Worker Sorting and the Gender Wage Gap

Giulia Vattuone*

June 16, 2023

Abstract

Around 15% of the gender wage gap is due to women sorting into firms that pay lower wages. Using French matched employer-employee data, I investigate whether these gender differences in sorting reflect differences in preferences or opportunities. I employ a finite mixture approach *à la* Lentz, Piyapromdee, and Robin (2023) to estimate a model of wages and mobility. Using information on wages, mobility, and observed characteristics, this model classifies workers and firms into a finite number of types and classes. Moves within and between firm classes separately identify the two key channels under the assumption that workers of the same gender and type are indifferent between firms that belong to the same class. I find that gender differences in preferences account for up to 70% of the sorting component of the gender wage gap, but with considerable differences across age groups.

JEL Codes: J16; J31; J63; J64

Keywords: Gender Wage Gap; Sorting; Job Preferences; Finite Mixture Models

* University of Warwick. Email: G.Vattuone@warwick.ac.uk. I am grateful to Roland Rathelot and Manuel Bagues for their invaluable guidance and infinite patience. For their incredible help and support, special thanks to Christine Braun, Mirko Draca, Emma Duchini, Roberto Pancrazi, and Natalia Zinovyeva. For insightful conversations and feedbacks, I thank Christian Belzil, Luis Candelaria, Julien Combe, James Fenske, Eric Renault, Seth Sanders, Benoit Schmutz, Thijs van Rens and participants in seminars at Warwick University, CREST-ENSAE, and at the 23rd IZA Summer School in Labor Economics. I thank CASD for data access and support. All errors are my own.

1 Introduction

The gender wage gap partly reflects differences in sorting across firms. Following the seminal work of Card et al. (2016), several studies confirm this finding.[1] A major debate is whether gender differences in sorting stem from differences in preferences or opportunities. In this paper, I address this question by estimating a model of wages and mobility by exploiting information on firm-to-firm transitions.

I employ a revealed preference argument in a random search framework, initially proposed by Sorkin (2018). Following this approach, data on observed firm-to-firm transitions are informative about offer arrival rates and worker preferences. The intuition is that, upon receiving an offer, a worker chooses to accept if the perceived value of the poacher is higher than the one of the incumbent. Workers may value something beyond wages in a way that guides where they sort. Throughout the paper, offer arrival rates represent employment opportunities to move to a specific firm, while perceived firm values represent worker preferences.

Sorkin (2017) studies revealed preferences through firm-to-firm mobility to estimate gender-specific firm-level values for workers, and compares these values to gender-specific firm-level earnings to study the role of compensating differentials in explaining wage inequality between men and women. The novelty of this paper is to allow for worker heterogeneity within as well as between genders in a framework that generates rich sorting patterns. I employ a finite mixture model recently proposed by Lentz et al. (2023) and Bonhomme et al. (2019), and rely on matched employer-employee monthly data for the region *Ile-de-France* (greater Paris) over the period 2016-2019.

Administrative data directly provide worker and firm matches, making it challenging to

1. It has been widely documented that unequal gender distributions across workplaces contribute to the gender wage gap (Blau 1977, Hirschman 2022). Card et al. (2016) is the first paper to comprehensively analyse the role of gender differences in worker-firm allocations in explaining the gender wage gap. Their approach, which builds on the log earnings model of Abowd et al. (1999), has been adopted using data from multiple countries: the US (Sorkin 2017), France (Coudin et al. 2018, Palladino et al. 2021), Germany (Bruns 2019), Italy (Casarico and Lattanzio 2022), Canada (Li et al. 2020), Brazil (Morchio and Moser 2020), Chile (Cruz and Rau 2022). The share of the gender wage gap due to differences in firm sorting ranges roughly between 15% and 25%, depending on country-specific data availability and labour market institutions. Differences in firm sorting are not related to a lack of skills. In general, Blau and Kahn (2017) stress that a substantial portion of the gender wage gap cannot be explained by conventional supply-side factors like human capital accumulation, psychological attributes or non-cognitive skills. See also Olivetti and Petrongolo (2016) for an extensive literature review of gender gaps.

disentangle choices from opportunities to move. The identification of the two key mobility channels requires additional assumptions. First, workers and firms are associated with a finite number of *types* and *classes*, respectively. Second, workers of a given type have common preferences over firms of a given class, up to an idiosyncratic utility draw specific to the worker-firm match. When choosing between two firms, workers consider the firm's common value, which is worker-type and firm-class specific, as well as the idiosyncratic utility draw. An interpretation of the idiosyncratic draw is that the choice to move may be influenced by moving costs.

To see how these two assumptions disentangle offer arrival rates from preferences, consider a simplified example. Suppose there is one type of worker and two classes of firms, A and B. Workers, in expectations, are indifferent between firms belonging to the same class. With no loss of generality, we can assume that when workers employed in a firm of class A draw an offer from another firm that also belongs to class A, half of them accept. The expected number of offers from class-A firms is, then, twice as much as the number of transitions that occur *within* class A. We identify the expected number of offers from class B with similar reasoning. Once we pin down the expected number of offers, we can look at the number of *between-class* moves to recover the expected share of workers choosing A over B, and vice versa. Choice probabilities reveal preferences under the following argument: conditional on receiving an offer, if a higher share of workers accepts offers from class A than offers from class B, then we can infer that workers prefer firms in class A. We can extend this simplified example to cases with multiple worker types and firm classes.

Under the fore-mentioned identifying assumptions, firm-to-firm transition probabilities are modelled as the product of an offer arrival rate and a choice probability. Transitions into non-employment and out of non-employment are left unrestricted. Finally, the framework allows for worker-firm wage complementarities, assuming that workers draw hourly wages from a distribution specific to worker types and firm classes. Similar to Abowd et al. (1999), mobility depends only on worker types and firm classes but not directly on wages.

I estimate the model in two steps as in Bonhomme et al. (2019). First, I group firms into classes employing a k-means algorithm. This algorithm uses firm data on size, gender-specific wage distributions and inflow/outflow rates. Second, conditional on the firm classes, I group workers into types and estimate the parameters of interest using an Expectation-Maximisation algorithm. This algorithm uses data on observed workers' wages and mobility patterns, allowing for flexible interactions between latent types and observed gender, tenure, and age groups. I follow the iterative process developed by Lentz et al. (2023) to deal with non-linearities in the mobility parameters.

Using the estimated mobility parameters, I obtain the stationary worker-firm allocations. Then, I perform several counterfactual exercises where I equate the mobility parameters of women to the ones of men in order to simulate scenarios where men and women share similar patterns in *i*) preferences over firms, *ii*) offer arrival rates, *iii*) entry rates into non-employment, or *iv*) re-entry rates into employment. I study how the gender wage gap changes under these multiple scenarios.

First, I find that the unequal employment distribution across firms accounts for 14 percent of the residualized gender wage gap. Gender differences in worker preferences, inferred from firm-to-firm transitions, are the primary determinants of this sorting component of the wage gap. Preferences account for up to 70% of this component. This result implies that amenities may play an important role in female sorting patterns and that firms differ in the amenity provision valued by male and female workers.

Second, gender differences in offer arrival rates do not contribute to the gender wage gap. On the contrary, if women sampled job offers at the same frequency as men, the gender wage gap would slightly increase, implying that women are more likely to draw job offers that would pay more. Although surprising, this result may be supported by the findings of a recent correspondence study run by the French *Institut des Politiques Publiques*.[2] The study carried out a large-scale experiment by sending fictitious CVs in response to several thousands job offers in eleven distinct professions. Callback rates in low-skilled occupations are significantly lower for women, while the opposite is observed for executive occupations with supervision, roles populated mainly by men. To the extent that the firm clustering captures differences in occupational compositions, my results align with these findings.

Third, differences in transitions into non-employment do not contribute to the gender

2. (Note IPP n°67) Discrimination à l'embauche selon le sexe: les enseignements d'un testing de grande ampleur. https://www.ipp.eu/actualites/note-ipp-n67-discrimination-a-l-embauche-selon-le-sexe-les-enseignements-d-un-testing-de-grande-ampleur/

wage gap.

Finally, I find that gender differences in patterns of re-employment transitions are the second most important determinant of the sorting component of the gender wage gap.

Do these sorting effects vary by age? Among early-career workers aged 25-35, gender differences in firm sorting explain 25 percent of the residualized wage gap. Up to 60% of this effect is accounted for by gender differences in sorting after a non-employment period. Although I do not observe maternity in my data, women in this age group are likely to have their first child. These results may thus reflect how the so-called child penalty (Kleven et al. 2019) affects differences in mobility that translate into gender wage differentials. Recent evidence shows that mothers opt for unemployment insurance benefits and forgo less generous standard parental leave programs (Zurla 2022). Based on my estimates, patterns in re-entry rates drive a sorting effect. As this type of transition is left unrestricted in the model, I cannot disentangle whether this result reflects differences in the offer distribution non-employed female and male workers face or differences in preferences for non-wage amenities.

Among 36-45 years old workers, the sorting component accounts for 20 percent of the residualized wage gap. In addition, perceived firm values become the principal determinant. Sorting based on preferences accounts for around one-fifth of the gender wage gap among workers of this age group. The analysis suggests that amenities matter mainly during a more mature stage of the career, when women may seek for better work-life balance to deal with family or caregiving responsibilities. Among workers aged 46-55, differences in where men and women work do not contribute to the gender wage gap, reflecting that wage profiles across firms become flat during a more senior age.

This paper relates to several strands of literature. First, long-standing literature has been studying gender differences in labour mobility in determining wage differentials (Loprest 1992, Bowlus 1997, Del Bono and Vuri 2011). Compared to this literature, I leverage detailed matched employer-employee data.

Second, I complement the literature that quantifies the sorting component of the gender wage gap. This literature starts with Card et al. (2016), and it builds on the pioneering work of Abowd et al. (1999), who estimate by Ordinary Least Squares a linear wage equation with additive worker and firm fixed effects and condition on observed worker characteristics.

Adopting the finite mixture model of Lentz et al. (2023) and Bonhomme et al. (2019) permits the explicit modelling of mobility, allowing me to gauge the relative importance of key mobility channels driving between-firm gender wage differentials. Casarico and Lattanzio (2022) show that women in firm-to-firm transitions are less likely to move towards firms with higher fixed effects estimates, that is, with higher wage policies. My paper complements their result by separating the role of offer arrival rates and worker preferences.

Most importantly, the revealed preference approach for estimating worker-perceived firm values connects my paper to Sorkin (2017) and Morchio and Moser (2020). Sorkin (2017) and Morchio and Moser (2020) conclude that there is an overall agreement between men and women on how they value firms. They also attribute differences in between-firm pay gaps to differences in the job offer distribution (Sorkin 2017) and to firms directing vacancies towards certain workers (Morchio and Moser 2020). The contribution of my paper is to allow for worker heterogeneity within as well as between genders. Based on my estimates, gender differences in preferences for non-wage amenities seem important in explaining differences in sorting across firms. The different conclusions of this paper may be due to French labour market specificities. At the same time, within-gender variation may swamp average gender differences in some mobility factors in a way that may underestimate the relevance of gender differences in sorting across different segments of the market.[3]

Finally, I relate to the important literature that points out that gender wage differentials may materialise as a result of differences in job search behaviour (Cortés et al. 2023, Braun and Figueiredo 2022), employer discrimination in hiring (Neumark et al. 1996, Xiao 2021, Kline et al. 2022), or as women have stronger preferences for shorter commuting time (Le Barbanchon et al. 2020, Petrongolo and Ronchi 2020, Fluchtmann et al. 2021), or for flexibility (Wiswall and Zafar 2018).[4] In this paper, I attempt to separate the relative importance of gender differences in offer distributions, which subsume worker and firm search

3. For example, Bertrand (2020) stresses the importance of within-gender variation in personal traits such as confidence, risk aversion, and willingness to negotiate. She reviews several meta-analyses that conclude that average gender gaps in these personal traits are minimal.

4. Concerning flexibility, evidence is mixed. Among low-skilled workers, Mas and Pallais (2017) do not find that differences in the value for flexibility translate into gender wage gaps. In a recent randomised experiment carried out in a large firm, Angelici and Profeta (2020) find that flexible time and space work improves the well-being and work-life balance of both male and female workers.

behaviour, and in preferences. I infer worker preferences from firm-to-firm transitions and do not focus on a specific preference mechanism. My estimates of worker-perceived firm values capture an overall bundle of firm characteristics valued by workers.[5] Throughout the paper, I do not take a stand on whether gender differences in perceived firm values arise from 'true' preferences or whether they reflect gender stereotypes or norms that influence the choices men and women make.

