Differences in positions along a hierarchy: Counterfactuals based on an assignment model

Laurent Gobillon* Dominique Meurs† Sébastien Roux‡

December 10, 2021§

Abstract

We propose an assignment model in which positions along a hierarchy are attributed to individuals depending on their characteristics. Our theoretical framework can be used to study differences in assignment and pay-offs across groups and we show how it can motivate decomposition and counterfactual exercises. In an application, we study gender disparities in the public and private sectors with a French exhaustive administrative dataset. The gender wage gap in the public sector is 13.3% and it increases by only 0.7 percentage points when workers are assigned to job positions according to the rules of the private sector.

Keywords: assignment, distributions, counterfactuals, wages, gender, public sector

JEL Classification: C51, J31, J45

*Paris School of Economics-CNRS. Address: Paris School of Economics (PSE), 48 Boulevard Jourdan, 75014 Paris, France. Email: laurent.gobillon@psemail.eu. Also affiliated to INED, CEPR and IZA.

†University of Nanterre (EconomiX), Chaire Travail (PSE) and INED. Address: Institut National d'Etudes Démographiques (INED), 133 Boulevard Davout, 75980 Paris Cedex 20, France. Email: dominique.meurs@ined.fr.

‡INSEE, CREST and INED. Centre de Recherche en Economie et Statistique (CREST), 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. Email: sebastien.roux@ensae.fr.

§This paper benefited from a grant provided by DAGFP-Défenseur des droits. Laurent Gobillon acknowledges the support of the EUR grant ANR-17-EURE-0001. Opinions expressed in this paper are solely those of the authors. Access to the data was carried through the CASD dedicated to researchers authorised by the French Comité du secret statistique. We thank Cecilia Garcia-Penalosa, Pauline Givord, Xavier d'Haultfoeuille, Arnaud Maurel, Roland Rathelot, Jean-Marc Robin, Christian Schluter, Antoine Terracol and two anonymous reviewers for useful comments and discussions. All remaining errors are ours.

1 Introduction

Disparities in outcomes across groups of individuals often result from a specific allocation of individuals to positions or entities determining their outcome. This is true on the labor market where gender wage disparities are mostly due to females occupying lower-paid jobs. Other examples include differences in health outcomes across income groups because the rich are able to be admitted in better-quality hospitals, and differences in post-education outcomes because good-quality schools select students with specific attributes.

In this paper, we propose an assignment model such that individuals are assigned to positions along a hierarchy depending on their characteristics and get the corresponding pay-offs. This model can be brought to the data to study differences in pay-offs between groups. In particular, our theoretical framework allows for decompositions and counterfactual exercises when using an alternative assignment rule.

In our model, positions are indexed by their rank in a hierarchy and the pay-off of an individual is derived solely from her assignment to a given position.[1] All individuals consider that the position at the highest rank is the most attractive. Observable characteristics, including the group, influence the chances of being selected for the position. Individuals who are not selected consider the position at the second highest rank, and so on, until all positions are filled. We show that the effects of observable characteristics on the propensity to get positions along the hierarchy can be estimated with a semi-parametric approach. We then propose new decompositions of the difference in pay-offs between groups and new ways of constructing counterfactuals for each group by fixing parameters underlying the propensity to get positions to alternative values.

Our approach is related to the literature on two-sided matching markets (Chiappori and Salanié, 2016) that involves in particular many recent applications to the marriage market (see for instance Choo and Siow, 2006; Galichon and Salanié, 2021). This literature is usually interested in stable matching such that it is impossible for any two entities on the two sides of the market to improve their pay-offs by matching together and for any entity to improve its pay-off by remaining single (Gale and Shapley, 1962). The outcome of our model is also a stable equilibrium. In particular, the surplus derived from a position of given rank cannot be increased by selecting any individual interested in the position, ie. any individual that has not been selected for a position of higher rank.

As our approach is based on a theoretical framework, it provides an alternative to more descriptive methods used to compute counterfactuals that usually involve linear decompositions, quantile decompositions or counterfactual distributions (Oaxaca and Ransom, 1994; Melly, 2005a; Machado and Mata, 2006; Firpo, Fortin and Lemieux, 2009 and 2011; Rothe, 2012; Chernozhukov, Fernandez-Val and Melly, 2013; Maasoumi and Wang, 2019). These approaches use conditional pay-off distributions from other samples as counterfactuals. Equilibrium effects leading to a reassignment of individuals across positions are not modeled. By contrast, in our setting, counterfactuals are

[1] The idea of a common index for all individuals is already present in Fortin and Lemieux (1998) who study the gender wage gap and consider that wages depend on a latent skill index common to males and females. Note however that their index relates to individual characteristics whereas we rather consider the rank in a hierarchy of positions. With this respect, our approach is more in line with Fortin, Bell and Boehm (2017) who study female shares and gender ratios in groups of centiles in the overall income distribution.

the result of equilibria obtained with alternative structural parameters determining the assignment of individuals to positions.

We show that it is possible to estimate a flexible semi-parametric version of the model by maximum likelihood to recover the influence of observable characteristics including the group on the propensity to get positions. Indeed, the probability of a position being filled with a specific individual rather than the other available individuals considering the position is similar to rank-ordered logit models used to recover individual tastes for goods or entities from a rank-ordered list. Applications of such models include the evaluation of tastes for cars from survey data (Beggs, Cardell and Hausman, 1981) and that of preferences for schools from lists of choices (Hastings, Kane and Staiger, 2007; Fack, Grenet and He, 2019). The logic behind our empirical specification is different since it is not based on rank-ordered lists of positions, but rather on the observed assignment of individuals to positions. We thus rather use the information on the identity of the individual allocated to each position.

An additional specificity of our logit specification for the probability of a position being filled by a given individual is that the coefficients of explanatory variables are group-specific polynomial series of the rank. We thus allow the propensity to get selected for the position to depend on observable characteristics and the rank.[2] Counterfactuals of pay-off distributions are generated by changing the coefficients of explanatory variables that capture differences in the rules of assignment to positions. Consistency when the number of individuals tends to infinity is proved by extending results in sampling theory proposed by Rosén (1972). Indeed, we show that it is possible to draw a parallel between the selection of individuals at each rank and the sampling of observations without replacement. We propose a simulation procedure that yields consistent estimators of these counterfactuals when the numbers of individuals and simulations tend to infinity.

We propose an application to gender wage differences in the public and private sectors in France which complements the cross-section literature showing the important role of the segregation of females in lower-paid jobs (Albrecht, Bjorklund and Vroman, 2003; Ponthieux and Meurs, 2015). It is said that females may be treated more fairly in the public sector because recruitment and promotions are based on competition, and labor unions are strong. The public sector is indeed characterized by a smaller wage gap which is consistent with these arguments. However, the wage dispersion is also smaller and may hide an assignment to job positions that is not favorable to females.

For each sector separately, we consider that job positions can be ranked according to the daily wage and we estimate parameters underlying the assignment rules of workers along the job hierarchy that depend on gender and other individual observable characteristics. We then conduct two counterfactual exercises. To quantify the role of gender differences in individual characteristics, we estimate counterfactuals of gender wage distributions in the public sector when considering that workers in that sector are assigned to jobs the same way whatever their gender.

[2] The literature using rank-ordered list of tastes rather consider interactions between the characteristics of individuals and those of goods or entities.

To assess the importance of the assignment rule, we study how gender wage differences change in the public sector when considering that the allocation of workers to positions follows the rules of the private sector. Our work adds to the literature on public-private cross-section differences which has mostly used standard Oaxaca decompositions and gender quantile decompositions (Melly, 2005b; Lucifora and Meurs, 2006; Depalo, Giordano and Papapetrou, 2015).

For our empirical application, we rely on the *DADS Grand Format - EDP* which is a unique administrative dataset recording all jobs in the public and private sectors for all workers born in the first four days of October. Estimations are conducted for the year 2011 on the sample of full-time jobs for workers aged 30-65 to avoid early-career and unstable job positions. We find that the profile of the gender probability ratio of getting a given job position along the wage distribution are rather similar in the public and private sectors, although the gap is slightly larger in the public sector in the $0.5 - 0.85$ rank interval (and slightly smaller above rank 0.85). In each sector, the overall contribution of observables (age, diploma, part-time history, work interruption history, and location in Paris region) to explaining gender differences in propensity to get job positions is small except for ranks below 0.5 in the public sector. Long part-time experience is the only factor that impedes females to get job positions to some extent.

The raw gender average wage gap in the public sector at 13.3% is smaller than that in the private sector which stands at 15.2%. Interestingly, when workers in the public sector are assigned to jobs according to the rules of the private sector, the gender wage gap does not increase much in the public sector since it reaches only 14.0%. This suggests that the gender wage gap difference between the two sectors due to differences in assignment rules would be rather small, around 0.7 percentage points, and the raw difference of 1.9 percentage points would be mostly due to the larger wage dispersion in the private sector. By contrast, the change in gender quantile gap at the last decile when assigning workers in the public sector with the rules of the private sector is large as it stands at 3.6 percentage points.

In Section 2, we detail the mechanisms underlying our approach and show to what extent our framework relates to the literatures on two-sided matching markets and pay-off counterfactuals. In section 3, we make a more formal presentation of our assignment model and explain how counterfactuals can be constructed by changing the parameters determining the propensities of individuals to get positions. We explain in Section 4 how the model can be estimated and empirical counterparts of counterfactuals can be obtained. Section 5 presents our data and discusses the results on the estimated parameters underlying the assignment rules and the counterfactuals. Finally, Section 6 concludes.

2 Mechanisms and literature

Our goal is to propose a way to study the assignment of individuals to positions according to their group, and show how this assignment and group-specific pay-off distributions are modified when changing the assignment rules. For that purpose, we propose an extension of the theoretical model proposed by Gobillon, Meurs and Roux (2015) that incorporates individual observed heterogeneity and can be brought to the data. In this section, we develop the ideas underlying our framework and show how our model fits in the literatures on two-sided matching markets and counterfactual distributions.

2.1 An assignment model with observed heterogeneity

We first present the version of our model in which workers differ only according to their gender j, with $j = f$ for females and $j = m$ for males (see Gobillon, Meurs and Roux, 2015). In our framework, there is a given set of positions which differ according to their pay-off. All the individuals rank the positions in decreasing order of pay-off. Consequently, each position can be characterized by its rank in the distribution of pay-offs, say u. When a gender-j individual is considered for a position of rank u, the potential match is characterized by a surplus that is additive in a term that depends on the gender $\ln \varphi(u|j)$ and a quality $\varepsilon(u)$ that is not observed by the econometrician. The surplus is given by $V(u) = \ln \varphi(u|j) + \varepsilon(u)$. Assignment occurs such that individuals try to maximize their pay-off and manager of each position tries to maximize their profit, which is the surplus net of the pay-off. We consider that the pay-off is the same whatever the individual occupying the position, ie. the pay-off is $w = Q_W(u)$ where $Q_W(\bullet)$ is the quantile function for pay-offs. Under this assumption, the manager actually tries to maximize the surplus. The assignment is such that the highest-ranked position is filled first with the individual maximizing the surplus for that position. The second-ranked position is then filled with one of the remaining individuals, etc. We consider for simplicity that the number of individuals is the same as the number of positions, so that all individuals get a position and all positions are filled. The assignment determines individual pay-offs.

Consider that there is a measure n_j of gender-j individuals such that $n_f + n_m = 1$. Under the assumption that qualities $\varepsilon(u)$ are random terms drawn independently across individuals and positions and following extreme value laws, the probability of rank-u position being filled with a given gender-j individual is given by:

$$\phi(u|j) = \frac{\varphi(u|j)}{n(u|f)\,\varphi(u|f) + n(u|m)\,\varphi(u|m)} \tag{1}$$

where $n(u|j)$ is the measure of gender-j individuals still available for the rank-u position after positions at higher ranks have been filled. On the right-hand side, the numerator is the propensity of the individual to occupy the position and the denominator is the propensity that the position is filled by anyone competing for it. An object of interest is the ratio $\phi(u|f)/\phi(u|m) = \varphi(u|f)/\varphi(u|m)$ that measures the relative propensities of a female and a male occupying the position of rank u. When this ratio is below one, females' access to the position if lower than that of males. Gobillon, Meurs and Roux (2015) show how to estimate this ratio.

In the current paper, we add observed heterogeneity to this framework and we now give more insights on this extension considering a simple case where workers only differ according to their gender and diploma ($D = 1$ for a high-level diploma and $D = 0$ for a low-level one). The probability of rank-u position being filled with a gender-j individual with diploma d is given by:

$$\phi(u\,|d,j) = \frac{\varphi(u\,|d,j)}{\sum\limits_{d\in\{0,1\}} [n(u\,|d,f)\,\varphi(u\,|d,f) + n(u\,|d,m)\,\varphi(u\,|d,m)]} \tag{2}$$

where $n(u\,|d,j)$ is the measure of gender-j individuals with diploma d available at rank u and $\varphi(u\,|d,j)$ is their propensity of occupying the rank-u position. The probability of rank-u position being filled with a gender-j individual then verifies:

$$\phi(u\,|j) = p(u\,|0,j)\,\phi(u\,|0,j) + p(u\,|1,j)\,\phi(u\,|1,j) \tag{3}$$

where $p(u\,|d,j) = n(u\,|d,j)\,/n(u\,|j)$ is the proportion of gender-j individuals with diploma d available at rank u.

As shown by equation (3), the probability of a gender-j worker getting the position at rank u is the weighted average of the probabilities for the two diploma groups of that gender, where the weights are their proportions. Note that these weights are endogenous and need to be computed as solutions of the model. The gender ratio of probabilities verifies:

$$\frac{\phi(u\,|f)}{\phi(u\,|m)} = \frac{p(u\,|0,f)\,\varphi(u\,|0,f) + p(u\,|1,f)\,\varphi(u\,|1,f)}{p(u\,|0,m)\,\varphi(u\,|0,m) + p(u\,|1,m)\,\varphi(u\,|1,m)} \tag{4}$$

Hence, it is equal to the gender ratio of the weighted average propensities to occupy the position. Importantly, studying the gender ratio without conditioning on the diploma level can mask the fact that females have a lower access to positions for a given diploma level and not for the other one. This happens for instance when $\varphi(u\,|0,m) = \varphi(u\,|0,f) = 0.5$, $\varphi(u\,|1,m) = 1$ but $\varphi(u\,|1,f) = \theta$ with $\theta \in]0.5;1[$ for every rank u. In that case, high-diploma females have a better access to any position than low-diploma ones, there is no gender difference in access to positions at every rank for low-diploma individuals since $\varphi(u\,|0,f)\,/\varphi(u\,|0,m) = 1$, but high-diploma females have a lower access than high-diploma males since $\varphi(u\,|1,m)\,/\varphi(u\,|1,m) = \theta < 1$. It is easy to show that the gender ratio of probabilities is then below one at every rank, suggesting that females have overall a lower access to positions than males. In a public policy perspective, disentangling the access by diploma level is necessary to know which diploma group might be targeted to improve their access to positions. Interestingly, equation (4) also shows that values chosen for the propensities to access positions generate a decreasing gender ratio of probabilities due to a filtering process such that high-diploma males occupy positions the fastest, followed by high-diploma females and then low-diploma individuals. When one moves recursively from rank 1 to rank 0, the gender ratio of probabilities tends to one. Indeed, high-diploma individuals are less and less available, and it only remains low-diploma individuals for whom access to positions is the same for females and males. Ignoring the diploma dimension and considering only the overall gender ratio of probabilities thus suggests that there is no gender difference in access to positions at lower rank, but it is more and more difficult for females to access positions as the rank increases. Of course, conditioning on diplomas tells a completely different story since gender differences in access to positions by diploma level is constant across all ranks, with access to positions being the same for low-diploma females and males, but being lower for high-diploma females than for high-diploma males. In the next sections, we will generalize this example to many observable characteristics X (instead of diploma D) and show how female access to positions is affected in counterfactual situations in which the effects of their characteristics on access are modified.

2.2 A two-sided matching framework

We now explain to what extent our model fits in the literature on two-sided matching markets surveyed by Chiappori and Salanié (2016). In our framework, there is a match between a manager and an individual when the individual is assigned to the position handled by the manager. The assignment rule is characterized by the way group and other individual characteristics enter the surpluses of positions. It is possible to characterize more precisely this assignment rule. Consider an individual in group j with observable characteristics X and denote by $\{X_-\}$ the set of observables for all other individuals. Also introduce ε the set of individual matches for every position and by $\{\varepsilon_-\}$ the set of matches for all other individuals. The random variable corresponding to the rank of the individual in the pay-off distribution of positions when using assignment rule r is given by:

$$U_{j,X|r} = f_r\left(X, \{X_-\}, \varepsilon_j, \{\varepsilon_-\}, j\right) \tag{5}$$

where $f_r(\cdot)$ is a deterministic function. Our assignment rule is a generalized version of the one proposed in Gobillon, Meurs and Roux (2015) in which individuals differ only according to their group such that $U_{j,X|r} = f_r\left(\varepsilon, \{\varepsilon_-\}, j\right)$. Denoting by W the pay-off random variable, we have the relationship:

$$W = Q_W\left(U_{j,X|r}\right) \tag{6}$$

where $Q_W(\cdot)$ is the unconditional pay-off quantile function. The pay-off is thus simply a transformation of the rank such that the transformation itself is not influenced by the assignment rule. An implicit assumption is that the pay-off attached to a given position does not depend on individual characteristics.

