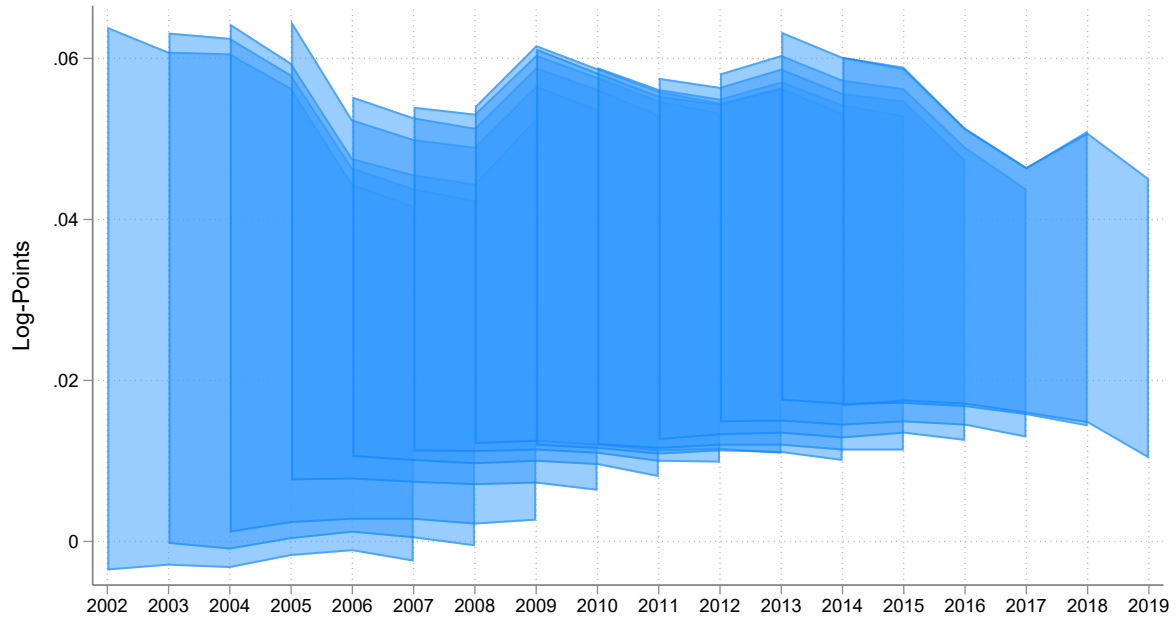


B. Firm Clustering



Notes: We plot the evolution of the between- and within-firm components of the gender gap in firm pay premiums over time using two different methodologies. Panel (a) displays the findings obtained through the rolling AKM (R-AKM) method, where we estimate the AKM model on six-year panels for every possible starting year in our data. For every thirteen periods, after estimating the model, we apply decomposition (2) separately for each year of each panel estimation. Panel (b) shows the results of the firm clustering exercise. We build sixteen 2-year subperiods. For each of the resulting subperiods, we partition firms into clusters, estimate the model, and apply decomposition (2).

FIGURE 3. Within-Firm component – Area Inbetween Lower and Upper Bounds – Rolling Panels