The remainder of the article is organised as follows. Section 2 presents the framework of analysis. Section 3 describes the estimation procedure. Section 4 and Section 5 present the data and results from the classification algorithms. Section 6 illustrates the empirical analysis. Finally, section 7 concludes and discusses some caveats.

2 Theoretical Framework

This section presents a theoretical framework with which to interpret the observed data. The objectives are twofold.

First, I want to predict worker mobility across firms, and into and out of non-employment. I model firm-to-firm mobility as a function of opportunities to move and preferences. Sorting is intended as the stationary worker-firm allocation, and it is obtained using the estimates of the mobility parameters.

Second, I am interested in predicting log-wage distributions of workers across firms. Estimates of worker-firm allocations and log-wage distributions allow me to document the relative importance of key mobility components driving gender imbalances in employment across firms that translate into gender wage differentials.

I employ a finite mixture model *à la* Lentz et al. (2023). In what follows, I describe in detail the framework of analysis and discuss the assumptions.

5. It is also important to stress that, in the absence of an experiment, estimating the willingness to pay for specific job attributes in an imperfectly competitive market has been proven difficult. Search frictions may entail small equilibrium wage differentials across jobs even in the presence of substantial preferences for amenities (Bonhomme and Jolivet 2009). As I estimate the average firm values perceived by workers of a given group, I leave unrestricted the way wages and amenities shape worker-perceived firm values. In the model I focus on, wages and firm values are separate parameters, and I can infer ex-post the importance of non-wage components by inspecting the stationary worker-firm allocations.

Agents

There are N workers and J firms. Workers are indexed by $i \in \{1, \ldots, N\}$ and firms by $j \in \{0, 1, \ldots, J\}$, where $j = 0$ is non-employment. Both firms and workers are heterogeneous.

Firms are associated with a finite number of K *classes*. The firm in which worker i is employed at time t is $j(i,t)$, and I denote as $k_{j(i,t)} \in \{1, \ldots, K\}$ the class of firm $j(i,t)$. The class of non-employment is $k_0 = 0$. I describe how I estimate firm classes in Section 3.1.

Workers differ in their initial observable characteristics (gender, age, tenure) and in their unobserved characteristics. Workers' unobserved heterogeneity is discrete and can be clustered into L groups. The *type* of worker i is thus a triplet $(l_i,\, g_i,\, x_{i1})$, where $l_i \in \{1, \ldots, L\}$ is the latent heterogeneity, $g_i \in \{F, M\}$ is an indicator for whether the worker is female (F) or male (M), and x_{i1} are combinations of age and tenure observed in the initial time period. Each t refers to a calendar month. I describe how I estimate worker types in Section 3.2.

Timing

In period 1, a worker with observed characteristics g_i and x_{i1} enters the panel being employed. Initial observed heterogeneity determines a particular distribution of initial matches $\Pr\Big(l, k_{j(i,1)} \;\Big|\; g_i, x_{i1}\Big)$, which is left unrestricted.

Job mobility between a firm at time t and another firm at time $t+1$ is denoted by $s_{it} = 1$. In every period $t \geq 1$, the worker changes employment status or firm class ($s_{it} = 1$ or 0) with a probability that depends on worker's type (l_i, g_i, x_{it}) and current firm class $k_{j(i,t)}$. I denote this probability as $\Pr\Big(k_{j(i,t+1)} \;\Big|\; k_{j(i,t)}, l_i, g_i, x_{it}\Big)$. Transitions into and from nonemployment are left unrestricted, while I model job-to-job transitions as the product between a job sampling probability and a choice probability as in Lentz et al. (2023). Whether a transition occurs in the last period is unknown.

The worker draws log-wages from a static distribution that depends on worker's types and firm's classes. The distribution of log-wages is $f\Big(y_{it} \;\Big|\; l_i, g_i, x_{it}, k_{j(i,t)}\Big)$, and it is assumed to be normal with (l, g, x, k)-specific means and variances.

I formally specify all parameters, along with their identification, in Section 2.1.

Discussion of the assumptions

The paper aims to assess to what extent the gender wage gap is explained by men and women being sorted differently across firms, to identify the key mobility components driving gender imbalances in employment across firms, and to quantify their relative importance in determining gender wage differentials. In practice, this translates into predicting worker-specific average wages across firms and their job mobility in the labour market. The high number of workers and firms makes the estimation of the parameters of interest burdensome. In addition, and most importantly, separating offer arrival rates from choice probabilities in matched employer-employee data for any worker-firm combination is not possible. The latent-type framework helps overcome these challenges.

Workers and firms are associated with latent *types/classes* that have an effect on earnings and mobility. Worker latent types interact with worker observed characteristics, allowing for a flexible relationship between their observed and unobserved heterogeneity. The interpretation is that, in expectations, workers of type (l, g, x) earn similar wages and have similar mobility patterns and that the heterogeneity of firms is captured at the level of the class to which firms belong.

Adopting a latent-type framework drastically reduces the number of parameters to be estimated, thus overcoming over-fitting issues that may be encountered in the fixed-effect estimation proposed by Abowd et al. (1999). The latent-type framework also improves on the fixed-effect estimation biases arising from the limited mobility of workers across firms (Bonhomme et al. (2023)). Importantly, it allows us to model mobility explicitly.

Workers in the first period earn log hourly wages drawn from a normal distribution specific to worker types and firm classes. When a new job spell starts, log hourly wages are drawn from the same distribution. Similar to Abowd et al. (1999), the wage distribution does not allow for wage dynamics. Similar to Bonhomme et al. (2019), the wage distribution does not impose separability between worker and firm heterogeneity. Similar to Lentz et al. (2023), stayers and movers share the same wage means.[6]

6. In Lentz et al. (2023), stayers draw wages from a dynamic distribution while movers draw wages from a static distribution. The two distributions share the same mean wages but have different variances. In Abowd et al. (1999), firm fixed effects are estimated only on movers. Bonhomme et al. (2019) estimate wage distribution parameters (worker- and firm-specific averages and variances) only on movers.

Mobility is a Markov process independent of wage realisations conditional on worker types and firm classes. This is the standard exogenous mobility assumption (Abowd et al. (1999)). Exogenous mobility implies that job assignment and job-to-job mobility depend only on observed and unobserved characteristics of workers and firms. Although it rules out mobility motivated by learning of new job opportunities or, more generally, driven by idiosyncratic shocks to earnings while on the job, it still allows for different sorting patterns. In particular, I can investigate sorting patterns based on worker-firm complementarities in wages separately from sorting patterns based on preferences for non-wage components.

In the model, workers make a firm-to-firm transition if they receive a job offer and if the value of the poacher is perceived as superior to the one of the incumbent. The model thus assumes that firm-to-firm mobility reveals preferences, allowing for differences in the opportunity to move (represented by the job offer rate). Workers of type (l, g, x), up to an i.i.d. idiosyncratic utility draw, value firms of class $k = 1, \ldots, K$ the same. When choosing between the poacher and the incumbent, workers take into account the common value of the firm as well as the idiosyncratic draw. The idiosyncratic utility draw is specific to the worker-firm match and may capture, for example, a mobility cost. The idiosyncratic draw is distributed type I extreme value with scale parameter 1. Under this distributional assumption, upon receiving an offer from a firm in class k', workers move to the poacher with a probability that increases in the ratio between the common value of the poacher and the one of the incumbent. Perceived firm values and wage distributions are separate parameters. There is no restriction on how firm values and average wages are related. This allows for the possibility that workers may value something beyond just wages in the firm.

If men and women care only about wages and earn higher wages at different firms, gender-based differences in worker-firm allocations may arise from a comparative advantage explanation. If women instead care about amenities more than wages, then they may be more likely to sort into firms that offer higher levels of amenities, which may not necessarily be the ones that would pay them more.

The job ladder based on utility levels closely mirrors the one proposed by Sorkin (2018), who analyses firm-to-firm transitions to estimate utility levels of working at a firm and compares it to firm-level earnings to find the role played by compensating differentials in explaining wage inequality. Sorkin (2017) and Morchio and Moser (2020) adopt Sorkin (2018)'s revealed preference estimation technique to study the gender wage gap in US and Brazil, respectively. The adoption of the finite mixture approach permits to have heterogeneous workers both within gender and across gender and to be more in line with key features of theoretical sorting models (see, for example, Bagger and Lentz (2019)).[7]

To sum up, from a theoretical labour perspective, the latent-type model relates to partial equilibrium on-the-job search models with heterogeneous workers and firms, wage posting, random preferences for job types, and worker-specific offer arrival rates. From an empirical labour perspective, if I were to impose additivity between worker and firm heterogeneity in the wage equation, the model reduces to a latent-type version of Abowd et al. (1999).

The following subsection presents the likelihood function and describes the specification of the parameters of interest, which formally outline all the model assumptions.

2.1 Likelihood Function

The observed data for worker i consist of sequences of firm identifiers $\Big(j(i,1),\ldots,j(i,T)\Big)$, log-hourly wages $\Big(y_{i1},\ldots,y_{iT}\Big)$, mobility indicators $\Big(s_{i,1},\ldots,s_{i,T-1}\Big)$, and observed characteristics g_i and x_{it}. The latent data consist of the unobserved heterogeneity types $l_i \in \{1,\ldots,L\}$ and $k_{j(i,t)} \in \{0,1,\ldots,K\}$, for $i \in \{1,\ldots,N\}$, $j \in \{0,1,\ldots,J\}$, and $t \in \{1,\ldots,T\}$.

Conditional on a classification C of firms into classes, on the initial characteristics x_{i1}, on gender g_i, and on a value θ of the parameters, the complete likelihood of worker i's history is:

$$\begin{aligned}\mathcal{L}_i(\theta|l_i,g_i,x_{i1},C) = \Pr\Big(l_i,k_{j(i,1)}\;\Big|\;g_i,x_{i1}\Big)\times\prod_{t=1}^{T-1}\Bigg\{&\Pr\Big(k_{j(i,t+1)}\;\Big|\;k_{j(i,t)},l_i,g_i,x_{it}\Big)^{\mathbb{1}\{s_{it}=1\}}\\ &\times\Pr\Big(\neg\;\Big|\;k_{j(i,t)},l_i,g_i,x_{it}\Big)^{\mathbb{1}\{s_{it}=0\}}\Bigg\}\\ &\times\prod_{t=1}^{T} f\Big(y_{it}\;\Big|\;l_i,g_i,x_{it},k_{j(i,t)}\Big)\end{aligned} \tag{1}$$

7. Bagger and Lentz (2019) view job-to-job moves as a revelation of preferences in a framework that allows for worker heterogeneity in skill levels. Taber and Vejlin (2020) also use a revealed preference argument, highlighting the importance of preferences for non-wage components in determining worker choices between two jobs.

The likelihood function factors into three parts: contributions from the initial matching distribution, contributions from the mobility processes, and contributions from hourly wages.

Initial matching distribution

At $t = 1$, worker i enters the panel being employed. Observed characteristics, (g_i, x_{i1}), determine the initial probability of worker-firm match $\Pr\Big(l_i, k_{j(i,1)} \,\Big|\, g_i, x_{i1}\Big)$. The worker's observed characteristics consist of her gender, $g_i \in \{F, M\}$, and combinations of short/long tenure status and age groups, x_{i1}. Following Lentz et al. (2023), short tenure is defined to be less than two years. I divide age into three groups: young (25-35), mid (36-45), senior (46-55). I therefore have $L \times 2 \times 6$ worker types. The initial matching parameter is left completely unrestricted and estimated using simple frequencies. For notational simplicity, from now onwards, I denote the initial matching distribution as $m_0\Big(l, k \,\Big|\, g, x\Big)$.

Within a latent class, each firm is equally likely to be selected. I do not explicit the factor that represents firm-specific sampling in the likelihood as, conditional on a firm classification into classes, it gets simplified out in the expectation step of the EM algorithm used to estimate the posterior probability that worker i is of type l, and it is a simple parameter that enters additively the log-likelihood in the maximisation step of the EM algorithm. In other words, the uniform-sampling assumption is not problematic. As I proceed in two steps, first clustering firms and then clustering workers conditional on the firm classification, in principle any assumption about the firm-specific sampling can be made. I further clarify this last point in section 3.2.