We now explain to what extent our setting is a standard two-sided matching framework. In corresponding models, there are two sets of individuals, say females and males, since many applications are on the marriage market. Each set is characterized by a given number of types. Each female (resp. male) maximizes utility by choosing a type of male (resp. female), or by remaining single. The objective of related papers is the estimation of match surpluses, and identification strategies usually rely on variations in the number of matches across the two sets of types (Choo and Siow, 2006; Menzel, 2015; Galichon and Salanié, 2021). Our setting is a restricted case such that there is only a single position available for each pay-off value, which yields that the number of matches does not vary across pay-off values and is always equal to one. Hence, variations in the number of matches across pay-offs cannot be used for identification, and in particular to recover how individuals rank positions depending on their characteristics. We therefore need to put more structure on the model. Our key identification assumption is a rank invariance assumption across individuals, ie. the fact that all the individuals rank positions in descending order of pay-off. This restriction gives an order in which positions are filled, and it is then possible to determine which individual occupies each position. Rank invariance also leads to a model structure that can be brought to the data.

A matching equilibrium is considered to be stable in the literature on matching models if no entity would rather be single and no pair of entities would both prefer to be matched together rather than remain in their current situation (Gale and Shapley, 1962). The equilibrium derived from our assignment rule is stable. Indeed, it is of no interest for a manager to fill her position with an individual assigned to a position of lower rank since the surplus resulting from the match would be lower. Also, it is of no interest for an individual occupying a position of higher rank to move to that position since pay-off would be lower. Finally, to keep things simple, we consider that pay-off in the benchmark version of our model is minus infinity when an individual occupies no position. In that case, all individuals want to be matched.

2.3 Counterfactuals

We will be interested in counterfactual gender pay-off distributions when changing the rules of assignment to positions, ie. when modifying the terms $\varphi(u\,|X, j\,)$ that affect the probabilities of the different individuals occupying rank-u position (see equation (3)). A counterfactual of interest is obtained considering that rules of assignment for females are the same as those of males, ie. $\varphi(u\,|X, f\,) = \varphi(u\,|X, m\,)$. The counterfactual assigment of females and males to positions, as well as the counterfactual gender pay-off distributions then differ only due to composition effects (in our simple example, the gender composition in terms of diplomas).

We can compare our setting to the statistical approach proposed by Chernozhukov, Fernandez-Val and Melly (2013), hereafter CFM, to study counterfactual group-specific distributions. CFM is interested in the gender-j counterfactual pay-off cumulative distribution generated from pay-off cumulative distributions conditional on individual characteristics under a counterfactual scenario r:

$$F_{W_{j|r}}(w) = \int_X F_{W_{j,X|r}}(w)\, dF_j(X)$$

where $W_{j|r}$ is the pay-off random variable for a gender-j individual under the counterfactual scenario and $F_{W_{j|r}}(\cdot)$ is its cumulative, $W_{j,X|r}$ is the pay-off random variable when additionally considering the individual characteristics X and $F_{W_{j,X|r}}(\cdot)$ is its cumulative, and $F_j(\cdot)$ is the gender-j cumulative distribution of observable characteristics. A particular counterfactual of interest is obtained when using for females, the conditional cumulative distributions of males. The counterfactual gender gap in pay-off distributions then reflects gender differences in characteristics.

Using the standard result that $U = F_{W_{j,X|r}}(W)$ follows a uniform law $[0,1]$, the statistical representation of counterfactual pay-offs consistent with CFM approach is:

$$W = Q_{W_{j,X|r}}(U) \tag{7}$$

where $Q_{W_{j,X|r}}(\cdot) = F^{-1}_{W_{j,X|r}}(\cdot)$ is the pay-off quantile function conditional on gender and other observable characteristics in the counterfactual scenario. There are several differences with our specification given by equation (6). First, CFM allows pay-offs at a given rank to depend on individual characteristics and counterfactual rules whereas, in our setting, pay-offs are attached to positions and thus independent of individual characteristics at a given rank. By contrast, we consider in our model that it is the ranks (and not pay-offs) that are affected by individual characteristics since those affect the chances of individuals being selected for positions. In CFM, ranks in the overall wage distribution are modified as a result of the counterfactual conditional wage distributions. Second, ranks in our model depend on draws of explanatory variables and matches for all individuals. This occurs because of competition for positions and it yields that ranks are correlated across individuals. Instead, ranks are independent across individuals in CFM. Third, the overall pay-off distribution is fixed in our framework whatever the counterfactual, whereas it depends on counterfactual rules in CFM. Indeed, in their setting, it is a weighted average of counterfactual conditional wage distributions which is given by $\int_{X,j} F_{W_{j,X|r}}(w)\, dF(X,j)$, where $F(X,j)$ is the cumulative distribution of gender and other observable characteristics.

We now explain in more details the differences between our approach and CFM in the way counterfactuals are constructed drawing from simple cases. Consider first the situation in which individuals only differ by their gender (and there is no other observable individual characteristic). In our setting, the couterfactual situation negating gender differences is such that females are given the same propensity of getting any position as that of males. Due to competition of females and males for positions, this modifies the assignment of both females and males to positions (see equation 1) and the two gender distributions of pay-offs, although the overall distribution of pay-offs is that of positions and thus remains unchanged. By contrast, in CFM framework, the counterfactual female pay-off distribution is simply that of males, and the couterfactual male distribution remains the observed one (see equation 7). Here, females simply get pay-offs as if they were males. There is no modelling of assignment to positions and thus no reassignment of males in the counterfactual situation due to females occupying their position. A consequence of that is also a modification of the overall distribution of pay-offs. Rather than being a mixture of the two gender distributions of pay-offs in the initial situation, it becomes the initial male pay-off distribution.

We can then turn to the case where individuals can have either a high-level diploma ($D = 1$) or a low-level one ($D = 0$). Our counterfactual is such that we negate gender differences for each diploma level. This can entail females replacing males for some positions whether they have the same diploma level as these males or a different one. Pay-off distributions conditional on gender and diploma are all affected by the change in the assignment rules. Counterfactual gender pay-off distributions are aggregates of gender distributions conditional on diploma levels. In CFM framework, the counterfactual female pay-off distribution conditional on a diploma level is that of males, and the couterfactual male distribution conditional on a diploma level remains the observed one. This implicitely means that there is no notion of competition for positions between females and males, whether they have the same or different diplomas.

3 Theoretical framework

We now make a formal presentation of our assignment model in the general case. We then show how our model can be used to motivate decompositions of differences in assignment to positions across groups of individuals and counterfactuals of allocations and outcomes when changing the assignment rules.

3.1 The model

In our assignment model, individuals are allocated to positions according to their observable characteristics. There is an infinite but countable number of individuals and we distinguish two groups, say males and females. There is a proportion $n(m)$ of males in the population, which we refer to as the *measure* of males for clarity hereafter, and a proportion $n(f) = 1 - n(m)$ of females. Individuals are characterized by observable attributes X which will affect their chances of getting a position. We focus on the case where attributes take a finite number of values $\{X^k\}_{k=1,\dots,K}$ with K the number of values. Denoting $n(X, j)$ the measure of gender-j individuals with characteristics X and $F_{X,j}(\cdot)$ the cumulative distribution of X for the population of individuals with that gender, we have $\int n(x, j)\, dF_{X,j}(x) = n(j)$. We assume that there exists a bijection between individuals and positions, such that all individuals are allocated to positions and no position is left empty.

Positions are heterogeneous in the pay-off they provide to individuals. For a given position, the pay-off is the same whoever occupies the position and we assume that two positions cannot provide the same pay-off. We also make the rank invariance assumption that every individual ranks positions in decreasing order of pay-off. Individuals are all in competition for the best ranked position whatever their characteristics. They decide whether or not to apply for that position, as conditions for holding the position may be too constraining for some individuals and consequently may deter them from applying. In particular, this may occur for some females who may not be able to work at the fixed hours imposed by the position because they need to take care of the kids. One individual among the applicants is chosen for the position by its manager while taking into account the observable attributes of all applicants. All the individuals not selected for the position turn to the second best ranked position, and so on until all positions are filled.[3]

More formally, the process leading to the choice of an applicant can be described in the following way. For a position of rank u in the distribution of pay-offs, all available individuals deciding to apply are screened. We denote by $n(u|X,j)$ the measure of gender-j individuals with characteristics X who are available for a position of rank u such that $n(1|X,j) = n(X,j)$ as all individuals are available for the first position, and $n(0|X,j) = 0$ as all individuals end up being allocated to a position. Denoting by $v(u|X,j)$ the exogenous share of available gender-j individuals with characteristics X who decide to apply for position u, the measure of applicants with characteristics X is given by $v(u|X,j)\, n(u|X,j)$. The surplus net of pay-off derived from an applicant i, denoted $V_i(u)$ and labelled "Individual value", is supposed to take the form:

$$V_i(u) = \ln \varphi(u|X_i, j(i)) + \varepsilon_i(u) \tag{8}$$

where $\varphi(u|X,j)$ is a fixed component that depends on the rank, the observables and the gender, and $\varepsilon_i(u)$ is a random component that captures the match quality and is drawn independently across individuals. In particular, the fixed component may capture taste discrimination against females that may affect the utility of the manager or a basic form of statistical discrimination such that the manager attributes to individuals the average perceived productivity based on their gender (see Gobillon, Meurs and Roux, 2015, for more details). Note that the strength of these mechanisms is allowed to vary across positions. The random component is observed by the manager but it is not observed by the econometrician. We consider that the applicant chosen for the position is the one with the highest value. The set of applicants to the position is the set of individuals not selected for a position of higher rank and interested in the position. This set can be defined recursively as:

$$\begin{aligned}\Omega(u) = \quad & \{i \text{ applying for rank-}u \text{ position}| \\ & \left.\text{for all } \widetilde{u} > u,\ i \text{ not applying for rank-}\widetilde{u} \text{ position or } V_i(\widetilde{u}) < \max_{k\in\Omega(\widetilde{u})} V_k(\widetilde{u})\right\}\end{aligned}$$

The set of applicants for the position, $\Omega(u)$, contains all the individuals who did not apply for the positions above rank u or did not draw a random match quality high enough to get selected for those positions.

For a given position, the choice of an applicant follows a multinomial specification with two specificities. First, the choice set consists in all applicants still available after better ranked positions have been filled. To avoid any selection of applicants based on the quality of the match with the position because of the filtering process at higher ranks, we assume that match qualities are drawn independently across positions. In particular, this assumption

[3] The framework needs to be extended when several positions provide the same pay-off as available individuals are interested in all the positions to the same extent. In particular, assumptions must be made on whether positions are filled simultaneously or sequentially.

rules out unobserved individual heterogeneity that may affect chances of getting selected at every rank and would then generate a correlation between all match qualities. Still, we will show below with simulations that unobserved individual heterogeneity is unlikely to qualitatively affect the findings in our application. The second specificity of our multinomial specification is that the choice set contains an infinite but countable number of individuals. We extend the extreme value assumption on the law of residuals that is associated with a logit specification to an infinite countable number of positions following Dagsvik (1994).[4] This assumption ensures that for any given position, the probability of selecting an individual follows a logit model.[5] Under this assumption, the probability that the individual chosen for the position of rank u is of gender j and has characteristics X verifies:

$$P\left(j\left(u\right)=j,X\left(u\right)=X\right)=n\left(u\left|X,j\right.\right)\phi\left(u\left|X,j\right.\right) \tag{9}$$

with:

$$\phi\left(u\left|X,j\right.\right)=\frac{\mu\left(u\left|X,j\right.\right)}{\int n\left(u\left|X,f\right.\right)\mu\left(u\left|X,f\right.\right)dX+\int n\left(u\left|X,m\right.\right)\mu\left(u\left|X,m\right.\right)dX} \tag{10}$$

where $\mu\left(u\left|X,j\right.\right)=v\left(u\left|X,j\right.\right)\varphi\left(u\left|X,j\right.\right)$ captures both the propensity to apply and the fixed component entering the value of individuals with characteristics X and gender j. In equations (9) and (10), gender intervenes in exactly the same way as the other individual characteristics. In fact, gender is singled out only because we are going to study gender differences below. The denominator on the right-hand side in equation (10) is a competition term which depends on the measures of applicants with the different genders and characteristics. We will evaluate the overall effects of explanatory variables on the propensity to get a position of rank u for a gender-j individual with characteristics X, $\mu\left(u\left|X,j\right.\right)$, that we label from now on the "conditional individual weight"since it weighs the measures of individuals depending on their chances of getting the position in the competition term. We will refer to $\phi\left(u\left|X,j\right.\right)$ as the conditional probability of getting the position of rank u, and to $\phi\left(u\left|X,m\right.\right)/\phi\left(u\left|X,f\right.\right)=\mu\left(u\left|X,m\right.\right)/\mu\left(u\left|X,f\right.\right)$ as the gender conditional probability ratio of getting the position for individuals with characteristics X.[6]

For gender j, we can derive a differential equation verified by the measure of individuals with characteristics X available for a position of rank u. Consider an arbitrarily small interval du in the unit interval. The proportion of positions in this small interval is du since ranks are equally spaced (and dense) in the unit interval. The measure of positions occupied by individuals of a given gender j with characteristics X is $n\left(u\left|X,j\right.\right)\phi\left(u\left|X,j\right.\right)du$. For these

[4] Put differently, we assume that the points of the sequence $\left\{j\left(i\right),X_i,\varepsilon_i\left(u\right)\right\}$, $i\in\Omega\left(u\right)$ are the points of a Poisson process with intensity measure $\frac{n(u|X,j)\mu(u|X,j)}{\int n(u|X,f)\mu(u|X,f)dX+\int n(u|X,m)\mu(u|X,m)dX}\exp(-\varepsilon)d\varepsilon$.

[5] This modelling is also used by Dupuy and Galichon (2014) in an application of two-sided matching models to the marriage market. They make the assumption that a woman can only match with men she met, their meeting process being random and following a Poisson process. In that case, the probability that a woman of a given type matches with a man of any given type also follows a continuous analog of the logit model.

[6] Note that the probability of getting the position at a given rank $\phi\left(u\left|X,j\right.\right)$ is conditional not only on the specific characteristics of a given applicant, but also on the characteristics of all applicants not hired for a position of higher rank and interested in the position. We do not write explicitly this second conditioning to keep notations simple.

gender and characteristics, the measure of individuals available for a position of rank $u - du$ can be deduced from the measure of individuals available for a position of rank u subtracting the individuals who get the positions of ranks between $u - du$ and u:

$$n\left(u - du\left|X, j\right.\right) = n\left(u\left|X, j\right.\right) - n\left(u\left|X, j\right.\right)\phi\left(u\left|X, j\right.\right)du \tag{11}$$

From this equation, we obtain when $du \to 0$:

$$n'\left(u\left|X, j\right.\right) = \phi\left(u\left|X, j\right.\right)n\left(u\left|X, j\right.\right) \tag{12}$$

This relationship states that the variations in the measure of gender-j individuals with characteristics X around rank u depend on the stock of gender-j individuals with these characteristics and their chances of getting a position. We also show in Appendix A that, under the initial conditions $n\left(1\left|X, j\right.\right) = n\left(X, j\right)$, the system of equations considering (12) for all X and j, where $\phi\left(u\left|X, j\right.\right)$ verifies (10), has a unique solution.

In the data, it is usually the assignment of individuals at every rank which is observed. Equivalently, it means that information is available on quantities $n\left(u\left|X, j\right.\right)$ and their derivatives with respect to the rank. Identification of conditional individual weights up to a multiplicative constant is then granted from the system of equations (10) and (12) considered for every gender and characteristics.

3.2 Decompositions

A matter of interest is the gender difference in allocation to positions which can be measured with the relative propensity of a female and a male getting a position at each rank, ie. the (unconditional) gender probability ratio of getting a position at each rank. Indeed, denote by $\phi\left(u\left|j\right.\right)$ the probability of an available gender-j individual getting a position at rank u. This probability verifies:

$$\phi\left(u\left|j\right.\right) = \int p\left(u\left|X, j\right.\right)\phi\left(u\left|X, j\right.\right)dX \tag{13}$$

with $p\left(u\left|X, j\right.\right) = n\left(u\left|X, j\right.\right)/n\left(u\left|j\right.\right)$ where $n\left(u\left|j\right.\right) = \int n\left(u\left|X, j\right.\right)dX$ is the proportion of gender-j individuals with characteristics X still available for a position at rank u. The (unconditional) gender probability ratio of getting a position of rank u is given by $\phi\left(u\left|f\right.\right)/\phi\left(u\left|m\right.\right)$. Importantly, we make use of the structure of the model to get an expression of this gender probability ratio as a function of conditional individual weights. Indeed, in equation (10), the denominator of the probability of being chosen at a given rank is the same across individuals whatever the characteristics (including gender) and this yields:

$$\frac{\phi\left(u\left|f\right.\right)}{\phi\left(u\left|m\right.\right)} = \frac{\int p\left(u\left|X, f\right.\right)\mu\left(u\left|X, f\right.\right)dX}{\int p\left(u\left|X, m\right.\right)\mu\left(u\left|X, m\right.\right)dX} \tag{14}$$

Suppose that we are able to construct estimators of conditional individual weights at any given rank, $\mu(u|X,j)$. It is possible to make a decomposition of the gender probability ratio of getting a position into the contribution of the gender differences in observable characteristics and the contribution of the gender differences in their returns. We introduce benchmark values for conditional individual weights that correspond to the situation in which there is no gender difference in propensity to get positions. These benchmark values, denoted $\mu^r(u|X)$, are fixed or estimable. For instance, they can be the conditional individual weights for males or for the overall population (in line with Oaxaca and Ransom, 1994). Taking the logarithm of the gender probability ratio (14) and rearranging the terms, we get:

$$\begin{aligned}\log\left[\phi(u|f)/\phi(u|m)\right] &= \int\left[p(u|X,f)-p(u|X,m)\right]\log\mu^r(u|X)\,dX \\ &+\int\left[\log\mu(u|X,f)-\log\mu^r(u|X)\right]p(u|X,f)\,dX \\ &-\int\left[\log\mu(u|X,m)-\log\mu^r(u|X)\right]p(u|X,m)\,dX+r(u) \qquad (15)\end{aligned}$$

where:

$$\begin{aligned}r(u) &= \left[\int p(u|X,f)\log\mu(u|X,f)\,dX-\log\left[\int p(u|X,f)\,\mu(u|X,f)\,dX\right]\right] \\ &-\left[\int p(u|X,m)\log\mu(u|X,m)\,dX-\log\left[\int p(u|X,m)\,\mu(u|X,m)\,dX\right]\right] \qquad (16)\end{aligned}$$

The first right-hand side term in (15) reflects the gender difference in propensity to get the position at a given rank for available individuals if conditional individual weights are the same for males and females, and fixed to the benchmark values. This gender difference is due only to gender differences in the composition of available individuals. The second (resp. third) right-hand side term reflects the gender difference in propensity to get the position if conditional individual weights of available females (resp. males) were modified to take the benchmark values. The fourth one is the residual due to the non-linearity introduced by the use of logarithms. All right-hand side terms can be computed replacing conditional individual weights by their estimators.