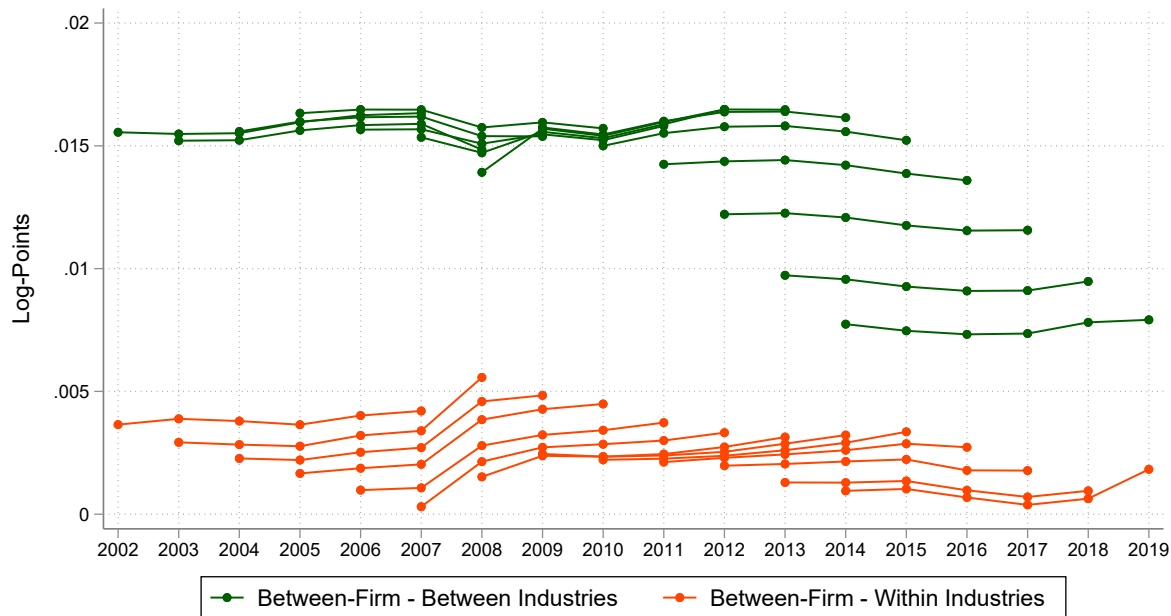


*Notes:* We plot the year-specific area between the lower and upper bound estimates of the within-firm component of the gender gap in firm pay premiums (see Section 4.2.2). The lower part of the shaded area represents our lower bound year-specific estimate in the rolling panels' specification (refer to Figure 2A). Conversely, the upper part of the shaded area corresponds to the upper bound.



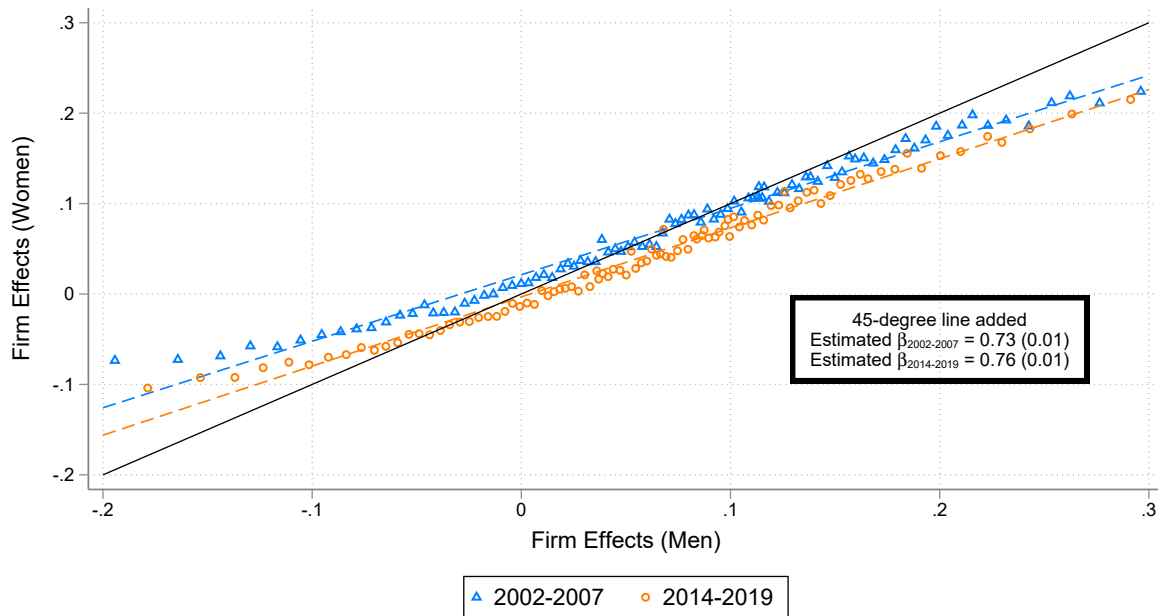
## Appendix Figures

FIGURE A1. Between- and Within-Industry Components of Sorting – Rolling Panels



Notes: We plot the evolution over time of the between- and within-industry components of the sorting estimate using the rolling AKM (R-AKM) method. This consists of estimating the AKM model on six-year panels for every possible starting year in our data. For every thirteen periods, after estimating the model, we apply decompositions (2) and (5) separately for each year of each panel estimation.