Mobility processes

At each period $t \in \{1, \ldots, T-1\}$, I observe whether the worker separates from the current firm, $s_{it} = 1$, or stays, $s_{it} = 0$. Mobility at $t = T$ is unknown. The worker changes employment status and firm class with a probability that depends on worker's type and current firm class, $\Pr\Big(k_{j(i,t+1)} \,\Big|\, k_{j(i,t)}, l_i, g_i, x_{it}\Big)$. For notational simplicity, denote the current firm class by k and subsequent firm class by k'. In addition, denote the transition probability by $m\Big(k' \,\Big|\, k, l, g, x\Big)$. The worker stays with probability $m\Big(\neg \,\Big|\, k, l, g, x\Big) = 1 - \sum_{k'=0}^{K} m\Big(k' \,\Big|\, k, l, g, x\Big)$.

Job-to-Job transitions. At any time period t, a (l, g, x)-type worker moves from a firm of class $k = 1, \dots, K$ to a firm of class $k' = 1, \dots, K$ if the worker receives an offer and if prefers the poacher over the incumbent. The poacher is preferred if the perceived value of the match (l, g, x, k') is higher than the perceived value of the match (l, g, x, k). The probability of a job-to-job transition is thus specified as the product between a job sampling probability and a choice probability:

$$m\Big(k' \;\Big|\; k, l, g, x\Big) = \lambda_{lgxk'} \, P_{lgx}\Big(k' \succ k\Big) = \lambda_{lgxk'} \, \frac{\gamma_{lgxk'}}{\gamma_{lgxk} + \gamma_{lgxk'}}$$

where $\lambda_{lgxk'}$ represents the probability that a worker of type (l, g, x) receives an offer by a different firm of class k'.[8] Upon receiving an offer, the worker evaluates both the current firm of class k and the potential poacher of class k'. The worker takes into account the values of the firms, common to worker types and firm classes, as well as an idiosyncratic utility draw. The worker moves if the firm of class k' is preferred over the firm of class k. Assuming the idiosyncratic draw is distributed according to a type I extreme value, the choice probability $P_{lgx}\Big(k' \succ k\Big)$ is increasing in the ratio of the two common values $\gamma_{lgxk'}/\gamma_{lgxk}$. The choice probability is therefore an increasing function of the ratio of the perceived common value of the poacher over the perceived common value of the incumbent. To be precise, γ_{lgxk} $\forall\, k \in \{1, \dots, K\}$ is a monotonic transformation of the firm values.

For the estimation, Lentz et al. (2023) see the choice probability, $P_{lgx}\Big(k' \succ k\Big)$, as a Bradley-Terry specification (Bradley and Terry (1952), Hunter (2004)).

The Bradley-Terry specification was initially introduced to model a situation in which individuals in a group are repeatedly compared with one another in pairs. As matched employer-employee data can be represented in a directed graph where the nodes are firms and the edges are non-negative integers of worker transitions between any pair of firms, the Bradley-Terry specification turns useful to estimate how workers value firms, under the assumption that they make only binary choices. Using information on relative flows between firms, it is possible to obtain a firm ranking that orders firms based on their value. The ranking is obtained for those firms such that there is a path from j to j', for all nodes j and j'. Under the latent-type framework the graph connectivity condition is likely to hold, and it

8. The different firm may belong to the same class of the firm in the current period.

does not require to focus on the set of strongly connected firms. Indeed, I end up having (l, g, x)-worker-type-specific $K \times K$ matrices where rows represent arrival firm classes and columns represent departing firm classes. Each cell contains information on the total number of transitions of (l, g, x)-type workers between firm classes. Two perceived value vectors γ_{lgx} and γ'_{lgx} are equivalent if one is a scalar multiple of the other. The firm values are thus normalised so that $\sum_{k=1}^{K} \gamma_{lgxk} = 1$.

It is worth highlighting that the estimates of γ_{lgxk} do not simply represent a ranking of preferences for firm classes. What matters is how much more a firm class is preferred over another. Ratios of firm class values determine how fast workers climb their specific job ladders.

Conditional on a firm classification, it is assumed that in expectations workers are indifferent between two firms belonging to the same class. With no loss of generality, the choice probability between two firms belonging to the same class is assumed to be one half. Under the discretisation of unobserved heterogeneities, the offer arrival rate parameter $\lambda_{lgxk'}$ and perceived value γ_{lgxk} are identified using information on the frequencies of transition probabilities $m\Big(k' \;\Big|\; k, l, g, x\Big)$, together with the normalisation $\sum_{k=1}^{K} \gamma_{lgxk} = 1$. First, λ_{lgxk} is identified for any combination (l, g, x, k) using data of within-class transitions $m\Big(k \;\Big|\; k, l, g, x\Big) = \lambda_{lgxk} \, \frac{1}{2}$.[9] Second, the choice probabilities $P_{lgx}\Big(k' \succ k\Big)$ are pinned down using information from the unrestricted transitions $m\Big(k' \;\Big|\; k, l, g, x\Big)$ and given knowledge of λ_{lgxk}, for any l, g, x, k. Finally, given the normalisation $\sum_{k=1}^{K} \gamma_{lgxk} = 1$, the ratios $\frac{\gamma_{lgxk'}}{\gamma_{lgxk}}$ follow:

$$\frac{P_{lgx}\Big(k' \succ k\Big)}{P_{lgx}\Big(k \succ k'\Big)} = \frac{\gamma_{lgxk'}}{\gamma_{lgxk}}$$

Conditional on meeting, if a higher number of workers of type (l, g, x) move from k to k' than from k' to k then we may infer that (l, g, x)-type workers prefer firms of class k' better than firms of class k. This is the basic principle behind the worker-specific firm values estimation, and this is what is intended by preferences throughout the paper.

Transitions to and from non-employment. At any time period $t \in \{1, \ldots, T-1\}$, a (l, g, x)-type worker moves from a firm of class $k = 1, \ldots, K$ to non-employment $k = 0$ with

9. Under the assumption of no zero cells in the worker-specific job-to-job transition matrices.

probability $m\Big(0 \;\Big|\; k, l, g, x\Big) = \delta_{lgxk}$. The worker moves from non-employment to a firm of class $k' = 1, \ldots, K$ with probability $m\Big(k' \;\Big|\; 0, l, g, x\Big) = \psi_{lgxk'}$.

Transitions to and from non-employment are left completely unrestricted, and are identified by simple frequencies. Moving into non-employment depends on worker type and on current firm class, moving into employment depends on worker type and new firm class. $m\Big(0 \;\Big|\; 0, l, g, x\Big) = 0$ as there are no transitions from non-employment into non-employment.

Given the specification of the transition probability parameters, it follows that the probability of staying into non-employment is:

$$m\Big(\neg \mid 0, l, g, x\Big) = 1 - \sum_{k'=1}^{K} \psi_{lgxk'}$$

For employed workers, $k \geq 1$, the probability of staying with the same firm is:

$$m\Big(\neg \mid k, l, g, x\Big) = 1 - \delta_{lgxk} - \sum_{k'=1}^{K} \left(\lambda_{lgxk'} \; \frac{\gamma_{lgxk'}}{\gamma_{lgxk} + \gamma_{lgxk'}} \right)$$

Hourly wage distributions

Hourly wages are drawn from a static worker-firm-specific log-normal distribution:

$$\ln f\Big(y_{it} \;\Big|\; l, g, x, k\Big) = -\ln(\sigma_{lgxk}) - \ln(\sqrt{2\pi}) - \frac{1}{2}\left(\frac{y_{it} - \mu_{lgxk}}{\sigma_{lgxk}}\right)^2$$

Hourly wages are recorded at annual frequency, that is there is only one payroll recorded for each employment spell in a year. I include just one hourly wage likelihood contribution for each spell-year observation, so that y_{it} refers to the unique wage observation for the first month of the current spell-year, and y_{it+1} refers to the unique wage observation for the first month of the subsequent spell-year.

Estimates of μ_{lgxk} and σ_{lgxk}, for any match (l, g, x, k), will be used to compute the gender wage gap in a framework that allows for earnings complementarities between workers and firms.

3 Estimation

Firm classes and worker types are unobserved. The mobility of workers across firm types makes it difficult to separate the complete log-likelihood across firm types. I, therefore, proceed with a two-step estimation as in Bonhomme et al. (2019). First, I cluster firms into classes using a k-means algorithm. Second, conditional on the firm clustering, I use an Expectation-Maximisation (EM) algorithm to iterate over (a) the calculation of the posterior probability that worker i is of type $l = 1, \ldots, L$, and (b) the maximisation of the expected log-likelihood with respect to the parameters of interest.[10]

3.1 K-Means Algorithm to Cluster Firms

In the model described in section 2, the initial matching distributions, log-wages, and mobility patterns depend on firm classes but not directly on firm identities. The idea is that unobserved firm heterogeneity is captured at the class level and not at the individual firm level. I, therefore, partition the J firms into a finite number of classes, K, solving a weighted k-means problem: I use as input characteristics of each firm male and female empirical cdfs of log-hourly wages, female shares, male and female inflow and outflow rates, and I weight by average firm size. As I want to estimate earnings distributions of male and female workers and their mobility patterns across firm types, I need the k-means algorithm to take care of firms' behaviour towards a specific gender.

I residualise log-hourly wages on a third-order polynomial in age and on 3-digit occupational dummies. I restrict the age profile to be flat at 40, attained by considering a cubic polynomial in (age $-$ 40). Age and occupational effects are obtained from the female sample only, and both male and female log-hourly wages are purged using these effects. I do so to control for observed workers' skills, proxied by occupations, without imposing similar (or different) returns between men and women. The EM algorithm is performed on log-hourly wages residualised on the same age and occupational effects estimates.

Using the elbow-method criterion to select the optimal number of clusters, I choose

10. In all fairness, Lentz et al. (2023) classify both firms and workers in the EM algorithm. They treat firm types as parameters to be estimated in the expected likelihood maximisation along with the other parameters. This has the advantage of fully using the information on both wages and mobility for both workers and firms.

$K = 10$. Appendix C presents more details about the k-means algorithm.

3.2 EM Algorithm to Classify Workers

Worker types are unobserved. The EM algorithm classifies workers into a discrete number of types L by iterating an expectation step and a maximisation step until convergence. I set $L = 3$ and show the fit of the model in Appendix A.[11]

Expectation step

For given parameters $\theta^{(m)}$ and a firm classification C, compute the posterior probability that worker i is of type $l = 1, \ldots, L$:[12]

$$p_i\Big(l \;\Big|\; \theta^{(m)}, g_i, x_{i1}, C\Big) = \frac{\mathcal{L}_i\Big(\theta^{(m)} \;\Big|\; l_i, g_i, x_{i1}, C\Big)}{\sum_{l=1}^{L} \mathcal{L}_i\Big(\theta^{(m)} \;\Big|\; l_i, g_i, x_{i1}, C\Big)} \tag{2}$$

Maximisation step

Maximise the expected log-likelihood with respect to the parameter of interest θ:

$$\sum_i \sum_t \sum_l \sum_{k,k'} p_i\Big(l \;\Big|\; \theta^{(m)}, g_i, x_{i1}, C\Big)\; ln\mathcal{L}_i\Big(\theta \;\Big|\; l_i, g_i, x_{i1}, C\Big)$$

where k refers to the firm class at t, and k' refers to the firm class $t+1$, for any $t = 1, \ldots, T$. The maximisation step gives the updated $\theta^{(m+1)}$, used to update the posterior probability in equation 2. Iterate between the expectation step and maximisation step until convergence.

The maximisation step updating formulas for the wage distributions are simple weighted averages and variances, using the posterior probabilities as weights. The maximisation step for the initial matching distribution and for unrestricted transition probabilities are simple frequencies. Transition probabilities for job-to-job mobility are non-linear in the parameters

11. Although the small number of points of support is computationally convenient, it can be shown that just few points of support approximate well the underlying distribution of unit fixed effects and their correlation with covariates. I thank Seth Sanders for the stimulating discussion.

12. Any individual firm sampling factor in equation 1, conditional on a firm classification C, would cancel out in equation 2.

of interest. In Appendix D I detail the minorisation-maximisation (MM) algorithm proposed by Lentz et al. (2023) to maximise the expected log-likelihood for the non-linear cases.

4 Data and Sample Selection

I use the French matched employer-employee data, *Déclarations Annuelles de Données Sociales* (DADS), over the period 2016-2019. The DADS datasets are extensive collections of mandatory employer reports of salaried employees, gathered by the French statistical institute *Institut National de la Statistique et des Etudes Economiques* (INSEE). The data contain job-spell-level information on worker-firm matches. Notably, working hours are reported, allowing me to control for gender differences in labour supply. In order to estimate the model described in Section 2, I use two data sources: the DADS-Postes, and the DADS-Panel.