Importantly, the set of individuals available at each rank is fixed and determined from the data. Individuals are thus not reassigned to positions when alternatively fixing the conditional individual weights to the benchmark values. We now show how to perform counterfactual exercises that involve a reassignment of individuals when changing conditional individual weights.

3.3 Counterfactuals

A matter of interest is the gender difference in propensity to get positions if individuals were attributed alternative conditional individual weights $\mu^*(u|X,j)$. Denote by $n^*(u|X,j)$ the counterfactual measures of individuals available at a given rank u which are obtained from the differential equation (12) where the conditional probability of getting a position has been replaced by its expression (10) and conditional individual weights by their counterfactuals. These counterfactual measures verify:

$$n^{*\prime}\left(u\left|X,j\right.\right)=\frac{n^{*}\left(u\left|X,j\right.\right)\mu^{*}\left(u\left|X,j\right.\right)}{\int n^{*}\left(u\left|X,f\right.\right)\mu^{*}\left(u\left|X,f\right.\right)dX+\int n^{*}\left(u\left|X,m\right.\right)\mu^{*}\left(u\left|X,m\right.\right)dX} \tag{17}$$

This differential equation is solved under the initial conditions $n^{*}\left(1\left|X,j\right.\right)=n\left(X,j\right)$.

The counterfactual of the gender probability ratio of getting a position can easily be obtained by replacing the conditional probabilities of getting this position by their expressions (13) where conditional individual weights have been replaced by their counterfactuals and the proportion of gender-j individuals with characteristics X still available for a position at rank u by the counterfactual $p^{*}\left(u\left|X,j\right.\right)=n^{*}\left(u\left|X,j\right.\right)/n^{*}\left(u\left|j\right.\right)$ with $n^{*}\left(u\left|j\right.\right)=\int n^{*}\left(u\left|X,j\right.\right)dX$. The counterfactual gender-j probability of getting the position is given by:

$$\phi^{*}\left(u\left|j\right.\right)=\int p^{*}\left(u\left|X,j\right.\right)\phi^{*}\left(u\left|X,j\right.\right)dX \tag{18}$$

In the counterfactual situation, individuals are reallocated across positions according to the alternative assignment rules while holding fixed the pay-off distribution of positions. This yields a change in the gender-specific distributions of pay-offs. The counterfactual of gender-j pay-off cumulative obtained when workers have the alternative individual conditional weights verifies:

$$F_{j}^{*}\left(w\right)=\frac{1}{n\left(j\right)}\int n^{*}\left(F\left(w\right)\left|X,j\right.\right)dX \tag{19}$$

where $F\left(\cdot\right)$ is the pay-off cumulative. The counterfactual gender-j cumulative is simply the proportion of gender-j individuals still available for positions below a given pay-off such that the set of gender-j available individuals was determined according to the alternative assignment rule. Note that the pay-off cumulative is kept the same in the counterfactual situation and changes in counterfactual gender cumulatives are only due to changes in the assignment rules. The derivation of relationship (19) gives the counterfactual of gender-j pay-off density:

$$f_{j}^{*}\left(w\right)=\frac{f\left(w\right)}{n\left(j\right)}\int n^{*\prime}\left(F\left(w\right)\left|X,j\right.\right)dX \tag{20}$$

where $f\left(\cdot\right)$ is the pay-off density. The derivative of the counterfactual measure of available workers can be replaced by its expression given by (17) to get the following expression:[7]

$$f_j^*(w) = \frac{f(w)}{n(j)} \frac{n^*(F(w)|j)\,\phi^*(F(w)|j)}{n^*(F(w)|f)\,\phi^*(F(w)|f) + n^*(F(w)|m)\,\phi^*(F(w)|m)} \tag{21}$$

The counterfactual gender-j pay-off density is proportional to the pay-off density of positions, the proportionality factor being the proportion of gender-j individuals getting positions at the pay-off which is considered.

4 Empirical strategy

4.1 Estimation of parameters

It is possible to quantify the influence of observable characteristics on conditional individual weights under semi-parametric assumptions. We make the assumption that conditional individual weights can be specified as:

$$\mu(u|X,j) = \exp\left[X\beta_j(u)\right] \tag{22}$$

where X now refers to a vector of attributes influencing the propensity to apply and the worker value (which includes the value one), and $\beta_j(\cdot)$ are some gender-specific functions of the rank that we choose to be polynomials of finite order.[8] This model makes an index assumption to decrease the dimensionality but coefficients are allowed to depend on the rank because the propensity to apply or the valuation of characteristics by the manager may depend on the position that is considered. In that setting, the empirical counterpart of the conditional probability of an individual i getting a position at rank u given by (10) is simply a logit model such that the latent variable associated to the individual is $X_i\beta_{j(i)}(u) + \eta_i(u)$ with $\eta_i(u)$ following independent extreme value laws. This latent variable looks like the individual value (8) except that the coefficients $\beta_j(u)$ do not measure the effects of explanatory variables on that value, but rather their joint effects on that value and the propensity to apply.[9]

The parameters of polynomial coefficients can be estimated by maximum likelihood. We first introduce some additional notations. Denote by u_i the rank of individual i and X_i the value of her observable attributes, $u^k = k/N$ the k^{th} rank, i_k the individual occupying the position at this rank, $\overrightarrow{X}_k = \left(X'_{i_1}, ..., X'_{i_k}\right)'$ and $\overrightarrow{j}_k = (j(i_1), ..., j(i_k))'$

[7] Indeed, substituting for $n^{*\prime}(F(w)|X,j)$ in (20) using (17) gives:

$$f_j^*(w) = \frac{f(w)}{n(j)} \frac{\int n^*(F(w)|X,j)\,\mu^*(F(w)|X,j)\,dX}{\int n^*(F(w)|X,f)\,\mu^*(F(w)|X,f)\,dX + \int n^*(F(w)|X,m)\,\mu^*(F(w)|X,m)\,dX}$$

From (18), we have that $\int n^*(F(w)|X,j)\,\mu^*(F(w)|X,j)\,dX = n^*(F(w)|j)\,\phi^*(F(w)|j)$, which can be inserted at the numerator and denominator in the previous expression of $f_j^*(w)$ and it gives equation (20).

[8] The definition of X is thus modified from being a value taken by the whole set of individual characteristics in the theoretical section, to the vector of values taken by every individual characteristics in the empirical part. We do not change the notation for simplicity as both refer to values taken by individual characteristics.

[9] Note that the residual $\eta_i(u)$ is not always equal to the match quality $\varepsilon_i(u)$ because of the heterogeneity in the propensities to apply for positions. We have $\eta_i(u) = \varepsilon_i(u)$ only when the propensities to apply are similar for all individuals.

respectively the observed characteristics and genders of the individuals occupying the k lowest ranked positions, and $\Omega\left(u^k\right)$ the set of workers available at rank u_k. The likelihood is given by:

$$
\begin{aligned}
L &= P\left(u_{i_1}=u^1, u_{i_2}=u^2, ..., u_{i_N}=u^N \middle| \overrightarrow{X}_N, \overrightarrow{j}_N\right) && (23)\\
&= P\left(u_{i_N}=u^N \middle| \overrightarrow{X}_N, \overrightarrow{j}_N\right)\prod_{k=1}^{N-1} P\left(u_{i_k}=u^k \middle| u_{i_{k+1}}=u^{k+1}, ..., u_{i_N}=u^N, \overrightarrow{X}_N, \overrightarrow{j}_N\right) && (24)\\
&= \prod_{k=1}^{N} P\left(u_{i_k}=u^k \middle| \{i_1, ..., i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right) && (25)
\end{aligned}
$$

where the last equality is obtained using the fact that random match qualities $\eta_i(u)$ are independently and identically distributed across ranks. Indeed, in that case, what matters for the selection of an individual at a given rank u_k is the set of available individuals in competition for the position and not the exact identity of individuals chosen at every higher rank since draws of random match qualities that determine their identity are not related to draws for the position at rank u_k. In equation (25), $P\left(u_{i_k}=u^k \middle| \{i_1, ..., i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right)$ is the empirical counterpart of $\phi\left(u_{i_k}\left|X_{i_k}, j\left(i_k\right)\right.\right)$ and it verifies:

$$
P\left(u_{i_k}=u^k \middle| \{i_1, ..., i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right) = \frac{\mu\left(u_{i_k}\left|X_{i_k}, j\left(i_k\right)\right.\right)}{\sum\limits_{\ell\leqslant k}\mu\left(u_{i_k}\left|X_{i_\ell}, j\left(i_\ell\right)\right.\right)} = \frac{\exp\left[X_{i_k}\beta_{j(i_k)}\left(u_{i_k}\right)\right]}{\sum\limits_{\ell\leqslant k}\exp\left[X_{i_\ell}\beta_{j(i_\ell)}\left(u_{i_k}\right)\right]} \qquad (26)
$$

The parameters of polynomial coefficients $\beta_j(u)$ are estimated by maximizing the logarithm of the likelihood $L=\frac{1}{N}\sum\limits_i P\left(u_{i_k}=u^k \middle| \{i_1, ..., i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right)$. In fact, the likelihood is the same as the partial likelihood obtained when estimating a Cox duration model with time-varying parameters.[10] Indeed, since $\beta_j(u)$ are polynomials, $X\beta_j(u)$ can be rewritten as interactions between X and terms u^p with $p\in\{1, ..., P\}$ with these interactions having constant coefficients (ie. coefficients that do not depend on rank). The asymptotic normality of estimated parameters for time-varying explanatory variables with constant constant coefficients has already been established in the duration model literature (see Andersen and Gill, 1982; Chernozhukov, Fernandez-Val and Melly, 2013).

[10] Cox duration models have already been used to model wage distributions (Donald, Green and Paarsch, 2000; Chernozhukov, Fernández-Val and Melly, 2013). However, unlike standard duration models, our approach considers an instantaneous probability of occupying a position that depends on a competition term (see equation 10) involving the characteristics and numbers of other individuals available for the position. In technical terms, our hasard rate at empirical rank u^k is of the form $\lambda(u^k, \overrightarrow{n}\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k)\mu\left(u^k\left|X_k, j_k\right.\right)$ where $\overrightarrow{n}\left(u^k\right)$ is a vector containing the terms $n\left(u^k\left|X, j\right.\right)$ for all X, j (and other terms are already defined in the text), rather than $\lambda(u^k)\mu\left(u^k\left|X_k, j_k\right.\right)$. Still, our likelihood ends up being similar to the partial likelihood of a Cox model. Other differences with the above references is that our model is based on ranks in the wage distribution rather than wages themselves, and the timeline goes from rank 1 downward, rather than from wage 0 upward.

4.2 Evaluation of decompositions and counterfactuals

It is also possible to make an empirical assessment of the decomposition of the gender probability ratio of getting each position given by (15). Assume that benchmark conditional individual weights $\mu^r(u|X)$ are also of the form (22) with estimable polynomial coefficients $\beta^r(u)$. In that case, the decomposition simplifies to:

$$\begin{aligned}\log\left[\phi(u|f)/\phi(u|m)\right] &= \left[E(X|f,u)-E(X|m,u)\right]\beta^r(u) \\ &\quad +E(X|f,u)\left[\beta_f(u)-\beta^r(u)\right]-E(X|m,u)\left[\beta_m(u)-\beta^r(u)\right] \\ &\quad +r(u) \qquad (27)\end{aligned}$$

where $E(X|j,u)$ are the average characteristics of gender-j individuals available for a position at rank u.[11] This expression is similar to the standard Oaxaca-Ransom decomposition: the gender probability ratio of getting a position of rank u depends on a linear composition effect related to the gender differences in average characteristics and two linear effects due to differences in returns to characteristics between each gender and the benchmark, as well as a residual due to non-linearities. Importantly, the decomposition is such that the gender probability ratio is affected only by gender-specific means of individual characteristics but not higher-order features of gender-specific distributions. In particular, suppose that the means are the same for the two genders, but females have a larger dispersion than males for a given individual characteristic. Females with extreme values for that characteristic may have extreme values for their probabilities of occupying position at some ranks, but these effects do not show up in the decomposition.

The gender probability ratio of getting positions on the left-hand side of equation (27) can be estimated non parametrically following Gobillon, Meurs and Roux (2015). Right-hand side terms can be obtained by replacing average characteristics with their empirical counterparts, and polynomial coefficients by their estimators. If the values used as a benchmark for the coefficients of explanatory variables are those of males, $\beta^r(u)\equiv\beta_m(u)$ which have already been estimated. If the values used as a benchmark are those of the overall population, $\beta^r(u)$ can be obtained by maximum likelihood fixing $\beta_f(u)=\beta_m(u)\equiv\beta^r(u)$ and adding a gender dummy to the specification to act as a control in line with Fortin (2008).

[11] The proof is the following. Using the expressions $p(u|X,j)=n(u|X,j)/n(u|j)$ and $\log\mu^r(u|X)=X\beta^r(u)$, we get that the explained part verifies:

$$\begin{aligned}\int\left[p(u|X,f)-p(u|X,m)\right]\log\mu^r(u|X)\,dX &= \left[\int n(u|X,f)\,XdX/n(u|f)-\int n(u|X,m)\,XdX/n(u|m)\right]\beta^r(u) \\ &= \left[\int_{i|j(i)=f,u_i\le u}X_i di/n(u|f)-\int_{i|j(i)=m,u_i\le u}X_i di/n(u|m)\right]\beta^r(u) \\ &= \left[E(X|f,u)-E(X|m,u)\right]\beta^r(u)\end{aligned}$$

which is the expression reported in the text. The expressions for the other right-hand side terms of (27) corresponding to the unexplained part can be established in the same way.

We now turn to the evaluation of the counterfactual gender-j probabilities of getting every position given by (18). For that purpose, we need to recover $n^*\left(u\,|X,j\right)$ when using the counterfactual conditional individual weights $\mu^*\left(u\,|X,j\right)$ which we suppose to be of the form (22) with polynomial coefficients $\beta_j^*\left(u\right)$. A direct method would consist in solving the system of non-linear differential equations given by (17). However, this is untractable in practice because the number of equations is equal to the number of values taken by the set of characteristics, which is very large. Consequently, we rather rely on a simulation approach.

First note that the finite discrete counterpart of the differential equation verified by the measures of available individuals (12) is:

$$N^*\left(u^k\,|X,j\right) = N^*\left(u^{k+1}\,|X,j\right) - D_{k+1}\left(X,j\right) \tag{28}$$

where $N^*\left(u\,|X,j\right)$ is the counterfactual number of gender-j individuals with characteristics X available at rank u, and $D_k\left(X,j\right)$ is a dummy taking the value one if an available individual in the set $\Omega_j^*\left(u^k,X\right)$ gets the position at rank u^k and zero otherwise. There is some randomness which comes from the choice of an individual from conditional probabilities of getting positions of the form (10). For a given rank v, a quantity of interest is $E\left[N^*\left(u^{\lfloor vN\rfloor+1}\,|X,j\right)\right]$ where $\lfloor\cdot\rfloor$ is the integer part. Indeed, we show in Online Appendix that $E\left[N^*\left(u^{\lfloor vN\rfloor+1}\,|X,j\right)\right]/N \underset{a.s.}{\rightarrow} n^*\left(v\,|X,j\right)$ for all $v\in\left]0,1\right[$ when $N\rightarrow+\infty$. The proof relies on the extension of a theorem on sampling without replacement proposed by Rosén (1972). Whereas in the original theorem, the influence of explanatory variables on the propensity of an individual to get the position does not vary across positions, in our case this propensity varies since $\exp\left[X\beta_j\left(u\right)\right]$ depends on the rank. We show that the proof of the original theorem can be generalized to the case where the influence of explanatory variables varies across positions.