FIGURE A2. Firm Effects – Men vs Women – Blocks



Notes: We plot the relationship between firm effects for men and women in the 2002-2007 and 2014-2019 periods. Each dot on the graph represents a percentile of the distribution of firm effects for men during a specific period. The men's firm pay premiums are divided into 100 equally sized bins with an equal number of men person-years. The period-specific slopes are reported along with their corresponding standard errors.

## Appendix Tables

TABLE A1. Selected estimates of the contribution of firm pay policies to the gender wage gap

	(2)	(3)	(4)	(5)	(6)	(7)
	<b>Data</b>	<b>Wage</b>	<b>Gender</b>	<b>Gender gap in</b>	<b>Within</b>	<b>Between</b>
	<b>source</b>	<b>measure</b>	<b>wage gap</b>	<b>firm pay premiums</b>	<b>(Bargaining)</b>	<b>(Sorting)</b>
Card et al. (2016)	Portugal (2002-2009)	hourly	0.23	0.050	0.003	0.047
Bruns (2019)	West Germany (1994-2001)	daily	0.25	0.028	-0.014	0.042
	West Germany (2001-2008)	daily	0.25	0.064	0.001	0.063
Casarico & Lattanzio (forthcoming)	Italy (1995-2015)	weekly	0.21	0.065	0.021	0.044
Coudin et al. (2018)	France (1995-2015)	hourly	0.17	0.014	-0.004	0.018
Morchio & Moser (2020)	Brazil (2007-2014)	hourly	0.14	0.084	0.020	0.064

*Notes:* The table shows estimates of papers that report both the between and within components of the gender gap in firm pay premiums. Column 4 reports the gender gap in log wages of the estimation sample. Column 5 shows the difference between average firm effects for males and average firm effects for females while columns 6 and 7 show the split between a within and a between components (using female effects for the between term and the male distribution for the within one, see equation 2).

TABLE A2. Between-Firm vs Between-Cluster variance of log-earnings

	Men			Women		
	Between-Firm	Between-Cluster	Ratio	Between-Firm	Between-Cluster	Ratio
2003	0.0843	0.0819	97.12%	0.0997	0.0959	96.22%
2004	0.0852	0.0826	96.96%	0.0995	0.0955	95.95%
2005	0.0861	0.0833	96.82%	0.0993	0.0957	96.38%
2006	0.0869	0.0845	97.26%	0.0997	0.0963	96.56%
2007	0.0889	0.0866	97.49%	0.1010	0.0976	96.65%
2008	0.0888	0.0861	96.97%	0.1065	0.1029	96.66%
2009	0.0919	0.0894	97.28%	0.1094	0.1059	96.79%
2010	0.0895	0.0874	97.60%	0.1074	0.1035	96.37%
2011	0.0909	0.0888	97.71%	0.1086	0.1052	96.92%
2012	0.0903	0.0881	97.63%	0.1084	0.1057	97.52%
2013	0.0920	0.0902	98.01%	0.1093	0.1063	97.29%
2014	0.0929	0.0910	97.98%	0.1094	0.1066	97.45%
2015	0.0896	0.0869	96.92%	0.1055	0.1023	96.89%
2016	0.0890	0.0862	96.82%	0.1025	0.0989	96.52%
2017	0.1008	0.0980	97.26%	0.1151	0.1109	96.30%
2018	0.1057	0.1032	97.70%	0.1221	0.1184	96.95%

*Notes:* The table shows the evolution of the between-firm and between-cluster variances of log earnings over time for men and women separately. The column “Ratio” reports the ratio between the two. Each year corresponds to a 2-year subperiod whose construction is described in Section 3.4.