DADS-Postes

The DADS-Postes dataset collects information on the universe of jobs in France. It does not provide a proper longitudinal dimension, as worker identifiers change every two years. As it provides complete yearly employment information for each firm, I use the DADS-Postes for the firm clustering described in section 3.1.

I select firms active in 2016, employing at least ten employees and employing each gender. I consider employment any job spell with positive wages and hours. Wages are reported at annual frequency.[13] I have information on firms' gender composition, wage distributions, workers' basic demographics, and occupations. I recover information on yearly inflow/outflow rates through worker job mobility over two consecutive years. In addition, I have information on firms' sectors and public/private status. I present ex-post tabulations of firms' observed characteristics by predicted latent class in section 5.

DADS-Panel

The DADS-Panel dataset contains information on the employment history of workers born in October. I use the DADS-Panel for the maximum likelihood estimation conditional on firm clustering.

13. I winsorize hourly wages at 0.001 and 0.999.

For each job spell, I have information on worker identifier, firm identifier, year, yearly earnings, hours worked, occupation, worker age, worker tenure in the firm, starting day of the spell in the year (between 1 and 360), ending day of the spell in the year (between 1 and 360). I construct a monthly panel of individual working trajectories. I obtain hourly wages by dividing yearly earnings by the hours worked. I keep a record of log hourly wages just for the first month of each spell-year. I consider job spells with positive wages and positive hours. I keep part-time and full-time contracts lasting at least a month, remove seasonal contracts, and retain workers who have never worked in the Agricultural sector. Non-employment is intended as the time in between different job spells.[14]

I select workers employed in January 2016, only working in *Ile-de-France* over the period 2016-2019, aged between 25 and 55, and only working in firms selected from the DADS-Postes dataset over the period 2016-2019. I track the selected workers over time.[15]

I focus on *Ile-de-France* simply to reduce the sample size. I consider a short panel of four years as I want to estimate time-invariant type and class effects and time-invariant offer arrival rates parameters. It would be implausible to assume that a worker unobserved ability is constant over long periods. The same reasoning holds for job arrival rates.

Table 1 presents the selected sample. Females' average annual earnings are 22% lower than the ones of males. Compared to men, women, on average, work 5% hours less, are more likely to work part-time jobs, and are less likely to be managers. Regarding mobility, men and women are similarly likely to do at least one monthly job-to-job transition over the period of interest. Women are slightly more likely than men to move from employment to non-employment between two consecutive months and more likely to move from non-employment to employment. The average monthly length of job spells over the period of interest is quite similar between men and women. Finally, women are 49% of the sample.

14. Non-employment does not include periods of maternity leave in a firm or retirement, but it could include periods of inactivity. As I do not observe education, I consider only workers aged 25+ to minimise the probability that non-employment gets confounded with periods of education.

15. Table B1 describes in details the sample selection.

5 Firm Classes and Worker Types

I now present ex-post tabulations of firm and worker observed heterogeneity by predicted latent classes k and types l, respectively. Latent classes and types are relabelled so that they are increasing in average log-hourly wages.[16] I therefore refer to higher latent groups as higher-paying firms and higher-wage workers.

Table 2 contains descriptive statistics for firms relative to the year 2016. Firm classes that, on average, pay more tend to pay more both men and women. Classes markedly differ in female shares, with some classes hiring mainly women and other classes hiring mainly men. Classes that hire mainly men tend to require longer work hours and concentrate more in the Construction and Commerce sectors. Classes that hire mainly women tend to operate in the Commerce, Education, Managing, and Public Administration sectors. The top-paying class hires an equal share of men and women and concentrates in the Commerce, Managing, and Finance sectors. Women are under-represented in the lowest-paying class, where Commerce, Hotel, and Managing are the dominant sectors.

Tables 3, 4, 5 describe how observed worker characteristics are distributed across latent worker types l. I tabulate by multiple combinations of age, gender, and tenure to zoom into worker characteristics and show how latent worker heterogeneity relates to observed characteristics. Within latent types, annual earnings are increasing in age and tenure, while long-tenured workers are less likely to experience at least one job transition than short-tenured workers. Across latent types, higher-wage workers are more likely to separate at least once than lower-wage workers. Compared to men, women earn on average less, work shorter hours, and are more likely to hold part-time jobs both within latent types and by different combinations of age-tenure groups. However, men and women tend to have similar observed characteristics. Regarding mobility, men and women are equally likely to make at least a job-to-job move over the sample period, both within the latent type and by age-tenure combinations. Young- and mid-age women are more likely to move into and from non-employment than young- and mid-age men. Senior-age women and senior-age men share similar mobility rates.

16. A formal description of class relabelling is provided in Section 6.

6 Worker Sorting and the Gender Wage Gap

In this section, I explore gender-based differences in worker sorting across firms. The goal is to assess the *sorting effect*, that is, the share of the gender wage gap explained by men and women being unequally distributed across firms, and to quantify the relative importance of patterns in offer arrival rates, exit rates, and perceived values of firm classes in driving the sorting effect. To this purpose, I obtain the stationary matching distribution from the predicted worker-firm stationary allocation and worker-type marginal distributions and run a series of counterfactual exercises.

The worker-firm stationary allocation

Sorting is the stationary allocation of worker types and firm classes, $\Pr^*\Big(k \;\Big|\; l, g, x\Big)$, for any $l \in \{1, \ldots, L\}$, $k \in \{0, 1, \ldots, K\}$, $g \in \{F, M\}$, and combinations of observed characteristics x (short/long tenure, and age groups). For each (l, g, x)-type, I build a $(K+1) \times (K+1)$ transition matrix, M_{lgx}, with k'-th row and k-th column cell corresponding to the (l, g, x)-specific estimated probability of moving from $k = 0, 1, \ldots, K$ to $k' = 0, 1, \ldots, K$. The stationary allocation is the eigenvector corresponding to an eigenvalue of 1 of the transition matrix M_{lgx}.

The stationary matching distribution

I obtain the empirical distribution of matches, $\Pr\Big(l, g, x, k\Big)$, using the sorting distribution and worker type frequencies:

$$\Pr\Big(l, g, x, k\Big) = \Pr^*\Big(k \;\Big|\; l, g, x\Big) \Pr\Big(l \;\Big|\; g, x\Big) \Pr\Big(g \;\Big|\; x\Big) \Pr\Big(x\Big)$$

where $\Pr^*\Big(k \;\Big|\; l, g, x\Big)$ is the stationary allocation of worker types and firm classes. $\Pr\Big(l \;\Big|\; g, x\Big)$ is the marginal distribution of worker types l for a given combination of observed characteristics. $\Pr\Big(g \;\Big|\; x\Big)$ and $\Pr\Big(x\Big)$ are simple empirical frequencies. I augment the initial sample size by L, and simulate a cross-sectional dataset from the empirical distribution of matches $\Pr\Big(l, g, x, k\Big)$, in which log-hourly wages are drawn using the estimated log-normal distributions centred in $\hat{\mu}_{lgxk}$ with variance $\hat{\sigma}^2_{lgxk}$. Workers earn no wage in non-employment.

The sorting effect and its determinants

The *sorting effect* is the percentage change difference between the gender wage gap estimated with gender-based differences in worker-firm allocations and the gender wage gap estimated under a counterfactual scenario in which men and women are equally distributed across firms. I obtain the counterfactual distribution of matches by equating the worker-firm allocations of women to the ones of men. For each $g \in \{F, M\}$, the counterfactual distribution of matches is:

$$\tilde{\Pr}\Big(l, g, x, k\Big) = \Pr^{*}\Big(k \;\Big|\; l, M, x\Big) \Pr\Big(l \;\Big|\; g, x\Big) \Pr\Big(g \;\Big|\; x\Big) \Pr\Big(x\Big)$$

In practice, I simulate two cross-sectional datasets: one from $\Pr\Big(l, g, x, k\Big)$, one from $\tilde{\Pr}\Big(l, g, x, k\Big)$. The percentage change difference between the gender wage gap estimated from the former dataset and the gender wage gap estimated from the latter is the *sorting effect*.

I quantify the relative importance of offer arrival rates while employed, exit rates, offer arrival rates while non-employed, and workers' perceived firm values by equating the estimated mobility parameters of female workers to the ones of male workers. Whenever I equate a mobility parameter, I predict the stationary worker-firm allocations, obtain the corresponding empirical matching distribution, simulate a cross-sectional dataset, draw wages, and compute the gender wage gap.

Latent types relabelling

Latent worker types l and latent firm classes k are per se meaningless, they simply capture unobserved heterogeneity of groups of workers and firms. To facilitate the interpretation of latent types and classes, I consider the standard two-way fixed effects projection of the estimated mean wages μ_{lgxk} with respect to the matching distribution $\Pr\Big(l, g, x, k\Big)$:

$$\mu_{lgxk} = \bar{\mu}_{gx} + \alpha_l + \psi_k + \tilde{\mu}_{lgxk}$$

where $\bar{\mu}_{gx}$ are interactions among observable characteristics, α_l is the worker effect, ψ_k is the firm effect, and $\tilde{\mu}_{lgxk}$ captures all remaining interactions. I relabel l and k so that α_l and ψ_k are increasing in l and k, respectively. The relabelling allows to interpret higher l as higher-wage worker types, and higher k as higher-paying firm classes.

Results

The top panel of Figure 1 compares how men and women are distributed across firm classes. Firm classes are relabelled so that $k = 1$ is the bottom-paying class and $k = 10$ is the top-paying class. The marginal distribution of firm classes for men is close to uniform, while some firm classes are more popular among women than other firm classes. 50% of men work in the top-paying half of firm classes, while 40% of women sort there. The gender composition across firm classes shown in the top panel of Figure 1 helps to roughly visualise the reallocation of women across firm classes in my counterfactual analyses. Whether and to what extent the reallocation would affect the gender wage gap depends on the wage complementarities between worker and firm groups.

The bottom panel of Figure 1 presents six counterfactual exercises run on the full cross-sectional dataset. I describe the results reading from left to right. *a*) The gender wage gap under the estimated worker-firm allocations, with men and women being unequally distributed across firms, is 9.4 log-points. *b*) If employed female workers receive job offers at the same rate as their male counterparts, the gender wage gap would be unaffected. If anything, it would slightly increase, indicating some worker-firm complementaries based on wages at the job offer stage. *c*) If employed female workers move into non-employment at the same rate as employed men, the gender wage gap is unaffected. The gender composition of employed worker types across firm classes would not change in a way that would affect the gender wage gap. *d*) If non-employed women receive job offers at the same rate as men, the gender wage gap would decrease to 8.7 log points. Non-employed women would be more likely to move to firm classes that would pay them better, and therefore the range over which the gender wage gap lies would shift downward. *e*) Conditional on drawing offers at a worker-firm-specific frequency, if women have the same preferences over firms as men, the gender wage gap would decrease to 8.5 log points. Recall that preferences and wages are two sets of separate parameters. Workers may value something beyond wages, and we infer from this result that amenities may play an important role in female sorting patterns. Under this scenario, the overall gender wage gap would reduce by 10%.[17] *f*) Finally, equating all the

17. Recall that the gender wage gap has been previously residualised on female occupational and age profile effects.

female sorting parameters to the ones of male workers, the gender wage gap would reduce to 8.1 log points. Overall, the sorting effect accounts for 14% of the gender wage gap.

Figure 2 replicates the exercises for different combinations of age groups. The overall sorting effect is entirely driven by young- and mid-age groups. The gender wage gap among 25-35 years old workers is 8.8 log points. If men and women in this group were equally distributed across firms, the gender wage gap would reduce to 6.6 log points. The major sorting determinant is offer arrival rates while non-employed. Where young women work after a period of non-employment affects the overall gender wage gap. The sorting effect for mid-age women is 20%. Gender differences in how mid-age workers value firms are the dominant sorting channel.

7 Discussion

This article documents the relative importance of multiple mobility patterns and channels in determining gender wage differentials. I focus on three types of transitions: *i*) firm-to-firm, *ii*) into non-employment, *iii*) out of non-employment. I adopt the finite mixture approach of Lentz et al. (2023) to preserve rich sources of two-sided heterogeneity. Firm-to-firm transition probabilities are modelled as the product between offer arrival rates and choice probabilities. Using a revealed preference argument as in Sorkin (2018), worker preferences guide job offer acceptance decisions: upon receiving a job offer, the higher the perceived value of the poacher, the higher the probability the worker decides to accept the offer. Offer arrival rates and worker preferences are separately identified under a discretisation of worker and firm heterogeneity and a distributional assumption on idiosyncratic factors affecting workers' mobility decisions. I estimate the model in two steps, similarly to Bonhomme et al. (2019).