The expectation $E\left[N^*\left(u^{\lfloor vN\rfloor+1}\,|X,j\right)/N\right]$ can be estimated using a simulation procedure that involves reassignments of individuals of the original sample to the positions based on the counterfactual assignment rules. Theoretical foundations of the simulation approach are detailed in Appendix B. A simulation is indexed by $s=1,...,S$ and we denote by $\Omega_j^s\left(u,X\right)$ the counterfactual sample of gender-j individuals with characteristics X available at rank u in simulation s. In practice, the counterfactual sets of individuals with given characteristics X available at the empirical rank u^{k-1}, $\left\{\Omega_j^s\left(u^{k-1},X\right)\right\}_{X,j}$, are deduced from the same sets at next empirical rank, $\left\{\Omega_j^s\left(u^k,X\right)\right\}_{X,j}$, by subtracting the individual who gets the position at rank u^k which is determined consistently with the model specification in the following way.

The empirical counterpart of the counterfactual conditional probability of getting a position at rank u^k can be rewritten as:

$$P\left(u_i = u^k\,\middle|\,\Omega^s\left(u^k\right),\overrightarrow{X}_k^s,\overrightarrow{j}_k^s\right) = \frac{\exp\left[X_i\beta_{j(i)}^*\left(u^k\right)\right]}{\sum_{k\in\Omega^s\left(u^k\right)}\exp\left[X_k\beta_{j(k)}^*\left(u^k\right)\right]} \tag{29}$$

with $j\left(i\right)$ the gender of individual i, $\Omega^s\left(u^k\right)$ the counterfactual sample of individuals available at rank u^k in simulation s, and $\left(\overrightarrow{X}_k^s,\overrightarrow{j}_k^s\right)$ their characteristics and genders. It thus corresponds to the probability of getting the position at rank u^k given by a multinomial logit model. Hence, it is possible to simulate which individual gets that position by computing, for each available individual, the sum $X_i\widehat{\beta}^*_{j(i)}\left(u^k\right)+\eta_i^s\left(u^k\right)$ where $\eta_i^s\left(u^k\right)$ has been drawn in an extreme value law. The individual getting the position is the one with the highest value of this sum. The sets of available individuals with given characteristics and gender at each empirical rank is obtained by recursively applying this procedure from the highest to the lowest rank. Finally, we obtain the counterfactual number of available individuals with given characteristics and gender for a simulation $N^s\left(u^k\,|X,j\right)=Card\ \Omega_j^s\left(u^k,X\right)$ and an estimator of the expected counterfactual number of available individuals with given characteristics and gender is $\widehat{N}^*\left(u^k\,|X,j\right)=\sum_{s=1}^S N^s\left(u^k\,|X,j\right)/S$. We show in Appendix B that when $S\to+\infty$, we have for all $v\in(0,1)$, $\widehat{N}^*\left(u^{\lfloor vN\rfloor+1}\,|X,j\right)\underset{a.s.}{\to}E\left[N^*\left(v\,|X,j\right)\right]$.

We now explain how to evaluate the probability of getting a position at each rank u^k for each gender. The empirical counterparts of the terms $p^*\left(u^k\,|X,j\right)$ which enter the counterfactual probability (18) are:

$$\widehat{p}^*\left(u^k\,|X,j\right)=\frac{\widehat{N}^*\left(u^k\,|X,j\right)}{\widehat{N}^*\left(u^k\,|j\right)} \tag{30}$$

where $\widehat{N}^*\left(u\,|j\right)=\sum_\ell\widehat{N}^*\left(u\left|X^\ell,j\right.\right)$ is an estimator of the expected counterfactual number of gender-j individuals available at rank u. An estimator of the counterfactual gender-j probability of getting a position at rank u^k is then given by:

$$\widehat{\phi}^*\left(u^k\,|j\right)=\sum\nolimits_\ell\widehat{p}^*\left(u^k\left|X^\ell,j\right.\right)\exp\left[X^\ell\widehat{\beta}^*_j\left(u^k\right)\right] \tag{31}$$

where $\widehat{\beta}^*_j\left(u\right)$ is an estimator of $\beta^*_j\left(u\right)$.

We can also recover the counterfactuals of gender-j cumulative and density of pay-offs. We first consider the original sample and sort positions in ascending order according to pay-offs, denoting by w^k the k^{th} pay-off. It is possible to construct an estimator of the counterfactual of gender-j cumulative at rank u^k, $F_j^*\left(w^k\right)$, from equation (19) as:

$$\widehat{F}_j^*\left(w^k\right)=\frac{\widehat{N}^*\left(\widehat{F}\left(w^k\right)|j\right)}{N\left(j\right)} \tag{32}$$

where $N(j)$ is the number of gender-j individuals and $\widehat{F}\left(w^k\right)$ is an estimator of the pay-off cumulative of positions computed at pay-off w^k. A counterfactual of gender-j density is obtained from equation (21) replacing right-hand terms by estimators:

$$\widehat{f}_j^*\left(w^k\right)=\frac{\widehat{f}\left(w^k\right)}{N\left(j\right)/N}\frac{\widehat{N}^*\left(F\left(w^k\right)|j\right)\widehat{\phi}^*\left(F\left(w^k\right)|j\right)}{\widehat{N}^*\left(F\left(w^k\right)|f\right)\widehat{\phi}^*\left(F\left(w^k\right)|f\right)+\widehat{N}^*\left(F\left(w^k\right)|m\right)\widehat{\phi}^*\left(F\left(w^k\right)|m\right)} \tag{33}$$

where $\widehat{f}\left(w^k\right)$ is an estimator of the pay-off density computed at pay-off w^k on the whole population.

5 Application

We use our assignment model to study the gender wage gap in the public and private sectors in France. Institutional details are relegated in Appendix C. Evidence shows a smaller gap in the public sector which is usually attributed to a fairer treatment of females. However, the wage dispersion is lower in the public sector (Melly, 2005b; Lucifora and Meurs, 2006; Depalo, Giordano and Papapetrou, 2015), and differences in gender wage gaps between the two sectors could simply reflect a difference in wage dispersion. In our application, we assess whether there are gender differences in the assignment of workers to well-paid jobs.

In line with our theoretical framework, we consider that the allocation of workers to job positions results from workers applying and being selected for these positions. We make the assumption that a fixed wage is associated to each job position through a contract. This wage is supposed not to depend on the gender or the other observable characteristics of applicants. Job positions are ranked along the wage hierarchy and workers are interested in positions yielding the highest wages. Individuals are heterogeneous in their labour supply and may not apply to every position because work conditions may be too constraining.

5.1 Data and stylized facts

Estimations are conducted on the *DADS Grand Format - EDP 2011* which is a panel dataset following all individuals born in the first four days of October and is constructed from two different sources (*Déclaration Annuelles des Données Sociales* i.e. *DADS* and *Echantillon Démographique Permanent*, i.e. *EDP*). The data record all their jobs in the public and private sectors since 1992. Jobs in the public sector can belong to three subsectors: central administration (including education), local government and public health.

The DADS are collected for tax purposes and contain details on job characteristics. They give the establishment identifier (*SIREN* number) from which we determine firm seniority since 1992 and the status (full-time or part-time) from which we reconstitute part-time history. In the public sector, firm tenure corresponds to the number of years spent in the subsector of the position that is currently occupied. We also compute the number of years individuals are absent from the data. Absence corresponds to an interruption in the salaried activity due to unemployment, exit from the labour force or self-employment. The full-time equivalent annual wage is reported. As we estimate a cross-section model, our analysis focuses on year 2011.

There are outliers with wage below the minimum wage and consequently we delete observations for which the monthly wage is below 1000 euros. The job duration during the year (in days) is reported and we only retain jobs occupied full time on July 1st in which workers stayed for at least 30 days during the year to focus on stable workers that are more likely to compete for all job positions. This also means that we keep at most one job per worker every year. Finally, we use the information on administration for public jobs to restrict the sample to those in the central and local administration.[12] We consider jobs only for workers aged 30-65 to avoid taking into account the frequent transitions between unstable job positions that often occur at the beginning of the career. Our final dataset contains 55, 881 observations and the proportion of females is 37.8%. Most individuals work in the private sector (82.6%) where the proportion of females (35.1%) is lower than in the public sector (50.9%).

Data are also used to construct our set of explanatory variables. We consider a dummy for firm tenure being larger or equal to 10 years and two dummies for part-time experience being respectively between 7% (the median) and 18% (the third quartile), and more than 18% (less than 7% being the reference). The location of job at the municipality level is used to construct a dummy for the job being located in the Paris region. We also consider two dummies for the age brackets 41-50 and 50-65 (31-40 being the reference). These variables are complemented with information on diploma and we construct three dummies corresponding to having a high-school diploma, spending two years or less in college, and spending more than two years in college. Finally, the number of children is taken into account with two dummies for having respectively no child, and three children and more (one or two children being the reference). Note that we are rather parsimonious in the number of categories. This is because we need to estimate the gender-specific coefficients of a polynomial function of ranks for each dummy in the empirical application and this makes the number of coefficients increase fast with the number of explanatory variables.

We then propose stylized facts on wages and individual characteristics for the two genders in the two sectors. Figure 1 shows that gender log-wage distributions in the private sector have fatter right-hand tails and lower peaks than in the public sector. In the two sectors, male distribution is slightly to the right of female distribution, especially in the public sector. We report descriptive statistics on wages by sector and gender in Table 1. They confirm that the public sector is characterized by a lower wage dispersion than the private sector. The average gender wage gap in the public sector is smaller (14% vs. 19%), and the gender quantile difference increases with the rank but more slowly than in the private sector.

[*Insert Figure* 1]

[*Insert Table* 1]

Turning to observable characteristics, Table 2 shows that in each sector females are more qualified than males, have less often three children or more, and have much more often long part-time experience and long work interruption. The gender gap in part-time experience is very large and is similar across sectors. The gender gaps in education and work interruption are smaller in size but remain sizable, and they are larger in the public sector. By contrast,

[12] We exclude teaching jobs because their management is very specific. We also exclude jobs in the health administration because some positions such as doctors are very specific and workers occupying them usually do not change job for a position in another administration of the public sector. These sample restrictions decrease the average wage in the public sector since excluded positions are paid above the average.

the gender gap in having three children, which also takes sizable values, is larger in the private sector.

[*Insert Table* 2]

In line with the literature, we then assess to what extent the gender quantile difference varies with the rank in the public and private sectors once observable characteristics have been taken into account. For that purpose, we run quantile regressions including a female dummy as well as the other observable characteristics which are used as controls. Figure 2 represents the estimated coefficient of the female dummy as a function of rank. It shows that in both sectors, there is a gender quantile gap at all ranks which increases with the rank. Whereas the gender quantile gap is similar in the two sectors at the lowest ranks, it is larger in the private sector above rank 0.05. The difference in gender quantile gap between the two sectors is rather stable above rank 0.2 at around 4 percentage points.

[*Insert Figure* 2]

We finally compute a non-parametric estimator of the gender probability ratio of getting each job for each sector following the procedure proposed by Gobillon, Meurs and Roux (2015). Figure 3 shows that above rank 0.05, females have a lower propensity than males to get any job position whatever the sector. In the private sector, the female propensity to get job positions decreases slowly until rank 0.45 and then increases before decreasing again after rank 0.65. Non-monotonic movements can be explained by the heterogeneity of job positions as very heterogeneous industries are pooled. At the highest ranks, female propensity to get job positions is very low: a female has 70% less chances of getting a job position than a male. In the public sector, the gender probability ratio of getting job positions decreases until rank 0.4, is nearly flat for ranks in the $0.4 - 0.9$ interval, and then decreases again. Interestingly, between ranks 0.5 and 0.85, female propensity to get job positions is lower in the public sector than in the private one.

[*Insert Figure* 3]

5.2 Estimation of the gender probability ratio of getting a job

We now turn to the estimation of the semi-parametric version of the model. We estimate specification (22) that involves category dummies for all our explanatory variables, including the female dummy (but not its interactions with other variables), and we fix the degree of polynomial coefficients to five for each dummy.[13] As we will see,

[13] Here and below, we consider only a limited number of categories for each variable for tractability. Still, we made robustness checks varying the number of categories for each variable successively. When increasing the number of categories for age to 6 (30-34, 35-39, 40-44, 45-49, 50-54, 55+), for diploma to 5 (secondary school, vocational training, high-school, college⩽2 years, college >2 years), for number of children to 4 (0, 1, 2, 3+), for firm tenure to 4 (<1 year, 1-3, 3-10, >10) and for part-time experience to 5 (0%, 0-7%, 7-18%,

this degree is enough to get a good fit of the specification with the data. Estimated coefficients (excluding that of the gender dummy) will be used as references in a counterfactual exercise in which conditional individual weights of the two genders are equalized. For this simple specification, the exponentiated effects of category dummies for a given variable capture the relative chances of getting a job position compared to the reference category. These exponentiated effects are represented as a function of rank in Figure 4 and we compare their values between the two sectors.

Variations across ranks of the estimated coefficients of the female dummy are consistent with the non-parametric estimators of the gender probability ratio of getting each job position, suggesting a lesser role of gender differences in observable characteristics. Interestingly, the estimated coefficients of the female dummy in the two sectors are similar for ranks in the 0.05-0.7 interval, and this contrasts with the gender differences obtained with quantile regressions (see Figure 2). The effect of every diploma is positive and increases with the rank, especially in the public sector. The higher the diploma, the larger the increase. The slope is particularly steep when spending more than two years in college, especially at the highest ranks. Not surprisingly, workers with no high-school diploma have nearly no chance of getting the best-paid job positions in the two sectors. Workers with three children and more have a higher probability of getting best-paid job positions in the two sectors, especially the private one.[14] Age profiles in the two sectors are consistent with larger chances of getting best-paid job positions when being older. The propensity to get theses job positions is also larger when living in the Paris region, consistently with a large concentration of high-paid job positions in that region. Short part-time experience is mostly detrimental in the public sector. This could be due to labor supply effects such that part-time workers in that sector are less career-oriented, or to career rules as part-time entry job positions impede sectoral seniority which is important for promotions in the public sector. By contrast, long part-time experience is detrimental in both sectors. Finally, the picture is similar when considering short and long work interruptions.

[*Insert Figure* 4]

We also consider a more complete semi-parametrization of conditional individual weights in which we additionally include the interactions between the female dummy and all the category dummies corresponding to observable characteristics. The degrees of all polynomial coefficients is again fixed to five. Figure 5 shows that, for each sector, the semi-parametric gender probability ratio of getting a job position at any given rank obtained using the procedure described in Section 3.2 is in the confidence interval of the non-parametric gender probability ratio obtained using 100 bootstrap replications. In fact, the curves obtained with the non-parametric and semi-parametric

18-44%, >44%), results are virtually unchanged.

[14] It is possible to check that this occurs mostly because of males. It is consistent with males with three children being more stable or being ready to work more when they are the main providers of the household. Firm tenure is associated with larger chances of getting positions, except at the highest ranks in the private sector. This can be explained by firm mobility to access some of the best-paid positions in the private sector.

approaches are nearly confounded, which suggests that the semi-parametric approach is reliable. For each observable characteristic other than gender, Figure 6 represents the exponentiated gender difference in the estimated coefficient of each category dummy. It corresponds to the conditional gender probability ratio of getting a given job position for the category while fixing other observable characteristics to their reference values. Except for a very few exceptions, exponentiated gender differences in estimated coefficients are well below one and take lower values at higher ranks. This suggests that females have a lower propensity to get job positions than males whatever the observable characteristics, and the gender difference is larger for high-paid job positions than for low-paid ones.

[*Insert Figures* 5 *and* 6]

5.3 Decompositions and counterfactuals

We then implement for each sector the Oaxaca decomposition of the gender probability ratio of getting any given job position using equation (27). Figure 7 shows that the explained part of this gender probability ratio is small at nearly all ranks in the private sector. By contrast, the explained part is rather large for ranks below 0.5 in the public sector before becoming small at higher ranks. Additional results in Appendix D.1 show that the explained part is mostly due to long part-time experience.

[*Insert Figure* 7]

We then compute the counterfactual gender probability ratios of getting any given job position in the two sectors when the two genders are given the same conditional individual weights which are fixed to their common reference. They capture gender differences in chances of getting job positions that are only related to gender differences in observable characteristics. A major difference between this approach and the Oaxaca decomposition is that workers are now reassigned to job positions. Interestingly, results represented in Figure 8 show that there are still gender differences in the counterfactual situation. They are due to the large gender differences in part-time experience. In the public sector, female propensity to get job positions is around 18% lower than that of males for ranks below 0.5 and then increases to reach equality around ranks $0.8 - 0.9$ before decreasing again to end up being lower by around 20% at the highest ranks. These variations across ranks differ from those computed for the explained part of the Oaxaca decomposition, which suggests a significant role of the equilibrium effects related to a reallocation of workers across job positions. Descriptive statistics in Table 3 and log-wage distributions in Figure 10 show that the counterfactual gender wage gap does not disappear as the gender average wage gap stands at $100 * [1 - \exp(-0.0258)] \approx 2.5\%$ and is significantly different from zero. The shape of the counterfactual gender probability ratio is similar in the private sector but it takes values closer to one. Indeed, the counterfactual female propensity to get job positions is only around 4% lower than that of males below rank 0.6 before increasing and reaching values above it at ranks in the $0.6 - 0.95$ interval. It then decreases again to end up being 20% lower than that of males at the highest ranks. The counterfactual gender average wage gap is now closer to zero at 0.2% and is not significant.