I use French monthly matched employer-employee data and focus on the Paris region from 2016-2019. Through counterfactual exercises, I find that if women and men were equally distributed across firms, the overall gender wage gap would reduce by 14 percent. Differences in preferences account for up to 70 percent of this sorting component of the gender wage gap, and they manifest among workers who are 36-45 years old.

Several caveats are worth discussing. First, the framework does not consider idiosyncratic shocks to wages and layoff notifications as explanations of moves observed in the data. I confound voluntary and involuntary transitions in the revealed preference argument. However, guessing how these omissions lead to biased estimates is not straightforward. I use all the flows made by groups of workers across groups of firms over the period of interest to extract a 'systematic' pattern of preferences.

Second, I focus on the cross-sectional average of the gender wage gap at different worker age groups, thus not considering age dynamics as done, for example, in Barth et al. (2021). Although I explicitly model mobility and do not solely rely on a log earnings model, my estimates document the role of multiple mobility patterns for three different age groups over a monthly panel of recent years.

Finally, more work on the occupational and spatial aspects should be done. These are substantive directions for future research.

Nevertheless, the novelty of this paper is to allow for both within and between gender heterogeneity so that the model generates rich sources of sorting between workers and firms. The sorting of men and women across different market segments may be based on wages or non-wage amenities. Sorting based on preferences for amenities account for 10 percent of the residualised gender wage gap. I do not take a stand on where these preferences come from. Suppose women are making choices so to balance caregiving responsibilities. In that case, the results of this paper should not imply that there is no need for any type of corrective action, as these choices may be endogenous to sticky gender stereotypes that blur what we define preferences (Bertrand 2020). This article suggests that combining administrative data with a flexible model incorporating worker heterogeneity in mobility decisions may contribute to our understanding of how male and female workers flow across firms and demands further research to design policies that enhance worker and firm allocations.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Angelici, M. and P. Profeta (2020): "Smart-Working: Work Flexibility Without Constraints," CESifo Working Paper 8165.

Bagger, J. and R. Lentz (2019): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86, 153–190.

Barth, E., S. P. Kerr, and C. Olivetti (2021): "The dynamics of gender earnings differentials: Evidence from establishment data," *European Economic Review*, 134, 103713.

Bertrand, M. (2020): "Gender in the Twenty-First Century," *AEA Papers and Proceedings*, 110, 1–24.

Blau, F. D. (1977): *Equal Pay in the Office*, Washington, DC: Lexington Books.

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

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023): "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, 41, 291–322.

Bonhomme, S. and G. Jolivet (2009): "The Pervasive Absence of Compensating Differentials," *Journal of Applied Econometrics*, 24, 763–795.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87, 699–739.

Bowlus, A. J. (1997): "A Search Interpretation of Male-Female Wage Differentials," *Journal of Labor Economics*, 15, 625–657.

Bradley, R. A. and M. E. Terry (1952): "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, 39, 324–345.

BRAUN, C. AND A. FIGUEIREDO (2022): "Labor Market Beliefs and the Gender Wage Gap," Working paper. https://doi.org/10.25397/eur.20391102.v1, Erasmus University Rotterdam (EUR).

BRUNS, B. (2019): "Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap," *American Economic Journal: Applied Economics*, 11, 74–113.

CARD, D., A. R. CARDOSO, AND P. 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, 633–686.

CASARICO, A. AND S. LATTANZIO (2022): "What Firms Do: Gender Inequality in Linked Employer-Employee Data," *Journal of Labor Economics*, forthcoming.

CORTÉS, P., J. PAN, L. PILOSSOPH, E. REUBEN, AND B. ZAFAR (2023): "Gender Differences in Job Search and the Earnings Gap: Evidence from the Field and Lab*," *The Quarterly Journal of Economics*, qjad017, https://doi.org/10.1093/qje/qjad017.

COUDIN, E., S. MAILLARD, AND M. TO (2018): "Family, Firms and the Gender Wage Gap in France," IFS Working Papers W18/01, Institute for Fiscal Studies.

CRUZ, G. AND T. RAU (2022): "The effects of equal pay laws on firm pay premiums: Evidence from Chile," *Labour Economics*, 75, 102135.

DEL BONO, E. AND D. VURI (2011): "Job mobility and the gender wage gap in Italy," *Labour Economics*, 18, 130–142.

FLUCHTMANN, J., A. GLENNY, N. A. HARMON, AND J. MAIBOM (2021): "The Gender Application Gap: Do Men and Women Apply for the Same Jobs?" Discussion Paper 14906, IZA.

HIRSCHMAN, D. (2022): "Controlling for What? Movements, Measures, and Meanings in the US Gender Wage Gap Debate," *History of Political Economy*, 54, 221–257.

HUNTER, D. R. (2004): "MM algorithms for generalized Bradley-Terry models," *The Annals of Statistics*, 32, 384 – 406.

Hunter, D. R. and K. Lange (2004): "A Tutorial on MM Algorithms," *The American Statistician*, 58, 30–37.

Kleven, H., C. Landais, and J. E. Søgaard (2019): "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11, 181–209.

Kline, P., E. K. Rose, and C. R. Walters (2022): "Systemic Discrimination Among Large U.S. Employers," *The Quarterly Journal of Economics*, forthcoming.

Le Barbanchon, T., R. Rathelot, and A. Roulet (2020): "Gender Differences in Job Search: Trading off Commute against Wage*," *The Quarterly Journal of Economics*, 136, 381–426.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2023): "The Anatomy of Sorting - Evidence from Danish Data," *Econometrica*, forthcoming.

Li, J., B. Dostie, and G. Simard-Duplain (2020): "What Is the Role of Firm-Specific Pay Policies on the Gender Earnings Gap in Canada?" IZA Discussion Paper 13907.

Loprest, P. J. (1992): "Gender Differences in Wage Growth and Job Mobility," *The American Economic Review*, 82, 526–532.

Mas, A. and A. Pallais (2017): "Valuing Alternative Work Arrangements," *American Economic Review*, 107, 3722–59.

Morchio, I. and C. Moser (2020): "The Gender Pay Gap: Micro Sources and Macro Consequences," Discussion Paper 16383, CEPR.

Neumark, D., R. J. Bank, and K. D. V. Nort (1996): "Sex Discrimination in Restaurant Hiring: An Audit Study," *The Quarterly Journal of Economics*, 111, 915–941.

Olivetti, C. and B. Petrongolo (2016): "The Evolution of Gender Gaps in Industrialized Countries," *Annual Review of Economics*, 8, 405–434.

Palladino, M. G., A. Roulet, and M. Stabile (2021): "The Gender Pay Gap: Micro Sources and Macro Consequences," Discussion Paper 16671, CEPR.

Petrongolo, B. and M. Ronchi (2020): "Gender gaps and the structure of local labor markets," *Labour Economics*, 64, 101819, european Association of Labour Economists, 31st annual conference, Uppsala Sweden, 19-21 September 2019.

Sorkin, I. (2017): "The Role of Firms in Gender Earnings Inequality: Evidence from the United States," *American Economic Review: Papers and Proceedings*, 107, 384–87.

——— (2018): "Ranking Firms Using Revealed Preference*," *The Quarterly Journal of Economics*, 133, 1331–1393.

Taber, C. and R. Vejlin (2020): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Econometrica*, 88, 1031–1069.

Wiswall, M. and B. Zafar (2018): "Preference for the Workplace, Investment in Human Capital, and Gender*," *The Quarterly Journal of Economics*, 133, 457–507.

Xiao, P. (2021): "Wage and Employment Discrimination by Gender in Labor Market Equilibrium," Working Paper 144, VATT Institute for Economic Research.

Zurla, V. (2022): "How Should We Design Parental Leave Policies? Evidence from Two Reforms in Italy," Working paper.

8 Tables

Table 1: Sample description

Gender:	Women	Men
Avg Annual Earnings	39,815	51,363
Avg Hours	1,679	1,771
Share Part-Time	16%	5%
Share doing JTJ	16%	15%
Share doing E-NE	21%	18%
Share doing NE-E	51%	42%
Length job spell (in months)	36	37
N	96,058	99,679

Notes: The table presents descriptive statistics for the selected sample over which the analysis is implemented. Data relative to mobility represents the share of workers doing at least one given transition. JTJ stands for Job-To-Job. E stands for Employment. NE stands for Non-Employment. Numbers in the table represent averages computed over the pooled period January 2016 - December 2019.

Table 2: Firm classes description

Firm class:	1	2	3	4	5	6	7	8	9	10
N firms	6,300	5,320	3,341	5,055	3,471	4,128	4,112	4,672	3,867	4,080
Female share	34%	68%	68%	24%	30%	65%	22%	63%	27%	47%
Avg N workers	80	157	173	189	313	409	241	206	211	116
Log hourly wage - women	2.53	2.66	2.68	2.72	2.74	2.77	2.88	2.94	3.01	3.26
Log hourly wage - men	2.57	2.71	2.73	2.81	2.83	2.86	2.95	3.10	3.14	3.50
Avg hours - women	1,275	1,284	1,134	1,445	1,045	1,397	1,526	1,387	1,528	1,478
Avg hours - men	1,330	1,320	1,116	1,516	1,170	1,458	1,592	1,431	1,602	1,542
% Women in top 10 occ	10%	11%	6%	11%	12%	9%	12%	17%	16%	24%
% Men in top 10 occ	11%	18%	9%	11%	12%	16%	13%	27%	20%	36%
% Firms in Hotel	17%	9%	11%	9%	28%	6%	4%	6%	5%	3%
% Firms in Admin Services	6%	4%	14%	5%	11%	5%	5%	6%	4%	5%
% Firms in Construction	6%	0%	0%	14%	4%	0%	21%	1%	21%	7%
% Firms in Commerce	21%	18%	14%	22%	11%	16%	22%	18%	23%	21%
% Firms in Education	3%	13%	18%	1%	4%	11%	1%	7%	2%	1%
% Firms in Managing	9%	11%	7%	8%	9%	9%	8%	19%	9%	18%
% Firms in Finance	1%	3%	2%	1%	3%	4%	2%	8%	5%	13%
% Firms in Pub Admin	0%	4%	2%	0%	1%	16%	1%	3%	1%	1%
% Firms in Health Accomm	2%	7%	13%	0%	0%	7%	0%	2%	0%	1%

Notes: Ex-post tabulations of observed firm characteristics by predicted latent firm classes. Firm classes are obtained using the k-means algorithm described in Section 3.1. Classes are relabelled so that they are increasing in average log-hourly wages. All numbers presented in the table correspond to year 2016.

Table 3: Worker types description - Aged 25-35

	Young, short-tenured					
Worker type:	1		2		3	
Gender:	F	M	F	M	F	M
Avg annual earnings	26,436	28,870	28,054	34,300	30,795	37,971
Avg hours	1,599	1,672	1,593	1,696	1,489	1,580
Avg age	29	29	29	29	29	29
Avg experience	7	7	6	7	6	7
Avg tenure	1	1	1	1	1	1
% with part-time	13%	8%	12%	4%	12%	7%
Share doing JTJ	26%	26%	25%	22%	33%	31%
Share doing E-NE	34%	28%	28%	28%	54%	46%
Share doing NE-E	60%	54%	45%	32%	53%	41%
Share	28%	35%	34%	32%	39%	32%
N	3,883	4,837	4,737	4,473	5,498	4,459

	Young, long-tenured					
Worker type:	1		2		3	
Gender:	F	M	F	M	F	M
Avg annual earnings	31,792	32,925	28,887	37,101	34,631	45,364
Avg hours	1,709	1,771	1,699	1,807	1,594	1,703
Avg age	30	31	31	31	31	31
Avg experience	9	9	9	9	9	9
Avg tenure	5	5	5	6	5	5
% with part-time	12%	4%	14%	3%	12%	5%
Share doing JTJ	15%	17%	16%	11%	27%	28%
Share doing E-NE	16%	14%	17%	12%	41%	33%
Share doing NE-E	59%	53%	59%	25%	50%	43%
Share	29%	35%	35%	34%	37%	31%
N	5,032	6,059	6,150	5,975	6,440	5,336

Notes: Ex-post tabulations of worker characteristics by predicted latent types. Workers are classified into types using the EM algorithm described in Section 3.2. Short tenure in the firm is defined to be less than two years. Data relative to earnings, hours, age, experience, tenure, and part-time contracts correspond to January 2016. Data relative to mobility represents the share of workers doing at least one given transition over the period January 2016 - December 2019. JTJ stands for Job-To-Job. E stands for Employment. NE stands for Non-Employment.

Table 4: Worker types description - Aged 36-45

	Mid, short-tenured					
Worker type:	1		2		3	
Gender:	F	M	F	M	F	M
Avg annual earnings	24,627	28,497	33,105	44,792	48,500	70,562
Avg hours	1,553	1,655	1,605	1,774	1,548	1,648
Avg age	40	40	39	39	39	40
Avg experience	15	16	15	16	15	15
Avg tenure	1	1	1	1	1	1
% with part-time	22%	9%	20%	3%	15%	7%
Share doing JTJ	20%	20%	20%	17%	27%	25%
Share doing E-NE	31%	30%	19%	16%	41%	37%
Share doing NE-E	56%	51%	59%	47%	50%	38%
Share	36%	41%	33%	30%	32%	28%
N	2,862	3,657	2,645	2,687	2,553	2,528

	Mid, long-tenured					
Worker type:	1		2		3	
Gender:	F	M	F	M	F	M
Avg annual earnings	32,966	36,546	38,997	47,939	49,981	76,244
Avg hours	1,710	1,802	1,734	1,827	1,645	1,752
Avg age	40	40	40	40	40	40
Avg experience	17	17	17	17	16	17
Avg tenure	8	8	9	9	8	8
% with part-time	20%	4%	19%	3%	19%	5%
Share doing JTJ	9%	10%	9%	7%	18%	20%
Share doing E-NE	11%	10%	7%	5%	29%	25%
Share doing NE-E	49%	45%	47%	42%	49%	40%
Share	37%	41%	33%	31%	30%	28%
N	10,248	11,995	9,122	9,075	8,228	8,064

Notes: Ex-post tabulations of worker characteristics by predicted latent types. Workers are classified into types using the EM algorithm described in Section 3.2. Short tenure in the firm is defined to be less than two years. Data relative to earnings, hours, age, experience, tenure, and part-time contracts correspond to January 2016. Data relative to mobility represents the share of workers doing at least one given transition over the period January 2016 - December 2019. JTJ stands for Job-To-Job. E stands for Employment. NE stands for Non-Employment.

Table 5: Worker types description - Aged 46-55

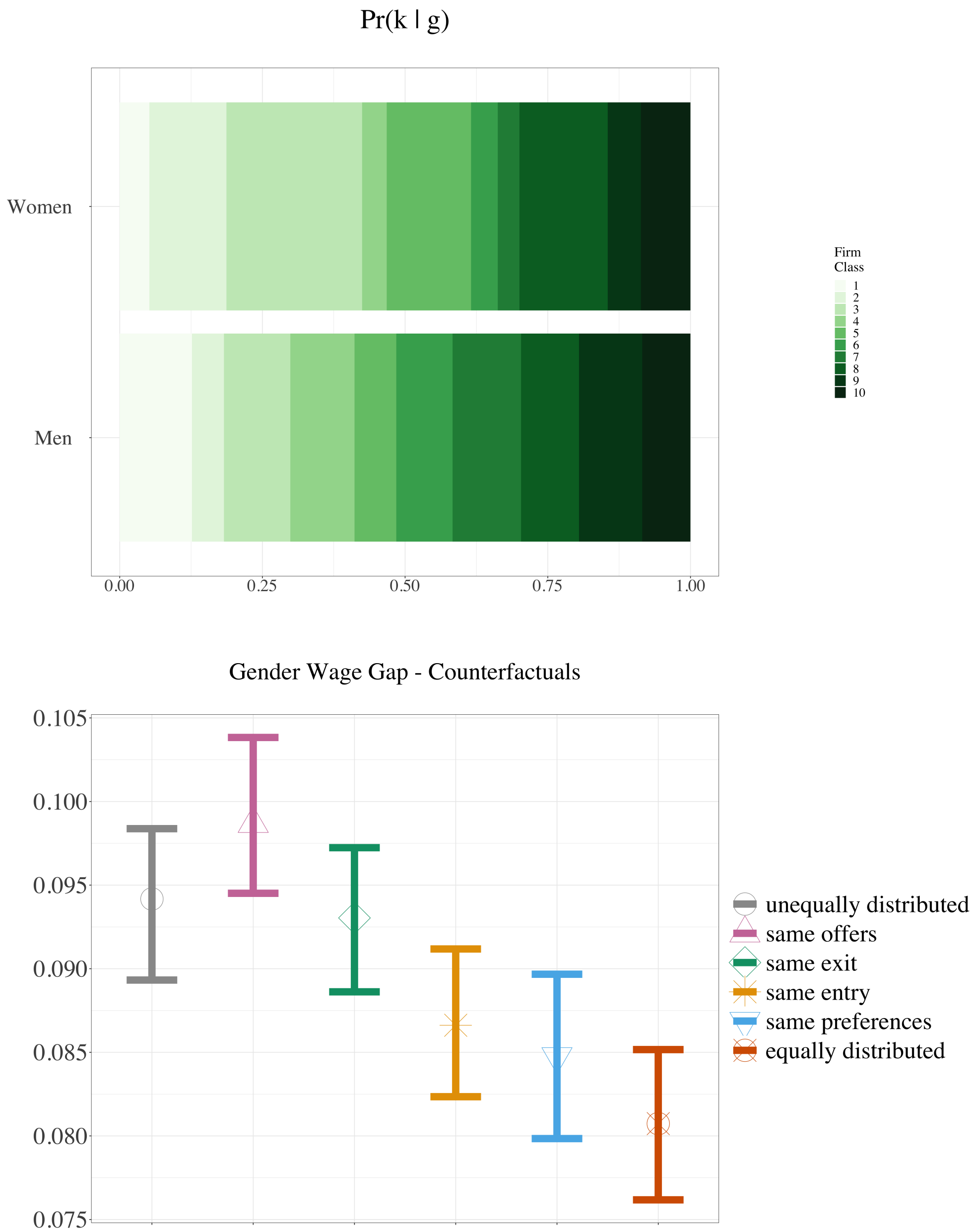
	Senior, short-tenured					
Worker type:	1		2		3	
Gender:	F	M	F	M	F	M
Avg annual earnings	25,020	30,521	39,902	50,381	57,309	92,956
Avg hours	1,536	1,694	1,684	1,790	1,556	1,645
Avg age	49	49	49	48	48	49
Avg experience	23	24	24	24	23	23
Avg tenure	1	1	1	1	1	1
% with part-time	26%	9%	16%	3%	19%	11%
Share doing JTJ	18%	16%	17%	12%	23%	22%
Share doing E-NE	25%	26%	13%	14%	36%	37%
Share doing NE-E	56%	52%	34%	45%	52%	36%
Share	43%	43%	29%	28%	29%	29%
N	1,982	2,276	1,329	1,479	1,350	1,499

	Senior, long-tenured					
Worker type:	1		2		3	
Gender:	F	M	F	M	F	M
Avg annual earnings	32,035	40,007	40,687	50,039	59,924	96,727
Avg hours	1,716	1,813	1,762	1,837	1,674	1,765
Avg age	49	49	49	49	49	49
Avg experience	25	25	25	26	24	25
Avg tenure	10	10	11	11	10	10
% with part-time	19%	4%	15%	3%	20%	7%
Share doing JTJ	6%	6%	5%	5%	13%	12%
Share doing E-NE	9%	9%	4%	5%	23%	24%
Share doing NE-E	41%	45%	29%	34%	43%	38%
Share	39%	39%	35%	35%	26%	26%
N	9,268	9,967	8,464	8,737	6,268	6,577

Notes: Ex-post tabulations of worker characteristics by predicted latent types. Workers are classified into types using the EM algorithm described in Section 3.2. Short tenure in the firm is defined to be less than two years. Data relative to earnings, hours, age, experience, tenure, and part-time contracts correspond to January 2016. Data relative to mobility represents the share of workers doing at least one given transition over the period January 2016 - December 2019. JTJ stands for Job-To-Job. E stands for Employment. NE stands for Non-Employment.

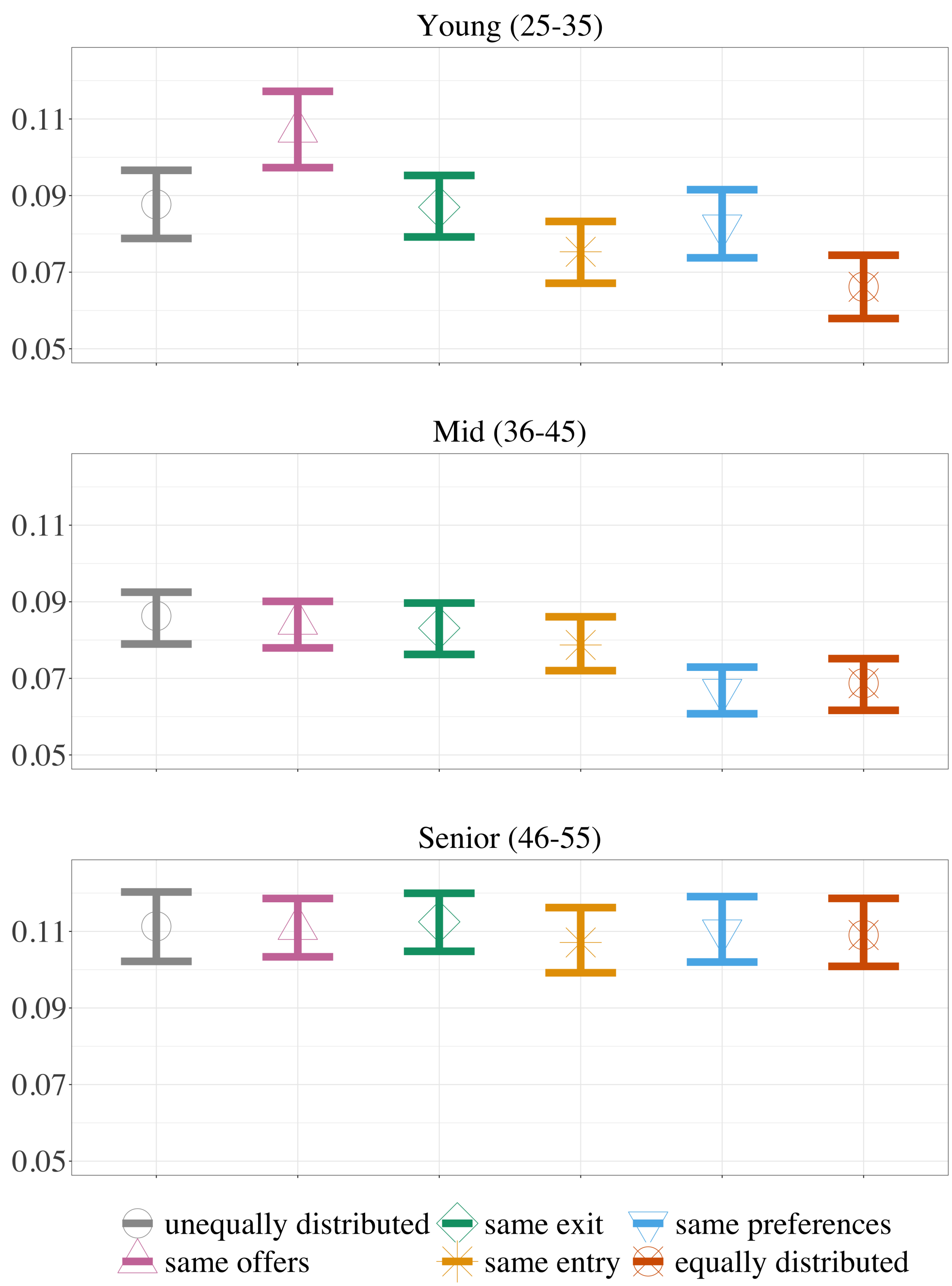
9 Figures

Figure 1: Employment distribution and counterfactual analysis



Notes: The top panel compares how men and women are distributed across firm classes. Firm classes are sorted in increasing average wages. The bottom panel presents six counterfactual exercises run on the stationary matching distribution $\Pr\big(l, g, x, k\big)$. The brackets indicate 5%-95% quantile bands obtained from 200 replications.