[*Insert Figure* 8]

[*Insert Table* 3]

We then consider the counterfactual situation in which workers in the public sector are allocated to job positions according to the assignment rules of the private sector. Figure 9 shows that the counterfactual gender probability ratio of getting a given job position in the public sector is larger than the initial ratio for ranks in the $0.5 - 0.85$ interval but lower for ranks above 0.85. In fact, it has a shape close to the one observed in the private sector. The counterfactual gender average wage gap in the public sector reaches $100 * [1 - \exp(-0.1514)] = 14.0\%$ and is only slightly higher than the raw wage gap which is 13.3%. It contrasts with the raw wage gap in the private sector which stands at 15.2%. Differences in assignment rules between the two sectors thus only explain 0.7 percentage points of the gender wage gap difference which stands to 1.9 percentage points. The rest of the gender wage gap difference between the two sectors can be explained by gender differences in observable characteristics and the larger wage dispersion in the private sector. Interestingly, there are also changes in the gender quantile gap when using the counterfactual assignment rule and they differ across ranks. Whereas the gender median wage gap in the public sector is 0.8 percentage points lower in the counterfactual situation, the gender wage gap at the last quartile and the last decile are 0.7 and 3.6 percentage points higher respectively. The gender gap in log-wage dispersion is also higher as the gender difference in standard deviation increases by 72% (ie. by 2.0 percentage points).

Conversely, the counterfactual gender probability ratio of getting a given job position in the private sector is lower than the initial ratio for ranks in the $0.5 - 0.85$ interval but higher for ranks above 0.85. Overall, its profile across ranks is similar to that of the public sector. The counterfactual gender average wage gap at 14.4% is lower than the original gap by 0.8 percentage points. There are also changes in the gender quantile gap that differ across ranks when changing the assignment rule, but variations are not exactly the opposite of those in the public sector. Indeed, the gender median wage gap increases by 1.3 percentage points, whereas the gender wage gaps at the last quartile and decile decrease by respectively 0.05 and 5.1 percentage points. The gender difference in standard deviation decreases by 27% (i.e. by 1.8 percentage points).

[*Insert Figures* 9 *and* 10]

Finally, we conduct robustness checks. When considering the sample that includes both full-time and part-time workers, and the hourly wage instead of the daily wage, results are similar as shown by Appendix D.2. We also assess whether ignored unobserved individual heterogeneity might bias the estimates and Appendix D.3 suggests that it is not the case as long as this heterogeneity is not excessively important.

5.4 Comparing counterfactuals with the existing literature

We now compare our results on counterfactual gender wage gaps in the two sectors with those obtained with the approach proposed by Chernozhukov, Fernández-Val and Melly (2013). First note that the two approaches differ in many aspects as explained in subection 2.3. In particular, our counterfactuals are generated by changing the conditional individual weights that capture the chances of getting job positions within a sector for the two genders, but the wage associated to every job position remains the same. The counterfactual reallocation of individuals thus leaves unchanged the sector log-wage distributions. By contrast, the application of CFM consists in computing the counterfactual gender log-wage gap in a given sector when using the gender log-wage distributions conditional on individual characteristics of the other sector. In that case, the sector log-wage distributions may well vary. Indeed, they result from gender distributions of individual characteristics within sectors (that remain unchanged) and gender conditional log-wage distributions that vary.

In practice, we implement CFM approach using their R package (Chen et al., 2016). For each sector and gender, we recover the first decile, first quartile, median, last quartile and last decile of the counterfactual log-wage distribution when using the conditional log-wage distributions of that gender in the other sector. We also compute the counterfactual gender gaps in each sector at those centiles, and their standard deviations using bootstrap with 100 replications. We compare the results that are reported in Table 4 with those obtained with our approach. Interestingly, they differ on two accounts. First, for a given sector, the differences in counterfactual gender gaps obtained with CFM when using alternatively conditional log-wage distributions in that sector and those in the other sector are larger than the differences obtained with our approach when changing conditional individual weights. This is not surprising since counterfactual sector log-wage distributions can vary with their approach but not with ours. Second, the counterfactual gender gap at some centiles moves in a different direction with CFM when using conditional log-wage distribution in the other sector. In particular, this occurs in the public sector at the first quartile and median. With CFM approach, the gender gap (in absolute terms) is more sizable when using the conditional log-wage distributions in the private sector rather than those in the public one. It is the opposite with our approach when rather changing the individual conditional weights. This occurs because the two approaches modify the gender log-wage distributions within sector in a different way. Whereas CFM change the conditional log-wage distributions, our approach re-assign females and males across job positions in a sector using the assignment rules of the other sector, and their new job positions result in different wages.

Overall, there are significant differences between the counterfactuals obtained with our approach and CFM. They are the consequences of different assumptions and mechanisms embedded in the two approaches.

[*Insert Table* 4]

6 Conclusion

In this paper, we show how to quantify pay-off gaps between groups with an assignment model that involves heterogeneous individuals and positions, as well as differences across groups in the propensity to get these positions. Individuals differ in their observable characteristics and they are primarily interested in the positions yielding the highest pay-offs. Some individuals do not apply because occupying these positions is too constraining. Individuals not selected turn to positions that are paid slightly less, and so on, until all positions are filled.

Our model can be estimated using a flexible semi-parametric approach. It is then possible to construct a counterfactual of the outcome distribution for each group when changing individuals' propensities to get positions conditional on their observable characteristics. Counterfactuals take into account the reallocation of individuals across positions and they can be compared across groups. Particular counterfactual situations of interest are obtained by fixing individuals' propensities to get positions to the same values for all groups or by fixing them to group-specific values corresponding to another context.

As an illustration, we use our approach to study gender wage differences in the public and private sectors for full-time workers using an original administrative dataset with accurate wage information. Whereas females are believed to be treated more fairly in the public sector, we find that the gender gap in propensity to get job positions along the wage distribution is rather similar in the two sectors. Results of a counterfactual exercise show that the gender average wage gap would be only slightly higher in the public sector if workers were attributed the same propensities to get job positions as in the private sector conditionally on their observable characteristics.

Our assignment framework can be extended in several dimensions. First, the model could be adapted to allow for the bunching of positions at some given values of pay-off. This is particularly relevant when considering a hierarchy of heterogeneous entities constituted of homogenous positions. Second, it could be of interest to introduce unobserved individual heterogeneity and assess under which assumptions the model can be estimated when using panel data. Third, principles of our assignment approach could be used to design a dynamic model such that individuals can move along the pay-off distribution by making transitions between available positions at each period.

References

[1] Albrecht J., Björklund A. and S. Vroman (2003), "Is There a Glass Ceiling in Sweden?", *Journal of Labour Economics*, 21(1), pp. 145-177.

[2] Andersen P. and R. Gill (1982), "Cox's regression model for counting processes: a large sample study", *The annals of statistics*, 10(4), pp. 1100-1120.

[3] Beggs S., Cardell S. and J. Hausman (1981), "Assessing the Potential Demand for Electrc Cars", *Journal of Econometrics*, 17(1), pp. 1-19.

[4] Chen M., Chernozhukov V., Fernández-Val I. and B. Melly (2016), "Counterfactual: an R package for counterfactual analysis".

[5] Chernozhukov V., Fernández-Val I. and B. Melly (2013), "Inference on counterfactual distributions", *Econometrica*, 81(6), pp. 2205-2268.

[6] Chiappori P.A. and B. Salanié (2016),"The Econometrics of Matching Models", *Journal of Economic Literature*, 54, pp. 832-861.

[7] Choo E. and A. Siow (2016), "Who Marries Whom and Why", *Journal of Political Economy*, 114(1), pp. 175-201.

[8] Dagsvik J. (1994), "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes", *Econometrica*, 62(5), pp. 1179-1205.

[9] Daussin-Benichou J., Koubi M., Leduc A. and B. Marc (2014), "Les carrières salariales dans le public et le privé", in *Emploi et salaires*, édition 2014, Insee, pp 47-60.

[10] Depalo D., Giordano R. and E. Papapetrou (2015), "Public-private wage differentials in euro-area countries: evidence from quantile decomposition analysis", *Empirical Economics*, 49(3), pp. 985-1015.

[11] Donald S., Green D. and H. Paarsch (2000), "Differences in Wage Distributions Between Canada and the United States: An Application of Flexible Estimator of Distribution Functions in the Presence of Covariates", *Review of Economic Studies*, 67, pp. 609-633.

[12] Doeringer P. and M. Piore (1985). *Internal labor markets and manpower analysis*, ME Sharpe, 248p.

[13] Dorothee O. and E. Baradji E. (2014), "L'emploi dans les trois versants de la fonction publique", *Point stat*, 15, DGAFP, Ministère de la décentralisation et de la fonction publique.

[14] Dorothee O., Le Faller P. and M. Treppoz (2013), “Les principales caractéristiques des agents de la fonction publique au 31 décembre 2011”, *Point stat*, 11, DGAFP, Ministère de la décentralisation et de la fonction publique.

[15] Dupuy A. and A. Galichon (2014), “Personality traits and the marriage market”, *Journal of Political Economy*, 122(6), pp. 1271-1319.

[16] Fack G., Grenet J. and Y. He (2019), “Beyond Truth-Telling: Preference Estimation with Centralized School Choice”, *American Economic Review*, 109(4), pp. 1486-1529.

[17] Firpo, N. Fortin and T. Lemieux (2009), “Unconditional quantile regressions”, *Econometrica*, 77(3), pp. 953-973.

[18] Fortin N. (2008), “The Gender Wage Gap among Young Adults in the United States The importance of Money versus People”, *Journal of Human Resources*, 43(4), pp.884-918.

[19] Fortin N. and T. Lemieux (1998), “Rank Regressions, Wage Distributions, and the Gender Gap”, *The Journal of Human Resources*, 33(3), pp. 610-643.

[20] Fortin N., Lemieux T. and N. Firpo (2011) , “Decomposition methods in economics”, in *Handbook of Labor Economics*, Ashenfelter O. and D. Card, Volume 4, Part A, pp. 1–102.

[21] Fortin N., Bell B. and N. Böhm (2017), “Top earnings inequality and the gender pay gap: Canada, Sweden, and the United Kingdom”, *Labour Economics*, 47, pp. 107-123.

[22] Gale D. and L. Shapley (1962), “College Admissions and the Stability of Marriage”, *American Mathematical Monthly*, 69 (1), pp. 9-15.

[23] Galichon A. and B. Salanié (2021), “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models”, *Review of Economic Studies*, forthcoming.

[24] Gobillon L., Meurs D. and S. Roux (2015), “Estimating gender differences in access to jobs”, *Journal of Labour Economics*, 33(2), pp. 317-363.

[25] Gourieroux C. and A. Monfort (1995), *Simulation-based econometric methods*, Oxford university press, 174p.

[26] Gregory R. and J. Borland (1999), “Recent developments in public sector labor markets”, in *Handbook of Labor Economics*, Ashenfelter O. and D. Card, Volume 3, Part C, pp. 3573–3630.

[27] Hart W. (1921), “The Cauchy-Lipschitz method for infinite systems of differential equations”, *American Journal of Mathematics*, 43(4), pp. 226-231.

[28] Hastings J., Kane T. and D. Staiger (2009), “Heterogeneous preferences and the efficacy of public school choice”, NBER Working Paper 2145.

[29] Maasoumi E. and L. Wang (2019), “The gender gap between earnings distributions”, *Journal of Political Economy*, 127(5), pp. 2438-2504.

[30] Machado J. and J. Mata (2005), “Counterfactual decomposition of changes in wage distributions using quantile regression”, *Journal of Applied Econometrics*, 20, pp. 445–465.

[31] Lucifora C. and D. Meurs (2006), “The Public Sector Pay Gap in France, Great Britain and Italy”, *Review of Income and Wealth*, 52(1), pp. 43–59.

[32] Melly (2005a), “Decomposition of differences in distribution using quantile regression”, *Labour Economics*, 12, pp. 577–590.

[33] Melly (2005b), “Public-private sector wage differentials in Germany: Evidence from quantile regression”, *Empirical Economics*, 30, pp. 505–520.

[34] Menzel K. (2015), “Large Matching Markets as Two-Sided Demand Systems”, *Econometrica*, 83(3), pp. 897-941. Menzel, Konrad. 2016. "Inference for Games with

[35] Oaxaca R. and M. Ransom (1994), “On discrimination and the decomposition of wage differentials”, *Journal of Econometrics*, 61, pp. 5–21.

[36] Ponthieux S. and D. Meurs (2015), “Gender inequality”, in *Handbook of Income Distribution*, Atkinson A. and F. Bourguignon, vol. 2, Elsevier, Chapter 12, pp. 981-1146.

[37] Rosén (1972), “Asymptotic Theory fo Successive Sampling with Varying Probabilities witout Replacement, I”, *The Annals of Mathematical Statistics*, 43(2), pp. 373-397.

[38] Rothe C. (2012), “Partial Distributional Policy Effects”, *Econometrica*, 80, pp. 2269–2301.

APPENDIX

A Solution of the model

We have the following theorem showing the existence and uniqueness of the solution:

Theorem 1 *Suppose that X can only take a finite number of values X^ℓ, $\ell = 1, ..., L$; $\mu\left(\cdot\left|X^\ell, m\right.\right)$ and $\mu\left(\cdot\left|X^\ell, j\right.\right)$ are C^1 on $(0,1]$ for each ℓ; and there is a constant $c > 0$ such that $\mu\left(u\left|X^\ell, m\right.\right) > c$ and $\mu\left(u\left|X^\ell, f\right.\right) > c$ for all $u \in (0,1]$ and all ℓ. Then there is a unique $2L$-uplet $\left\{n\left(\cdot\left|X^\ell, f\right.\right), n\left(\cdot\left|X^\ell, m\right.\right)\right\}_{\ell=1,\ldots,L}$ verifying (12) where $\phi\left(u\left|X, j\right.\right)$ is given by (10).*

Proof. The proof revolves around the application of the Cauchy-Lipschitz theorem. Plugging (10) into (12), we get for any j and k:

$$n'\left(u\left|X^k, j\right.\right) = \frac{n\left(u\left|X^k, j\right.\right)\mu\left(u\left|X^k, j\right.\right)}{\sum\limits_{\ell,g} n\left(u\left|X^\ell, g\right.\right)\mu\left(u\left|X^\ell, g\right.\right)} \tag{34}$$

Introduce the vectors

$$\overline{\mu}(u) = \left[\mu\left(u\left|X^1, f\right.\right), .., \mu\left(u\left|X^L, f\right.\right), \mu\left(u\left|X^1, m\right.\right), ..., \mu\left(u\left|X^L, m\right.\right)\right]' \tag{35}$$

$$\overline{n}(u) = \left[n\left(u\left|X^1, f\right.\right), .., n\left(u\left|X^L, f\right.\right), n\left(u\left|X^1, m\right.\right), ..., n\left(u\left|X^L, m\right.\right)\right]' \tag{36}$$

A stacked version of (47) is given by:

$$\overline{n}'(u) = g\left(u, \overline{n}(u)\right) \tag{37}$$

with:

$$g\left(u, \overline{n}(u)\right) = \frac{\overline{n}(u) .* . \overline{\mu}(u)}{\langle \overline{n}(u), \overline{\mu}(u)\rangle} \tag{38}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidian scalar product and for any two vectors V_1 and V_2 of same dimension, $V_1 . * . V_2$ is the vector where any element i is the product of the elements i of V_1 and V_2.

The equation (50) is a first-order differential equation. The denominators of all elements of $g(\cdot, \cdot)$ are strictly positive on $\widetilde{\Phi} = (0,1] \times \left[0, n\left(X^1, f\right)\right] \times ... \times \left[0, n\left(X^L, f\right)\right] \times \left[0, n\left(X^1, m\right)\right] \times ... \times \left[0, n\left(X^L, m\right)\right]$ where $n\left(X^\ell, j\right)$ is the measure of gender-j individuals with characteristics X^ℓ. This is because there is a constant $c > 0$ such that $\mu\left(u\left|X^\ell, m\right.\right) > c$ and $\mu\left(u\left|X^\ell, f\right.\right) > c$ for all ℓ and all $u \in (0,1]$. As $\mu\left(\cdot\left|X^\ell, m\right.\right)$ and $\mu\left(\cdot\left|X^\ell, f\right.\right)$ are C^1 on $(0,1]$ for all ℓ, it is then straightforward to show that $g(\cdot, \cdot)$ is C^1 on $\widetilde{\Phi}$. This yields that on any compact set $[\varepsilon, 1] \times \left[0, n\left(X^1, f\right)\right] \times ... \times \left[0, n\left(X^L, f\right)\right] \times \left[0, n\left(X^1, m\right)\right] \times ... \times \left[0, n\left(X^L, m\right)\right]$, $g(\cdot, \cdot)$ is Lipshitzienne and (50) has a unique solution for $\overline{n}(\cdot)$ on $[\varepsilon, 1]$ according to the Cauchy-Lipshitz theorem. As this is true for ε arbitrarily close to zero, (50) has a unique solution for $\overline{n}(\cdot)$ on $(0,1]$.

Note that, in our theorem, we make the assumption that attributes are discrete to end up with a finite system of first-order differential equations to which we can apply the Cauchy-Lipschitz theorem. Our theorem can be extended to the case where individual characteristics X take an infinite but countable number of values since Cauchy-Lipschitz theorem has been extended to a infinite but countable system of first-order differential equations (see Hart, 1921). We cannot extend our theorem to the case where there is at least one continuous explanatory variable as we could not find any version of the Cauchy-Lipschitz theorem for an infinite continous system of first-order differential equations.