Figure 2: Counterfactual analysis by age groups

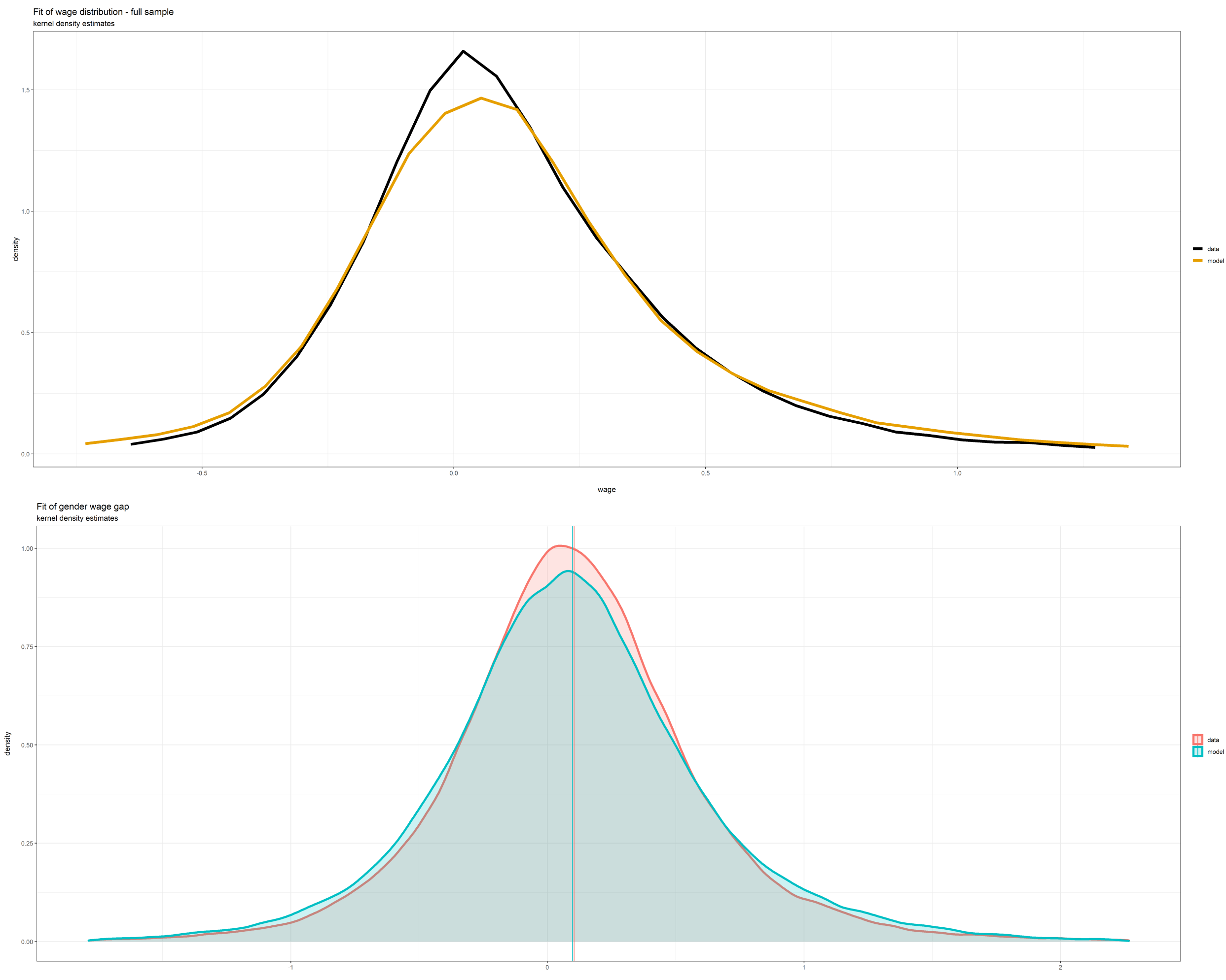


Notes: The figure presents estimates of the gender wage gap for six counterfactual exercises, by different combinations of age groups. The counterfactuals simulate scenarios in which men and women have same patterns in multiple mobility transitions. 'Young' refers to workers aged 25-35, 'Mid' refers to age 36-45, 'Senior' refers to age 46-55. The brackets indicate 5%-95% quantile bands obtained from 200 replications.

Appendix

A Model Fit

Figure A1: Fit of wage distribution and gender wage gap



Notes: The figure shows the model fit of the wage distribution (top panel) and of the gender wage gap (bottom panel).

B Sample selection

Table B1: Worker selection

Selection Step	N Women	N Men
Main job	1,318,122	1,386,365
Ile de France only	283,468	302,629
30+ days contracts	282,350	300,800
+ wages, + hours	281,787	299,956
Employed in Jan 2016	192,352	200,672
Never in Agriculture	192,217	200,423
Part-time & Full-time only	180,711	196,914
Aged 25-55	129,643	142,412
Never in seasonal/internship/domicile	120,471	131,611
Only in firms from DADS-Postes	96,058	99,679

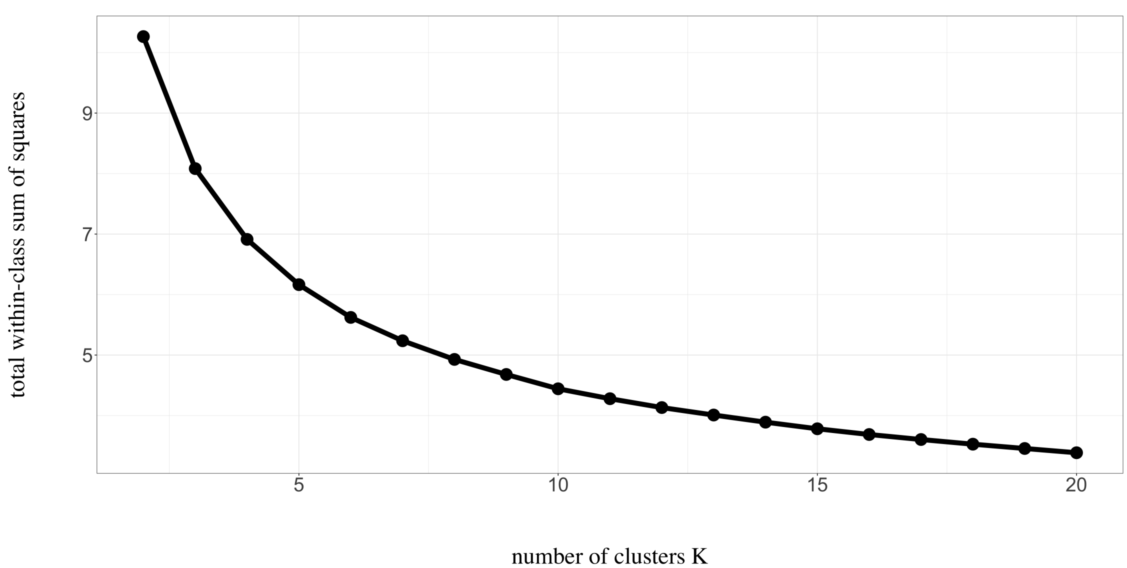
Notes: The table shows the number of workers by gender at each selection step. The data source is DADS-Panel.

C K-means algorithm for firm clustering

I partition the J firms into a finite number K of classes solving a weighted k-means problem: I use as input characteristics firms' male and female empirical cdfs of log-hourly wages, female shares, male and female inflow and outflow rates, and I weight by firms' average size. For a given K, I initiate the algorithm with 100 initial guesses and select the classification with the smallest residual sum of squares.

In order to select the optimal number of clusters K I apply the elbow method, that is I pick the number of classes such that the total intra-class variation is minimised. Figure C1 plots the curve of within-class sum of squares for a given number of classes K. I choose $K = 10$.

Figure C1: Elbow method to select optimal number of firm classes



Notes: Elbow method for pinning down the optimal number of firm clusters. Increasing number of clusters K are shown on the x-axis. On the y-axis the corresponding total within-class sum of squares (rescaled by the total number of firms).

D EM and MM algorithm

This section describes in detail the algorithm of Lentz et al. (2023) to classify workers into a finite number of groups, and to estimate the parameters of interest. The expectation-maximization (EM) algorithm is an iterative method used to numerically find local maximum likelihood parameters in statistical models that commonly involve latent variables in addition to unknown parameters and observed data points. It is particularly attractive and effective for maximum likelihood estimation because, at each iteration, it consistently increases the likelihood by maximizing a simple surrogate function for the log-likelihood. The EM algorithm is a special case of a more general class of optimization algorithms, minorization-maximization (MM) algorithms, that exploit concavity in finding a surrogate function for maximization (Hunter and Lange (2004)).

The basic idea of the MM algorithm is to look for a minorizing function that makes the maximisation step easier. Let the real-valued objective function be $f(\theta)$. A real-valued function $g(\theta|\theta^{(s)})$ is said to be minorizing $f(\theta)$ if $g(\theta|\theta^{(s)}) \leq f(\theta)\ \forall\theta$ and $g(\theta^{(s)}|\theta^{(s)}) = f(\theta^{(s)})$. $\theta^{(s)}$ is the parameter vector obtained at the current iteration, and $g(.)$ is the surrogate function being maximised in the M-step of the algorithm. If $\theta^{(s+1)}$ is the local maximizer of $g(\theta|\theta^{s})$, then $f(\theta^{(s+1)}) \geq f(\theta^{s})$.[18]

The log-likelihood relative to the likelihood in equation 1 is not linear in the mobility components of the parameter vector θ. In order to ease the maximization step I consider a surrogate function proposed by Lentz et al. (2023). The surrogate function is linear in the parameters of interest and therefore it favors a straightforward maximization. In this Appendix, I describe in details the surrogate function of the log-likelihood as well as the first order conditions with respect to the parameters of interest.

As it is a local estimation, I initiate the full estimation with 20 random guess for the parameters of interest. I select the repetition with the highest likelihood.

18. Indeed, $g(\theta^{(s+1)}|\theta^{(s)}) \geq g(\theta^{(s)}|\theta^{(s)})$ by definition. Together with the definition of the function $g(\theta|\theta^{(s)})$, that is $g(\theta|\theta^{(s)}) \leq f(\theta)\ \forall\theta$ and $g(\theta^{(s)}|\theta^{(s)}) = f(\theta^{(s)})$, it is straightforward to see that the following inequality is true.

$$f(\theta^{(s+1)}) = g(\theta^{(s+1)}|\ \theta^{(s)}) + f(\theta^{(s+1)})) - g(\theta^{(s+1)}|\ \theta^{(s)}) \geq g(\theta^{(s)}|\ \theta^{(s)}) + f(\theta^{(s)})) - g(\theta^{(s)}|\ \theta^{(s)}) = f(\theta^{(s)})$$

D.1 Initial matching distribution

The M-step updating formula for the initial matching distribution is:

$$m_0\Big(l, k \;\Big|\; g\Big) = \frac{\sum_i p_i\Big(l \mid \theta, g\Big)\mathbb{1}\Big\{k_{j(i,1)} = k\Big\}}{\sum_l \sum_k \sum_i p_i\Big(l \mid \theta, g\Big)\mathbb{1}\Big\{k_{j(i,1)} = k\Big\}}$$

D.2 Wage Distribution Parameters

Wages are assumed to be log-normal with gender-type-class specific mean and variance.

$$\ln f(y_{it}|l, g, k) = -\ln(\sigma_{lgk}) - \ln(\sqrt{2\pi}) - \frac{1}{2}\left(\frac{y_{it} - \mu_{lgk}}{\sigma_{lgk}}\right)^2$$

The wage segment of the expected log-likelihood writes:

$$W = \sum_i \sum_l \sum_k \sum_{t=1}^{T_i} p_i(l|\theta^{(m)}, g)\mathbb{1}\Big\{k_{j(i,t)} = k\Big\}\ln f(y_{it}|l, g, k)$$

Taking derivatives with respect to μ_{lgk} and σ_{lgk} we obtain the M-step updating formulas for the wage parameters:

$$\mu_{lgk}^{(m+1)} = \frac{\sum_i p_i(l|\theta^{(m)}, g) \sum_{t=1}^{T_i} \mathbb{1}\Big\{k_{j(i,t)} = k\Big\}y_{it}}{\sum_i p_i(l|\theta^{(m)}, g) \sum_{t=1}^{T_i} \mathbb{1}\Big\{k_{j(i,t)} = k\Big\}} \tag{D.1}$$

$$\sigma_{lgk}^{(m+1)} = \sqrt{\frac{\sum_i p_i(l|\theta^{(m)}, g) \sum_{t=1}^{T_i} \mathbb{1}\Big\{k_{j(i,t)} = k\Big\}(y_{it} - \mu_{lgk})^2}{\sum_i p_i(l|\theta^{(m)}, g) \sum_{t=1}^{T_i} \mathbb{1}\Big\{k_{j(i,t)} = k\Big\}}} \tag{D.2}$$

D.3 Mobility Parameters

The mobility segment of the expected log-likelihood is:

$$
\begin{aligned}
Q = & \sum_i \sum_l \sum_k \sum_t p_i(l|\theta^{(m)}, g) \mathbb{1}\Big\{k_{j(i,t)} = k, D_{it} = 0\Big\} \ln M_{lgk\neg} + \\
& \sum_i \sum_l \sum_{k,k'} \sum_t p_i(l|\theta^{(m)}, g) \mathbb{1}\Big\{k_{j(i,t)} = k, k_{j(i,t+1)} = k', D_{it} = 1\Big\} \ln M_{lgkk'}
\end{aligned}
$$

Let k be the firm class of the current period, and k' be the firm class of the subsequent period. Recall the parametric specification:

- Unemployment to employment transition probabilities: $M_{lg0k'} = \psi_{lgk'}$ for $k' \geq 1$
- Employment to unemployment transition probabilities: $M_{lgk0} = \delta_{lgk}$ for $k \geq 1$
- Job-to-Job transition probabilities: $M_{lgkk'} = \lambda_{lgk'} P_{lgkk'}$ for $k, k' \geq 1$
  where $P_{lgkk'} = \dfrac{\gamma_{lgk'}}{\gamma_{lgk} + \gamma_{lgk'}}$.