B Theoretical foundations of the simulation approach

The finite discrete counterpart of the differential equation verified by the measures of available individuals (12) can be rewritten in vector form piling up (28) in the (X, j) dimension as:

$$\overrightarrow{N}^* \left(u^k\right) = \overrightarrow{N}^* \left(u^{k+1}\right) - \overrightarrow{D}_{k+1} \tag{39}$$

where $\overrightarrow{N}^* (u) = \left[N^* \left(u \left| X^1, j_1 \right.\right), ..., N^* \left(u \left| X^L, j_2 \right.\right)\right]'$ and $\overrightarrow{D}_k = \left[D_k \left(X^1, j_1\right), ..., D_k \left(X^L, j_2\right)\right]'$ with $\{j_1, j_2\} = \{f, m\}$. It is straightforward to show recursively that:

$$E\left[\overrightarrow{N}^* \left(u^k\right)\right] = \overrightarrow{N} - E\left(\overrightarrow{M}_{k+1}\right) \tag{40}$$

where $\overrightarrow{M}_k = \sum_{\ell=k}^{N} \overrightarrow{D}_\ell$ and $\overrightarrow{N} = \left[N\left(X^1, j_1\right), ..., N\left(X^L, j_2\right)\right]'$ where $N(X, j)$ is the number of gender-j individuals with characteristics X in the sample. We need a strategy to estimate the second right-hand side term. We have:

$$E\left(\overrightarrow{M}_k\right) = E_{\overrightarrow{M}_k, \overrightarrow{M}_{k+1}, ..., \overrightarrow{M}_N}\left(\overrightarrow{M}_k\right) \tag{41}$$

The expectation $E\left(\overrightarrow{M}_k\right)$ can thus be computed by simulation, averaging across iterations the values of $\overrightarrow{M}_k$ obtained when drawing values of $\overrightarrow{M}_k, \overrightarrow{M}_{k+1}, ...,$ and $\overrightarrow{M}_N$ which joint law verifies:

$$P\left(\overrightarrow{M}_k = \overrightarrow{m}_k, \overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}, ..., \overrightarrow{M}_N = \overrightarrow{m}_N\right) \tag{42}$$

$$= \left[\prod_{\ell=k+1}^{N-1} P\left(\overrightarrow{M}_k = \overrightarrow{m}_k \left| \overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}\right.\right)\right] P\left(\overrightarrow{M}_N = \overrightarrow{m}_N\right) \tag{43}$$

Draws can thus be made first drawing in the law of $\overrightarrow{M}_N$, and then sequentially in the law of $\overrightarrow{M}_k$ conditionally on the simulated value of $\overrightarrow{M}_{k+1}$ denoted $\overrightarrow{m}^s_{k+1}$ (Gourieroux and Monfort, 1996). The law of $\overrightarrow{M}_N = \overrightarrow{D}_N$ is simply that of a multinomial logit which probabilities are given by:

$$P\left(1\left|X,j\right.\right)=\frac{N\left(X,j\right)\exp\left[X\beta_j^*\left(1\right)\right]}{\sum_{\ell,g}N\left(X^\ell,g\right)\exp\left[X^\ell\beta_g^*\left(1\right)\right]} \tag{44}$$

where $N\left(X,j\right)$ is the number of gender-j individuals with characteristics X in the sample. Denote by $\Omega_j^*\left(u,X\right)$ the random set that contains all available gender-j individuals with characteristics X at rank u and $\overrightarrow{\Omega}^*\left(u\right)=\left\{\Omega_{j_1}^*\left(u,X^1\right),...,\Omega_{j_2}^*\left(u,X^L\right)\right\}$. The law of $\overrightarrow{M}_k\left|\overrightarrow{M}_{k+1}=\overrightarrow{m}_{k+1}^s\right.$ is simple because $\overrightarrow{m}_{k+1}^s$ contains all the information necessary to determine the realization of $\overrightarrow{\Omega}^*\left(u\right)$ that we denote $\overrightarrow{\Omega}^s\left(u\right)$. We have:

$$P\left(\overrightarrow{M}_k=\overrightarrow{m}_k\left|\overrightarrow{M}_{k+1}=\overrightarrow{m}_{k+1}^s\right.\right)=P\left[\overrightarrow{D}_\ell=\overrightarrow{m}_k-\overrightarrow{m}_{k+1}^s\left|\overrightarrow{\Omega}^*\left(u^k\right)=\overrightarrow{\Omega}^s\left(u^k\right)\right.\right] \tag{45}$$

where $\overrightarrow{m}_k-\overrightarrow{m}_{k+1}^s$ is a vector where there is only one element that takes the value one and this occurs at the position associated to the characteristics $\left(X,j\right)$ of the individual who gets the position, and other elements of the vector take the value zero. The probability that the element corresponding to a given $\left(X,j\right)$ takes the value one (and other elements take value zero) is given by:

$$P^s\left(u^k\left|X,j\right.\right)=\frac{N^s\left(u^k\left|X,j\right.\right)\exp\left[X\beta_j^*\left(u^k\right)\right]}{\sum_{\ell,g}N^s\left(u^k\left|X^\ell,g\right.\right)\exp\left[X\beta_g^*\left(u^k\right)\right]} \tag{46}$$

where $N^s\left(u\left|X,j\right.\right)=Card\ \Omega_j^s\left(u,X\right)$ with $\Omega_j^s\left(u,X\right)$ the set of available gender-j individuals with characteristics X.

For a given simulation iteration, we first draw for rank 1 in the law of a multinomial logit using formula (57), and we then draw sequentially for ranks $u^k=\left(k-1\right)/\left(N-1\right)$ with $k=N-1,...,1$ in the laws of multinomial logits using formula (59). A simulated value of $\overrightarrow{M}_k$ is denoted by $\overrightarrow{M}_k^s$. A consistent estimator of $E\left[\overrightarrow{N}^*\left(u^k\right)\right]$ when the number of simulations tends to infinity is then given by $\overrightarrow{N}-\sum_{s=1}^S\overrightarrow{M}_{k+1}^s/S$.

C The public sector in France

The French public sector accounts for around 20% of total salaried employment. Most public employees are females (61%) whereas this is not the case in the private sector (44%) (Dorothee, Le Faller and Treppoz, 2013). The public sector is divided into three subsectors: central administration (44% of employment in the public sector) which includes education, local government (35%) and public health (21%). The share of local government in employment has increased over the last 10 years in line with the decentralization process occurring during that period (Dorothée and Baradji, 2014).

In France, the public sector has a highly centralized pay setting compared to the private sector. A common pay scale is applied to all subsectors, which means that the nominal value of the basic wage is the same at any given grade through the entire public sector. However, individual differences in earnings may arise from bonuses which are mostly related to the type of occupation. Due to budget constraints, the basic wage is constant in nominal terms since 2010, and there has been no major change in pay scale by occupation in the past decade. As a consequence, advancement along the pay grid is currently the main way to get a pay rise.

The French public sector is very close to the model of internal labor market proposed by Doeringer and Piore (1985). The main recruitment process is a competitive exam with diploma requirements specific to the type of occupation. There are as many competitive exams as types of occupations (policemen, judges, teachers, nurses, clerks, academics, etc.), the most prestigious one leading to careers in top management in the public administration through ENA (*Ecole Nationale d'Administration*). Once recruited, civil servants start their career at the bottom of the pay scale specific to their occupation. Mobility between the public and private sectors is very limited and occurs mainly at early stage of careers (Daussin-Benichou et al., 2014). Wage increases depend much on seniority despite recent reforms which aim at taking into account individual performance and/or local work conditions. Promotions are seldom events which also depend on seniority to some extent and are most often obtained through competitive or professional exams.

D Additional results

D.1 Contribution of explanatory variables to Oaxaca decomposition

In the Oaxaca decomposition, it is possible to decompose the explained part of the gender probability ratio of getting a given job into the contributions of each category dummy of every explanatory variable. The contribution of a given category dummy D to the explained part at a given rank u is $\left[E\left(D\,|f,u\right)-E\left(D\,|m,u\right)\right]\beta_D^r\left(u\right)$, where $\beta_D^r\left(u\right)$ is the polynomial coefficient of the category dummy. Figure D.1 suggests that it is mostly long part-time experience that has some explanatory power. In the two sectors, long part-time experience significantly decreases females' relative chances of getting job positions along the wage distribution. In the private sector, its contribution represents around 30% of the gender probability ratio of getting a given job position up to rank 0.8. In the public sector, long part-time experience explains most of the gender probability ratio up to rank 0.2, but its importance then decreases with the rank and becomes small after rank 0.6. Diplomas have a small explanatory power and increase only slightly females' propensity to get job positions above rank 0.6, especially in the private sector. We then decompose the unexplained part of gender probability ratio of getting a given job position into the contribution of each category dummy of every explanatory variable. The contribution of a given category dummy D to the unexplained part at a given rank u is $E\left(D\,|f,u\right)\left[\beta_{fD}\left(u\right)-\beta_D^r\left(u\right)\right]-E\left(D\,|m,u\right)\left[\beta_{mD}\left(u\right)-\beta_D^r\left(u\right)\right]$, where $\beta_{jD}\left(u\right)$ is the gender-j polynomial coefficient of the category dummy. Figure D.2 shows that the contribution of the gender difference in the returns to every category dummy is negligible or very small. Overall, the Oaxaca decomposition suggests that gender differences in propensity to get job positions cannot be explained by composition effects or differences in the returns of observables.

D.2 Estimation results using the hourly wage when both full-time and part-time workers are included in the sample

So far, we have conducted our analysis on the subsample of full-time workers, but this can lead to biases in the estimates if there are selection effects. It is possible that in a given sector a larger share of females ends up in part-time job positions and ignoring them leads to an underestimation of the gender differences in the propensity to get full-time job positions. Moreover, we use the daily wage to rank job positions whereas there can be gender differences in the number of hours worked. If females work on average less hours in some positions than males, their daily wage is likely to be lower and their rank in the hierarchy of job positions is underestimated. Not taking into account the number of hours worked leads to an overestimation of the gender differences in the propensity to get job positions at some ranks, and an underestimation of this gender difference at ranks below them.

We repeat our analysis considering the hourly wage instead of the daily wage as our outcome of interest. Estimations are conducted on a sample that now includes part-time workers in addition to full-time ones. Results on counterfactuals in the different scenarios are qualitatively similar but quantitatively slightly different from those in our benchmark case. As shown by Table D.1, the gender average wage gap and gender wage difference at the last decile in the public sector when using the assignment rules in the private sector are now respectively 2.4 and 6.5 percentage points higher than when considering the assignment rules in the public sector (instead of 0.7 and 3.6 percentage points respectively in the benchmark case).

D.3 Influence of unobserved individual heterogeneity

We also assess to what extent the estimated gender probability ratio of getting a job might be influenced by unobserved individual heterogeneity ignored from the specification as it may bias our results. For that purpose, we first estimate the conditional individual weights for each gender using our semi-parametric approach. We then add an unobserved individual heterogeneity term to the logarithm of the conditional individual weights of each worker and reassign workers to positions in each sector using the same simulation approach as when we constructed counterfactuals. Finally, we assess to what extent the non-parametric estimators of the gender probability ratio of getting a given job position obtained after the reassignment of workers differ depending on the dispersion of unobserved individual heterogeneity terms. In our simulations, these terms are drawn identically and independently in a centered normal law with variance $k^2.V\left(X_i\overline{\beta}_{j(i)}\right)$, where $\overline{\beta}_j = \int_0^1 \beta_j(u)\,du$ and k is a parameters that affects the scale of the variance. Figures D.3 shows that increasing k leads to a decrease in the slope of the estimated gender probability ratio of getting a job position both in the public and private sectors. Nevertheless, k must be large (above 1) for the decrease in the slope to be significant.

APPENDIX

A Solution of the model

We have the following theorem showing the existence and uniqueness of the solution:

Theorem 2 *Suppose that X can only take a finite number of values X^ℓ, $\ell = 1, ..., L$; $\mu\left(\cdot\left|X^\ell, m\right.\right)$ and $\mu\left(\cdot\left|X^\ell, j\right.\right)$ are C^1 on $(0,1]$ for each ℓ; and there is a constant $c > 0$ such that $\mu\left(u\left|X^\ell, m\right.\right) > c$ and $\mu\left(u\left|X^\ell, f\right.\right) > c$ for all $u \in (0,1]$ and all ℓ. Then there is a unique $2L$-uplet $\left\{n\left(\cdot\left|X^\ell, f\right.\right), n\left(\cdot\left|X^\ell, m\right.\right)\right\}_{\ell=1,\ldots,L}$ verifying (12) where $\phi\left(u\left|X, j\right.\right)$ is given by (10).*

Proof. The proof revolves around the application of the Cauchy-Lipschitz theorem. Plugging (10) into (12), we get for any j and k:

$$n'\left(u\left|X^k, j\right.\right) = \frac{n\left(u\left|X^k, j\right.\right)\mu\left(u\left|X^k, j\right.\right)}{\sum\limits_{\ell,g} n\left(u\left|X^\ell, g\right.\right)\mu\left(u\left|X^\ell, g\right.\right)} \tag{47}$$

Introduce the vectors

$$\overline{\mu}(u) = \left[\mu\left(u\left|X^1, f\right.\right), .., \mu\left(u\left|X^L, f\right.\right), \mu\left(u\left|X^1, m\right.\right), ..., \mu\left(u\left|X^L, m\right.\right)\right]' \tag{48}$$

$$\overline{n}(u) = \left[n\left(u\left|X^1, f\right.\right), .., n\left(u\left|X^L, f\right.\right), n\left(u\left|X^1, m\right.\right), ..., n\left(u\left|X^L, m\right.\right)\right]' \tag{49}$$

A stacked version of (47) is given by:

$$\overline{n}'(u) = g\left(u, \overline{n}(u)\right) \tag{50}$$

with:

$$g\left(u, \overline{n}(u)\right) = \frac{\overline{n}(u) .* .\overline{\mu}(u)}{\left\langle \overline{n}(u), \overline{\mu}(u)\right\rangle} \tag{51}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidian scalar product and for any two vectors V_1 and V_2 of same dimension, $V_1 .* .V_2$ is the vector where any element i is the product of the elements i of V_1 and V_2.

The equation (50) is a first-order differential equation. The denominators of all elements of $g(\cdot, \cdot)$ are strictly positive on $\widetilde{\Phi} = (0,1] \times \left[0, n\left(X^1, f\right)\right] \times ... \times \left[0, n\left(X^L, f\right)\right] \times \left[0, n\left(X^1, m\right)\right] \times ... \times \left[0, n\left(X^L, m\right)\right]$ where $n\left(X^\ell, j\right)$ is the measure of gender-j individuals with characteristics X^ℓ. This is because there is a constant $c > 0$ such that $\mu\left(u\left|X^\ell, m\right.\right) > c$ and $\mu\left(u\left|X^\ell, f\right.\right) > c$ for all ℓ and all $u \in (0,1]$. As $\mu\left(\cdot\left|X^\ell, m\right.\right)$ and $\mu\left(\cdot\left|X^\ell, f\right.\right)$ are C^1 on $(0,1]$ for all ℓ, it is then straightforward to show that $g(\cdot, \cdot)$ is C^1 on $\widetilde{\Phi}$. This yields that on any compact set $[\varepsilon, 1] \times \left[0, n\left(X^1, f\right)\right] \times ... \times \left[0, n\left(X^L, f\right)\right] \times \left[0, n\left(X^1, m\right)\right] \times ... \times \left[0, n\left(X^L, m\right)\right]$, $g(\cdot, \cdot)$ is Lipshitzienne and (50) has a unique solution for $\overline{n}(\cdot)$ on $[\varepsilon, 1]$. As this is true for ε arbitrarily close to zero, (50) has a unique solution for $\overline{n}(\cdot)$ on $(0,1]$.

B Theoretical foundations of the simulation approach

The finite discrete counterpart of the differential equation verified by the measures of available individuals (12) can be rewritten in vector form piling up (28) in the (X, j) dimension as:

$$\overrightarrow{N}^*\left(u^k\right) = \overrightarrow{N}^*\left(u^{k+1}\right) - \overrightarrow{D}_{k+1} \tag{52}$$

where $\overrightarrow{N}^*(u) = \left[N^*\left(u\left|X^1, j_1\right.\right), ..., N^*\left(u\left|X^L, j_2\right.\right)\right]'$ and $\overrightarrow{D}_k = \left[D_k\left(X^1, j_1\right), ..., D_k\left(X^L, j_2\right)\right]'$ with $\{j_1, j_2\} = \{f, m\}$. It is straightforward to show recursively that:

$$E\left[\overrightarrow{N}^*\left(u^k\right)\right] = \overrightarrow{N} - E\left(\overrightarrow{M}_{k+1}\right) \tag{53}$$

where $\overrightarrow{M}_k = \sum_{\ell=k}^N \overrightarrow{D}_\ell$ and $\overrightarrow{N} = \left[N\left(X^1, j_1\right), ..., N\left(X^L, j_2\right)\right]'$ where $N(X, j)$ is the number of gender-j individuals with characteristics X in the sample. We need a strategy to estimate the second right-hand side term. We have:

$$E\left(\overrightarrow{M}_k\right) = E_{\overrightarrow{M}_k, \overrightarrow{M}_{k+1}, ..., \overrightarrow{M}_N}\left(\overrightarrow{M}_k\right) \tag{54}$$

The expectation $E\left(\overrightarrow{M}_k\right)$ can thus be computed by simulation, averaging across iterations the values of $\overrightarrow{M}_k$ obtained when drawing values of $\overrightarrow{M}_k, \overrightarrow{M}_{k+1}, ...,$ and $\overrightarrow{M}_N$ which joint law verifies:

$$P\left(\overrightarrow{M}_k = \overrightarrow{m}_k, \overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}, ..., \overrightarrow{M}_N = \overrightarrow{m}_N\right) \tag{55}$$

$$= \left[\prod_{\ell=k+1}^{N-1} P\left(\overrightarrow{M}_k = \overrightarrow{m}_k \left|\overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}\right.\right)\right] P\left(\overrightarrow{M}_N = \overrightarrow{m}_N\right) \tag{56}$$

Draws can thus be made first drawing in the law of $\overrightarrow{M}_N$, and then sequentially in the law of $\overrightarrow{M}_k$ conditionally on the simulated value of $\overrightarrow{M}_{k+1}$ denoted $\overrightarrow{m}_{k+1}^s$ (Gourieroux and Monfort, 1996). The law of $\overrightarrow{M}_N = \overrightarrow{D}_N$ is simply that of a multinomial logit which probabilities are given by:

$$P\left(1\left|X, j\right.\right) = \frac{N(X, j) \exp\left[X\beta_j^*(1)\right]}{\sum_{\ell,g} N\left(X^\ell, g\right) \exp\left[X^\ell \beta_g^*(1)\right]} \tag{57}$$

where $N(X, j)$ is the number of gender-j individuals with characteristics X in the sample. Denote by $\Omega_j^*(u, X)$ the random set that contains all available gender-j individuals with characteristics X at rank u and $\overrightarrow{\Omega}^*(u) = \left\{\Omega_{j_1}^*\left(u, X^1\right), ..., \Omega_{j_2}^*\left(u, X^L\right)\right\}$. The law of $\overrightarrow{M}_k \left|\overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}^s\right.$ is simple because $\overrightarrow{m}_{k+1}^s$ contains all the information necessary to determine the realization of $\overrightarrow{\Omega}^*(u)$ that we denote $\overrightarrow{\Omega}^s(u)$. We have:

$$P\left(\overrightarrow{M}_k = \overrightarrow{m}_k \left|\overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}^s\right.\right) = P\left[\overrightarrow{D}_\ell = \overrightarrow{m}_k - \overrightarrow{m}_{k+1}^s \left|\overrightarrow{\Omega}^*\left(u^k\right) = \overrightarrow{\Omega}^s\left(u^k\right)\right.\right] \tag{58}$$

where $\overrightarrow{m}_k - \overrightarrow{m}^s_{k+1}$ is a vector where there is only one element that takes the value one and this occurs at the position associated to the characteristics (X, j) of the individual who gets the position, and other elements of the vector take the value zero. The probability that the element corresponding to a given (X, j) takes the value one (and other elements take value zero) is given by:

$$P^s\left(u^k\left|X, j\right.\right) = \frac{N^s\left(u^k\left|X, j\right.\right)\exp\left[X\beta^*_j\left(u^k\right)\right]}{\sum_{\ell,g} N^s\left(u^k\left|X^\ell, g\right.\right)\exp\left[X\beta^*_g\left(u^k\right)\right]} \quad (59)$$

where $N^s\left(u\left|X, j\right.\right) = Card\ \Omega^s_j\left(u, X\right)$ with $\Omega^s_j\left(u, X\right)$ the set of available gender-j individuals with characteristics X.

For a given simulation iteration, we first draw for rank 1 in the law of a multinomial logit using formula (57), and we then draw sequentially for ranks $u^k = (k-1)/(N-1)$ with $k = N-1, ..., 1$ in the laws of multinomial logits using formula (59). A simulated value of $\overrightarrow{M}_k$ is denoted by $\overrightarrow{M}^s_k$. A consistent estimator of $E\left[\overrightarrow{N}^*\left(u^k\right)\right]$ when the number of simulations tends to infinity is then given by $\overrightarrow{N} - \sum_{s=1}^S \overrightarrow{M}^s_{k+1}/S$.

C The public sector in France

The French public sector accounts for around 20% of total salaried employment. Most public employees are females (61%) whereas this is not the case in the private sector (44%) (Dorothee, Le Faller and Treppoz, 2013). The public sector is divided into three subsectors: central administration (44% of employment in the public sector) which includes education, local government (35%) and public health (21%). The share of local government in employment has increased over the last 10 years in line with the decentralization process occurring during that period (Dorothée and Baradji, 2014).

In France, the public sector has a highly centralized pay setting compared to the private sector. A common pay scale is applied to all subsectors, which means that the nominal value of the basic wage is the same at any given grade through the entire public sector. However, individual differences in earnings may arise from bonuses which are mostly related to the type of occupation. Due to budget constraints, the basic wage is constant in nominal terms since 2010, and there has been no major change in pay scale by occupation in the past decade. As a consequence, advancement along the pay grid is currently the main way to get a pay rise.

The French public sector is very close to the model of internal labor market proposed by Doeringer and Piore (1985). The main recruitment process is a competitive exam with diploma requirements specific to the type of occupation. There are as many competitive exams as types of occupations (policemen, judges, teachers, nurses, clerks, academics, etc.), the most prestigious one leading to careers in top management in the public administration through ENA (*Ecole Nationale d'Administration*). Once recruited, civil servants start their career at the bottom of the pay scale specific to their occupation. Mobility between the public and private sectors is very limited and occurs mainly at early stage of careers (Daussin-Benichou et al., 2014). Wage increases depend much on seniority despite recent reforms which aim at taking into account individual performance and/or local work conditions. Promotions are seldom events which also depend on seniority to some extent and are most often obtained through competitive or professional exams.

D Additional results

D.1 Contribution of explanatory variables to Oaxaca decomposition

In the Oaxaca decomposition, it is possible to decompose the explained part of the gender probability ratio of getting a given job into the contributions of each category dummy of every explanatory variable. The contribution of a given category dummy D to the explained part at a given rank u is $\left[E\left(D\,|f,u\right)-E\left(D\,|m,u\right)\right]\beta^r_D\left(u\right)$, where $\beta^r_D\left(u\right)$ is the polynomial coefficient of the category dummy. Figure D.1 suggests that it is mostly long part-time experience that has some explanatory power. In the two sectors, long part-time experience significantly decreases females' relative chances of getting job positions along the wage distribution. In the private sector, its contribution represents around 30% of the gender probability ratio of getting a given job position up to rank 0.8. In the public sector, long part-time experience explains most of the gender probability ratio up to rank 0.2, but its importance then decreases with the rank and becomes small after rank 0.6. Diplomas have a small explanatory power and increase only slightly females' propensity to get job positions above rank 0.6, especially in the private sector. We then decompose the unexplained part of gender probability ratio of getting a given job position into the contribution of each category dummy of every explanatory variable. The contribution of a given category dummy D to the unexplained part at a given rank u is $E\left(D\,|f,u\right)\left[\beta_{fD}\left(u\right)-\beta^r_D\left(u\right)\right]-E\left(D\,|m,u\right)\left[\beta_{mD}\left(u\right)-\beta^r_D\left(u\right)\right]$, where $\beta_{jD}\left(u\right)$ is the gender-j polynomial coefficient of the category dummy. Figure D.2 shows that the contribution of the gender difference in the returns to every category dummy is negligible or very small. Overall, the Oaxaca decomposition suggests that gender differences in propensity to get job positions cannot be explained by composition effects or differences in the returns of observables.

D.2 Estimation results using the hourly wage when both full-time and part-time workers are included in the sample

So far, we have conducted our analysis on the subsample of full-time workers, but this can lead to biases in the estimates if there are selection effects. It is possible that in a given sector a larger share of females ends up in part-time job positions and ignoring them leads to an underestimation of the gender differences in the propensity to get full-time job positions. Moreover, we use the daily wage to rank job positions whereas there can be gender differences in the number of hours worked. If females work on average less hours in some positions than males, their daily wage is likely to be lower and their rank in the hierarchy of job positions is underestimated. Not taking into account the number of hours worked leads to an overestimation of the gender differences in the propensity to get job positions at some ranks, and an underestimation of this gender difference at ranks below them.

We repeat our analysis considering the hourly wage instead of the daily wage as our outcome of interest. Estimations are conducted on a sample that now includes part-time workers in addition to full-time ones. Results on counterfactuals in the different scenarios are qualitatively similar but quantitatively slightly different from those in our benchmark case. As shown by Table D.1, the gender average wage gap and gender wage difference at the last decile in the public sector when using the assignment rules in the private sector are now respectively 2.4 and 6.5 percentage points higher than when considering the assignment rules in the public sector (instead of 0.7 and 3.6 percentage points respectively in the benchmark case).

D.3 Influence of unobserved individual heterogeneity

We also assess to what extent the estimated gender probability ratio of getting a job might be influenced by unobserved individual heterogeneity ignored from the specification as it may bias our results. For that purpose, we first estimate the conditional individual weights for each gender using our semi-parametric approach. We then add an unobserved individual heterogeneity term to the logarithm of the conditional individual weights of each worker and reassign workers to positions in each sector using the same simulation approach as when we constructed counterfactuals. Finally, we assess to what extent the non-parametric estimators of the gender probability ratio of getting a given job position obtained after the reassignment of workers differ depending on the dispersion of unobserved individual heterogeneity terms. In our simulations, these terms are drawn identically and independently in a centered normal law with variance $k^2.V\left(X_i\overline{\beta}_{j(i)}\right)$, where $\overline{\beta}_j = \int_0^1 \beta_j\left(u\right)du$ and k is a parameters that affects the scale of the variance. Figures D.3 shows that increasing k leads to a decrease in the slope of the estimated gender probability ratio of getting a job position both in the public and private sectors. Nevertheless, k must be large (above 1) for the decrease in the slope to be significant.

Table 1: Descriptive statistics on wages (in euros) by gender

	Public sector				Private sector			
	All	Males	Females	% diff.	All	Males	Females	% diff.
Mean	70.6	76.2	65.2	-14.4%	79.2	84.9	68.7	-19.1%
Standard Deviation	32.4	36.0	27.4	-23.8%	69.9	77.6	51.0	-34.3%
5th centile	41.8	44.4	40.1	-9.8%	39.8	41.4	38.2	-7.6%
First decile	45.3	47.9	43.5	-9.2%	43.1	45.0	40.7	-9.4%
First quartile	51.1	54.8	48.9	-10.9%	50.2	52.9	46.7	-11.7%
Median	62.2	67.2	57.2	-14.9%	63.0	66.1	56.8	-14.0%
Last quartile	79.6	84.7	73.3	-13.4%	86.4	92.1	76.3	-17.2%
Last decile	102.2	112.6	93.7	-16.7%	126.7	137.1	105.9	-22.7%
95th centile	126.6	139.7	115.3	-17.4%	165.8	180.3	133.3	-26.1%
N	9,732	4,781	4,951		46,149	29,964	16,185	

Note: Statistics are computed for the daily wage. The % difference is the female value minus the male value divided by the male value.

Table 2: Descriptive statistics on explanatory variables by gender

	Public sector				Private sector			
	All	Males	Females	% diff.	All	Males	Females	% diff.
Female	0.509				0.351			
Diploma								
<High-School	0.521	0.544	0.399	-8.3%	0.515	0.563	0.425	-24.6%
High-School	0.207	0.200	0.213	+6.9%	0.187	0.165	0.226	+36.7%
College $\leq$ 2 years	0.128	0.112	0.144	+29.3%	0.157	0.133	0.202	+52.0%
College $>$ 2 years	0.144	0.144	0.143	-0.8%	0.142	0.139	0.147	+6.4%
Children								
0	0.208	0.224	0.193	-14.1%	0.217	0.222	0.209	-6.0%
1 or 2	0.572	0.543	0.599	+10.3%	0.580	0.552	0.632	+14.4%
$\geq$3	0.220	0.233	0.208	-10.5%	0.203	0.226	0.160	-29.4%
Age								
31-40	0.245	0.274	0.218	-20.7%	0.331	0.336	0.323	-3.9%
41-50	0.353	0.355	0.351	-1.2%	0.369	0.369	0.369	-0.1%
$\geq$51	0.402	0.371	0.432	+16.4%	0.300	0.296	0.309	+4.5%
Paris region								
Inside	0.234	0.211	0.256	+20.8%	0.235	0.224	0.255	+13.8%
Outside	0.766	0.786	0.744	-5.6%	0.765	0.776	0.745	-4.0%
Firm tenure								
$\leq$10 years	0.421	0.413	0.428	+3.7%	0.687	0.684	0.692	+1.2%
>10 years	0.579	0.587	0.572	-2.6%	0.313	0.316	0.308	-2.5%
Part-time experience								
$\leq$7%	0.459	0.600	0.322	-46.3%	0.500	0.582	0.348	-40.2%
>7% and $\leq$18%	0.202	0.215	0.189	-12.2%	0.239	0.243	0.231	-5.1%
>18%	0.339	0.184	0.489	+164.8%	0.261	0.175	0.421	+141.2%
Work interruption								
$\leq$ 1 year	0.233	0.237	0.229	-3.3%	0.203	0.207	0.196	-5.6%
$>$ 1 and $\leq$ 3 years	0.184	0.207	0.162	-21.4%	0.274	0.289	0.246	-14.7%
$>$ 3 and $\leq$ 6 years	0.239	0.256	0.223	-13.1%	0.271	0.277	0.259	-6.5%
$>$ 6 years	0.344	0.300	0.386	+28.4%	0.252	0.227	0.299	+31.7%

Note: Figures in the table correspond to proportions except % difference which is the female value minus the male value divided by the male value.

Table 3: Observed and counterfactual log-wage gaps obtained in different scenarios

	Public sector				Private sector			
	Observed	Public weights	Equal weights	Private weights	Observed	Private weights	Equal weights	Public weights
Mean								
Males	4.2596	4.2602	4.1999	4.2638	4.2953	4.2954	4.2383	4.2922
Females	4.1165	4.1159	4.1741	4.1124	4.1305	4.1304	4.2360	4.1363
F-M Gap	-0.1431	-0.1444	-0.0258	-0.1514	-0.1648	-0.1649	-0.0023	-0.1559
		(0.0069)	(0.0058)	(0.0064)		(0.0048)	(0.0031)	(0.0090)
Standard Deviation								
Males	0.3584	0.3571	0.3528	0.3663	0.4716	0.4706	0.4571	0.4653
Females	0.3284	0.3294	0.3484	0.3178	0.4007	0.4028	0.4506	0.4164
F-M Gap	-0.0300	-0.0277	-0.0044	-0.0484	-0.0710	-0.0678	-0.0066	-0.0489
		(0.0073)	(0.0049)	(0.0056)		(0.0052)	(0.0029)	(0.0105)
First decile								
Males	3.8698	3.8717	3.8298	3.8724	3.8059	3.8069	3.7660	3.8012
Females	3.7729	3.7734	3.7964	3.7731	3.7071	3.7087	3.7561	3.7120
F-M Gap	-0.0969	-0.0983	-0.0334	-0.0994	-0.0987	-0.0982	-0.0099	-0.0892
		(0.0070)	(0.0047)	(0.0049)		(0.0038)	(0.0020)	(0.0074)
First quartile								
Males	4.0044	4.0041	3.9455	4.0012	3.9676	3.9675	3.9184	3.9692
Females	3.8893	3.8886	3.9207	3.8884	3.8430	3.8416	3.9130	3.8471
F-M Gap	-0.1151	-0.1155	-0.0248	-0.1128	-0.1246	-0.1259	-0.0054	-0.1220
		(0.0057)	(0.0039)	(0.0051)		(0.0038)	(0.0023)	(0.0064)
Median								
Males	4.2070	4.2066	4.1370	4.2021	4.1906	4.1923	4.1407	4.1987
Females	4.0461	4.0438	4.1202	4.0483	4.0403	4.0419	4.1469	4.0335
F-M Gap	-0.1609	-0.1628	-0.0168	-0.1538	-0.1503	-0.1504	0.0062	-0.1653
		(0.0087)	(0.0064)	(0.0076)		(0.0051)	(0.0030)	(0.0078)
Last quartile								
Males	4.4388	4.4436	4.3837	4.4541	4.5230	4.5228	4.4534	4.5134
Females	4.2952	4.2934	4.3714	4.2954	4.3346	4.3355	4.4666	4.3327
F-M Gap	-0.1435	-0.1502	-0.0123	-0.1587	-0.1884	-0.1873	0.0132	-0.1807
		(0.0104)	(0.0078)	(0.0087)		(0.0070)	(0.0047)	0.0155
Last decile								
Males	4.7236	4.7267	4.6493	4.7510	4.9207	4.9213	4.8478	4.9016
Females	4.5403	4.5407	4.6110	4.5207	4.6628	4.6652	4.8340	4.7089
F-M Gap	-0.1833	-0.1860	-0.0383	-0.2303	-0.2579	-0.2561	-0.0138	-0.1927
		(0.0176)	(0.0144)	(0.0123)		(0.0104)	(0.0076)	(0.0235)

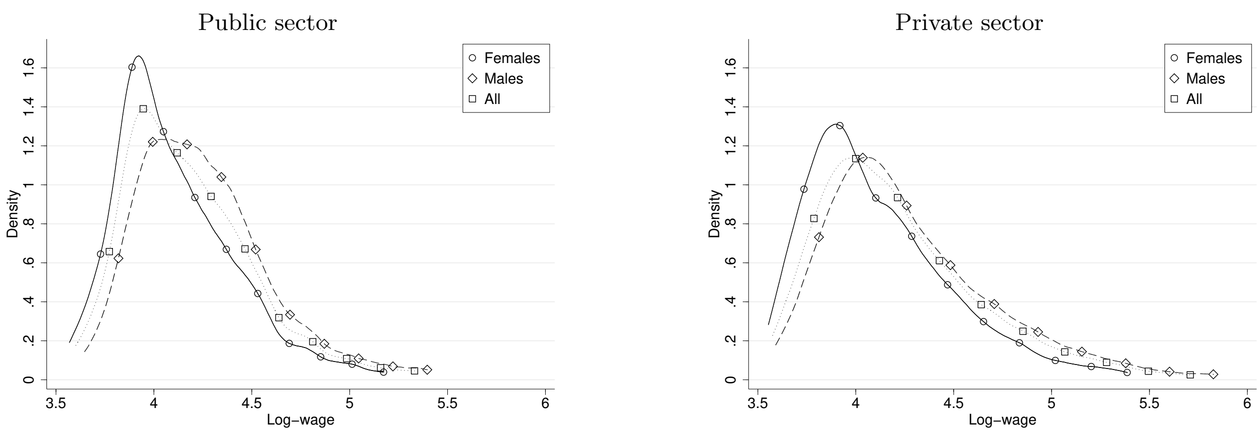
Note: Statistics are computed for the logarithm of daily wage. Column headings mention either that statistics are computed directly from the data (label "Observed") or that they are derived from counterfactuals using the conditional individual weights of the public sector, the private sector, or the same conditional individual weights for the two genders in the sector that is considered (respective labels: "Private weights", "Public weights", ''Equal weights"). For counterfactual exercises, the standard deviation of the gender gap obtained by bootstrap is reported in parentheses.