For $k, k' \in \{0, 1, \dots, K\}$, define:

- $n^{(m)}_{lgk\neg} = \sum_i p_i(l|\theta^{(m)}, g) \sum_{t=1}^{T_i} \mathbb{1}\Big\{k_{j(i,t)} = k, D_{it} = 0\Big\}$
- $n^{(m)}_{lgkk'} = \sum_i p_i(l|\theta^{(m)}, g) \sum_{t=1}^{T_i} \mathbb{1}\Big\{k_{j(i,t)} = k, k_{j(i,t+1)} = k', D_{it} = 1\Big\}$

D.3.1 UE transition probabilities

We can obtain the M-step updating formulas for the unemployment to employment transition probabilities by deriving with respect to $\psi_{lgk'}$ the following segment of the expected log-likelihood:

$$
\sum_l n^{(m)}_{lg0\neg} \ln\Big(1 - \sum_{k'=1}^{K} \psi_{lk'}\Big) + \sum_l \sum_{k'=1}^{K} n^{(m)}_{lg0k'} \ln(\psi_{lgk'})
$$

$$
\psi^{(m+1)}_{lgk'} = \frac{n^{(m)}_{lg0k'}}{n^{(m)}_{lg0\neg} + \sum_{k'=1}^{K} n^{(m)}_{lg0k'}} \tag{D.3}
$$

D.3.2 EU and JTJ transition probabilities

The remaining segment of the expected log-likelihood writes:

$$\sum_{l}\sum_{k=1}^{K} n_{lgk\neg}^{(m)} \ln(M_{lgk\neg}) + \sum_{l}\sum_{k=1}^{K}\sum_{k'=0}^{K} n_{lgkk'}^{(m)} \ln(M_{lgkk'})$$

Under the parametric specification provided above, this segment of the expected log-likelihood is not linear in the parameters of interest (specifically, the one related to job-to-job transitions). I therefore consider the minorising function proposed by Lentz et al. (2023).

For $k \in \{1, \ldots, K\}$, we can write:

$$M_{lgk\neg} = 1 - \delta_{lgk} - \sum_{k'=1}^{K} \lambda_{lgk'} P_{lgkk'} = 1 - \delta_{lgk} - \sum_{k'=1}^{K} \lambda_{lgk'} + \sum_{k'=1}^{K} \lambda_{lgk'}(1 - P_{lgkk'})$$

In words, lg-type worker stays in the same firm class k if either she does not receive an offer/layoff or if she receives an offer from k' but prefers to stay in k. In order to build the minorising function we first notice that the following equality holds true.

$$\begin{aligned} M_{lgk\neg} =& \frac{1 - \delta_{lgk}^{(s)} - \sum_{k'=1}^{K} \lambda_{lgk'}^{(s)}}{M_{lgk\neg}^{(s)}} \frac{M_{lgk\neg}^{(s)}}{1 - \delta_{lgk}^{(s)} - \sum_{k'=1}^{K} \lambda_{lgk'}^{(s)}} \left(1 - \delta_{lgk} - \sum_{k'=1}^{K} \lambda_{lgk'}\right) + \\ & \sum_{k'=1}^{K} \frac{\lambda_{lgk'}^{(s)}(1 - P_{lgkk'}^{(s)})}{M_{lgk\neg}^{(s)}} \frac{M_{lgk\neg}^{(s)}}{\lambda_{lgk'}^{(s)}(1 - P_{lgkk'}^{(s)})} \lambda_{lgk'} \left(1 - P_{lgkk'}\right) \end{aligned}$$

Exploiting the concavity of the logarithm, the following inequality holds true.

$$
\ln(M_{lgk\neg}) = \ln\left(1 - \delta_{lgk} - \sum_{k'=1}^{K} \lambda_{lgk'} + \sum_{k'=1}^{K} \lambda_{lgk'}(1 - P_{lgkk'})\right) \geq
$$

$$
\frac{1 - \delta_{lgk}^{(s)} - \sum\limits_{k'=1}^{K} \lambda_{lgk'}^{(s)}}{M_{lgk\neg}^{(s)}} \ln\left(\frac{1 - \delta_{lgk} - \sum\limits_{k'=1}^{K} \lambda_{lgk'}}{1 - \delta_{lgk}^{(s)} - \sum\limits_{k'=1}^{K} \lambda_{lgk'}^{(s)}} M_{lgk\neg}^{(s)}\right) +
$$

$$
\sum_{k'=1}^{K} \frac{\lambda_{lgk'}^{(s)}(1 - P_{lgkk'}^{(s)})}{M_{lgk\neg}^{(s)}} \ln\left(\frac{\lambda_{lgk'}(1 - P_{lgkk'})}{\lambda_{lgk'}^{(s)}(1 - P_{lgkk'}^{(s)})} M_{lgk\neg}^{(s)}\right) \equiv \ln(\underline{\mathrm{M}}_{lgk\neg})
$$

The inequality becomes an equality if $\lambda_{lgk'}^{(s)} = \lambda_{lgk'}$ and $P_{lgkk'}^{(s)} = P_{lgkk'}$: $\ln(\underline{\mathrm{M}}_{lgk\neg})$ minorizes $\ln(M_{lgk\neg})$. We can thus consider $\ln(\underline{\mathrm{M}}_{lgk\neg})$ instead of $\ln(M_{lgk\neg})$ and the MM algorithm maximizes:

$$
H(M|\theta^{(m)}) = \sum_{l} \sum_{k=1}^{K} n_{lgk\neg}^{(m)} \ln(\underline{\mathrm{M}}_{lgk\neg}) + \sum_{l} \sum_{k=1}^{K} \sum_{k'=0}^{K} n_{lgkk'}^{(m)} \ln(M_{lgkk'})
$$

Given $\theta^{(m)}$ obtained at the m-step of the EM algorithm, I update δ, γ, λ by maximising $H(M|\theta^{(m)})$ using an iterative procedure described below. First define:

- $\tilde{n}_{lgkk'}^{(s)} = n_{lgk\neg}^{(m)} \dfrac{\lambda_{lgk'}^{(s)}(1 - P_{lgkk'}^{(s)})}{M_{lk\neg}^{(s)}}$ the predicted number of lg-type stayers that receive an offer from k' but prefer to stay in k.

- $\hat{n}_{lgk}^{(s)} = n_{lgk\neg}^{(m)} \dfrac{1 - \delta_{lgk}^{(s)} - \sum\limits_{k'=1}^{K} \lambda_{lk'}^{(s)}}{M_{lgk\neg}^{(s)}}$ the predicted number of lg-type stayers that stay because they receive no offer/layoff.

We plug $\tilde{n}_{lgkk'}^{(s)}$ and $\hat{n}_{lgk}^{(s)}$ into $H(M|\theta^{(m)})$ and update γ_{lgk} maximising

$$
\sum_{l} \sum_{k} \sum_{k'} \tilde{n}_{lgkk'}^{(s)} \ln \frac{\gamma_{lgk}}{\gamma_{lgk} + \gamma_{lgk'}} + \sum_{l} \sum_{k} \sum_{k'} n_{lgkk'}^{(m)} \ln \frac{\gamma_{lgk'}}{\gamma_{lgk} + \gamma_{lgk'}}
$$

With a simple change of indices and by focusing on a specific worker type:

$$
\sum_{k} \sum_{k'} \tilde{n}_{lgkk'}^{(s)} \ln \frac{\gamma_{lgk}}{\gamma_{lgk} + \gamma_{lgk'}} + \sum_{k} \sum_{k'} n_{lgk'k}^{(m)} \ln \frac{\gamma_{lgk}}{\gamma_{lgk} + \gamma_{lgk'}}
$$

Following Hunter (2004) we note that:

$$-\ln(\gamma_{lgk}+\gamma_{lgk'}) \geq 1-\ln(\gamma_{lgk}^{(s)}+\gamma_{lgk'}^{(s)})-\frac{\gamma_{lgk}+\gamma_{lgk'}}{\gamma_{lgk}^{(s)}+\gamma_{lgk'}^{(s)}}$$

With an additional change of indices and with simple algebra we update γ_{lgk}:

$$\gamma_{lgk}^{(s+1)} = \frac{\sum_{k'=1}^{K}(\tilde{n}_{lgkk'}^{(s)}+n_{lgk'k}^{(m)})}{\sum_{k'=1}^{K}\left(\frac{\tilde{n}_{lgkk'}^{(s)}+n_{lgkk'}^{(m)}+\tilde{n}_{lgk'k}^{(s)}+n_{lgk'k}^{(m)}}{\gamma_{lgk}^{(s)}+\gamma_{lgk'}^{(s)}}\right)} \tag{D.4}$$

The part of the expected log-likelihood to update $\lambda_{lgk'}$ and δ_{lgk} is:

$$\sum_{l}\sum_{k=1}^{K}\hat{n}_{lgk}^{(s)}\ln\Big(1-\delta_{lgk}-\sum_{k'=1}^{K}\lambda_{lgk'}\Big)+\sum_{l}\sum_{k=1}^{K}\sum_{k'=1}^{K}\tilde{n}_{lgkk'}^{(s)}\ln(\lambda_{lgk'})$$
$$+\sum_{l}\sum_{k=1}^{K}n_{lgk0}^{(m)}\ln(\delta_{lgk})+\sum_{l}\sum_{k=1}^{K}\sum_{k'=1}^{K}n_{lgkk'}^{(m)}\ln(\lambda_{lgk'})$$

We update $\lambda_{lgk'}$ and δ_{lgk} as follows:

$$\lambda_{lgk'}^{(s+1)} = \frac{\sum_{k=1}^{K}\left(\tilde{n}_{lgkk'}^{(s)}+n_{lgkk'}^{(m)}\right)}{\sum_{k=1}^{K}n_{lgk0}^{(m)}+\sum_{k=1}^{K}\hat{n}_{lgk}^{(s)}+\sum_{k=1}^{K}\sum_{k'=1}^{K}\tilde{n}_{lgkk'}^{(s)}+\sum_{k=1}^{K}\sum_{k'=1}^{K}n_{lgkk'}^{(m)}} \tag{D.5}$$

$$\delta_{lgk}^{(s+1)} = \frac{n_{lgk0}^{(m)}\Big(1-\sum_{k'=1}^{K}\lambda_{lgk'}^{(s+1)}\Big)}{n_{lgk0}^{(m)}+\hat{n}_{lgk}^{(s)}} \tag{D.6}$$

For given value of $\theta^{(m)}$, the sequence $H(M|\theta^{(m)})$ increases at each iteration step s of the MM algorithm. It is thus not strictly necessary to wait for convergence, the algorithm can be stopped at any time. I iterate the MM algorithm 200 times before it delivers the updated values $\delta^{(m+1)}$, $\gamma^{(m+1)}$, and $\lambda^{(m+1)}$.