Table 4: Counterfactual log-wage gaps obtained using Chernozhukov, Fernández-Val and Melly (2013)

	Public sector		Private sector	
Conditional wage distributions	Public	Private	Private	Public
First decile				
Males	3.8706	3.8357	3.8063	3.8458
Females	3.7744	3.7200	3.7091	3.7698
F-M Gap	-0.0962	-0.1158	-0.0972	-0.0760
	(0.0065)	(0.0057)	(0.0039)	(0.0069)
First quartile				
Males	4.0055	4.0050	3.9709	3.9737
Females	3.8939	3.8602	3.8499	3.8836
F-M Gap	-0.1116	-0.1448	-0.1211	-0.0901
	(0.0066)	(0.0053)	(0.0044)	(0.0072)
Median				
Males	4.2119	4.2355	4.1950	4.1601
Females	4.0483	4.0685	4.0518	4.0257
F-M Gap	-0.1636	-0.1671	-0.1431	-0.1344
	(0.0102)	(0.0093)	(0.0052)	(0.0077)
Last quartile				
Males	4.4444	4.5593	4.5352	4.3936
Females	4.3005	4.3556	4.3417	4.2662
F-M Gap	-0.1440	-0.2037	-0.1935	-0.1274
	(0.0105)	(0.0107)	(0.0065)	(0.0110)
Last decile				
Males	4.7345	4.9413	4.9244	4.6801
Females	4.5414	4.6676	4.6855	4.5229
F-M Gap	-0.1931	-0.2737	-0.2389	-0.1572
	(0.0162)	(0.0185)	(0.0129)	(0.0187)

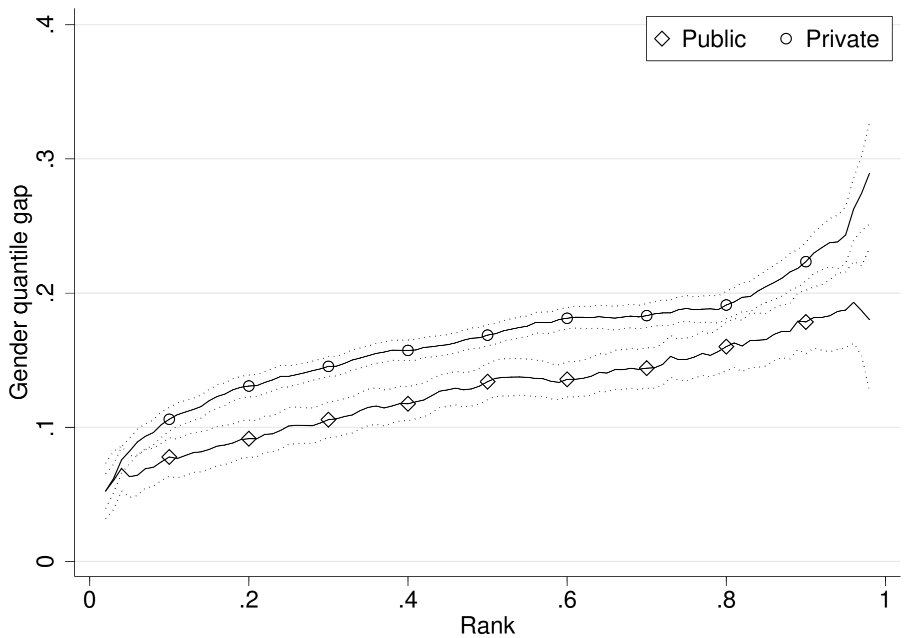
Note: Statistics are computed for the logarithm of daily wage. Column headings give the sector of interest and mentions whether counterfactuals are constructed from conditional log-wage distributions in the public or the private sector. The standard deviations of gender gaps obtained by bootstrap with 100 replications are reported in parentheses.

Figure 1: Gender log-wage distributions in the public and private sectors



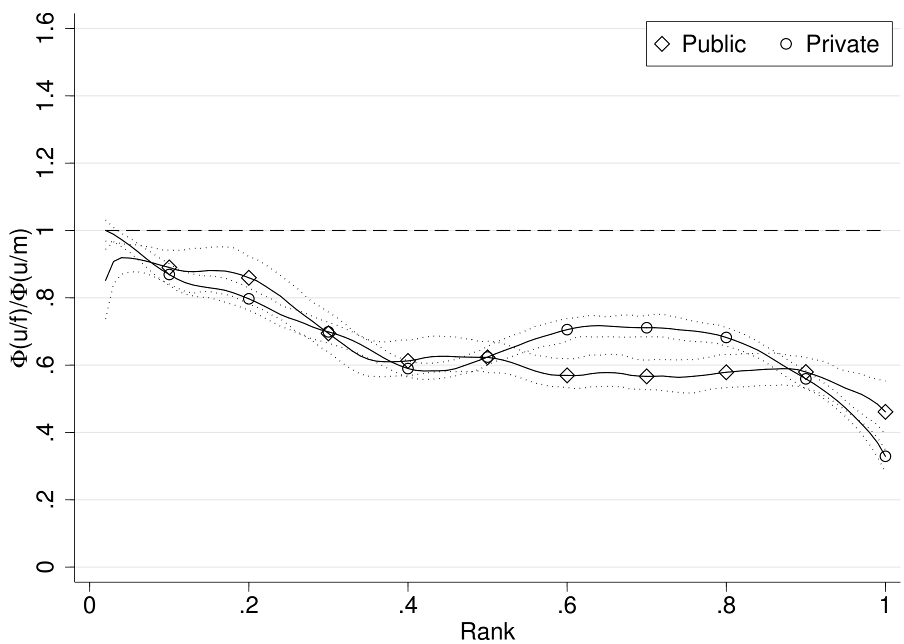
Note: Densities are computed for the logarithm of daily wage.

Figure 2: Gender quantile gap as a function of rank in the public and private sectors



Note: The gender quantile gap is the estimated coefficient of a female dummy introduced in quantile regressions of the logarithm of daily wage evaluated at each centile by sector. These regressions also include as controls the category dummies for all the other individual characteristics, ie. age, diploma, number of children, part-time experience, work interruption, firm tenure and being located in the Paris region. Confidence intervals at the 5% level are reported in dotted lines.

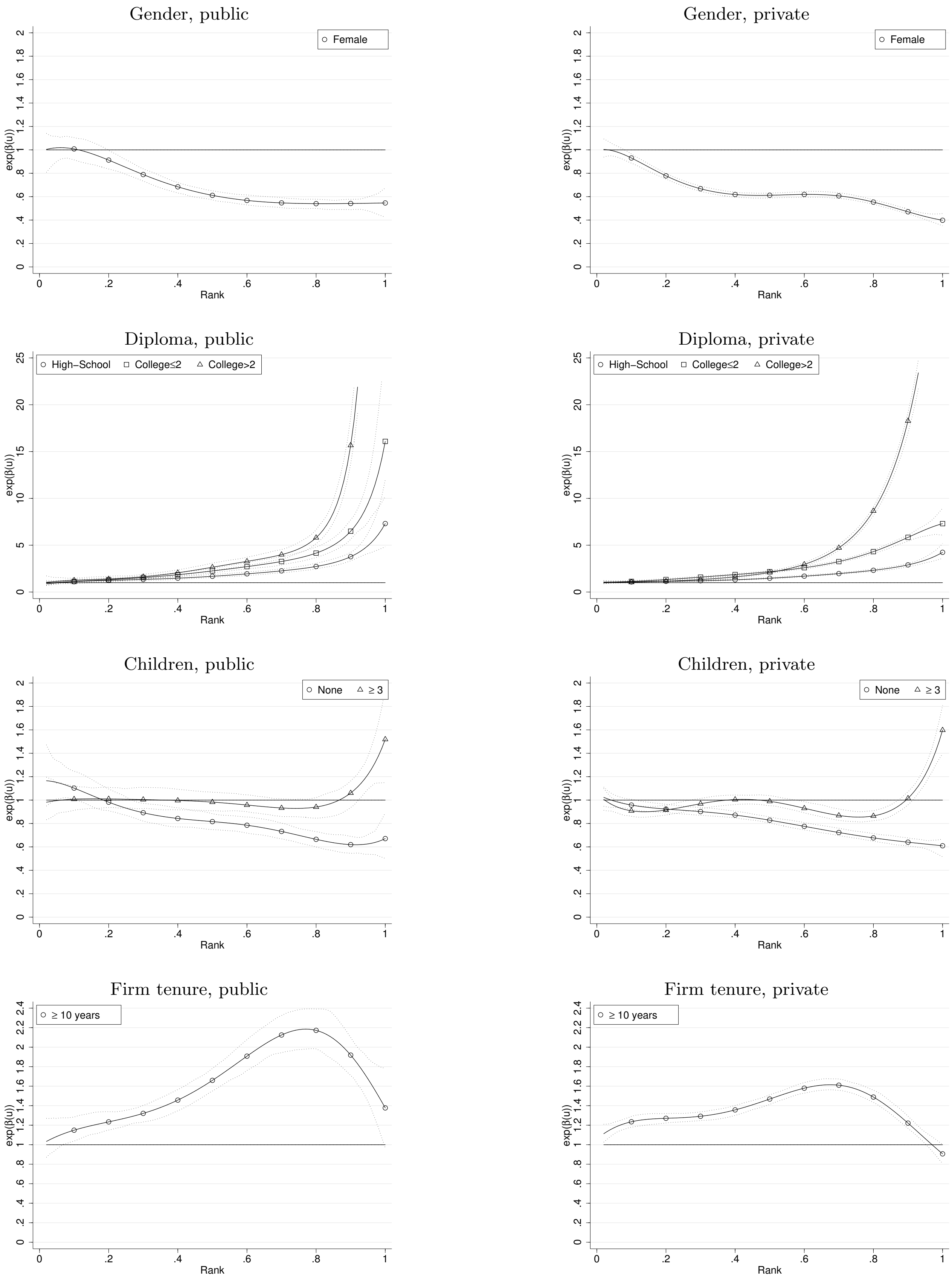
Figure 3: Gender probability ratio of getting a given job position in the public and private sectors



Note: Confidence intervals at the 5% level obtained by bootstrap using 100 replications are reported in dotted lines.

Figure 4: Exponentiated effects of category dummies on individual values

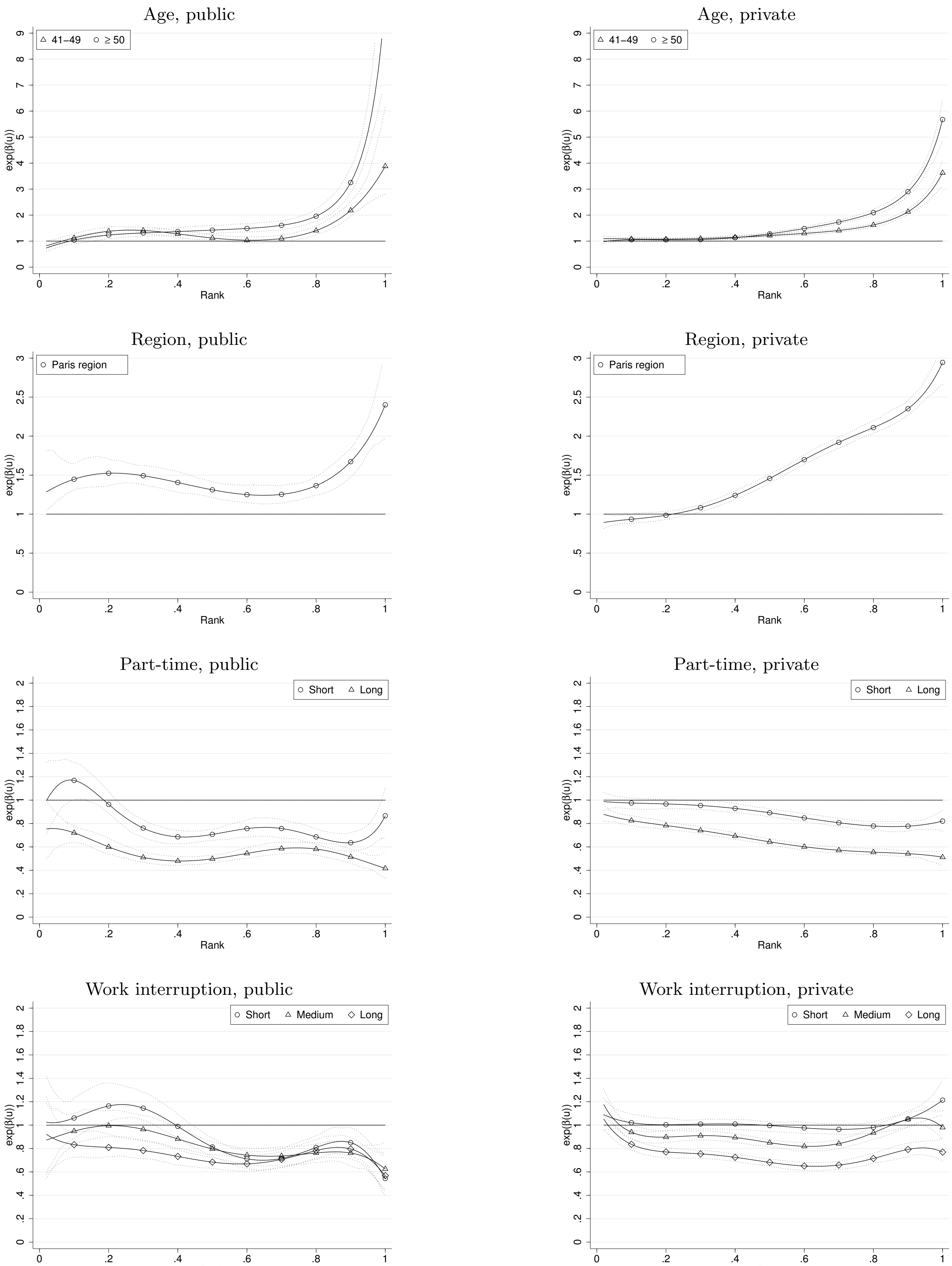
for each individual characteristic



Note: The graph title gives the individual characteristic for which the exponentiated coefficients of category dummies are graphed, as well as the sector. Reference categories for individual characteristics are “Male” for gender, “<High-School” for diploma, “1 or 2” for children, “<10 years” for firm tenure.

Figure 4 (Cont.): Exponentiated effects of category dummies on individual values

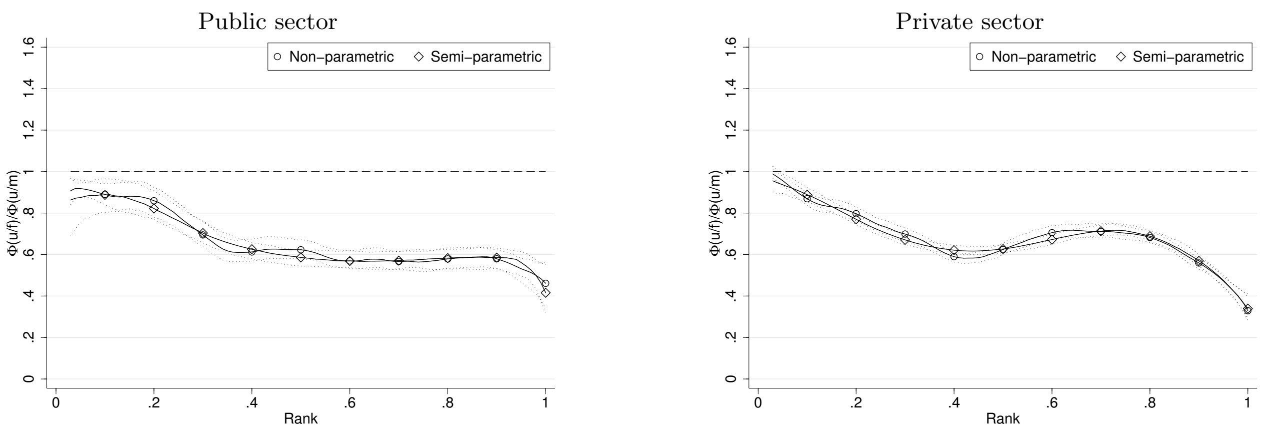
for each individual characteristic



Note: The graph title gives the individual characteristic for which the exponentiated coefficients of category dummies are graphed, as well as the sector. Reference categories for individual characteristics are "31-39" for age, "Paris region" for region, "None" for part-time, "None" for work interruption.

Figure 5: Non-parametric and semi-parametric estimators of

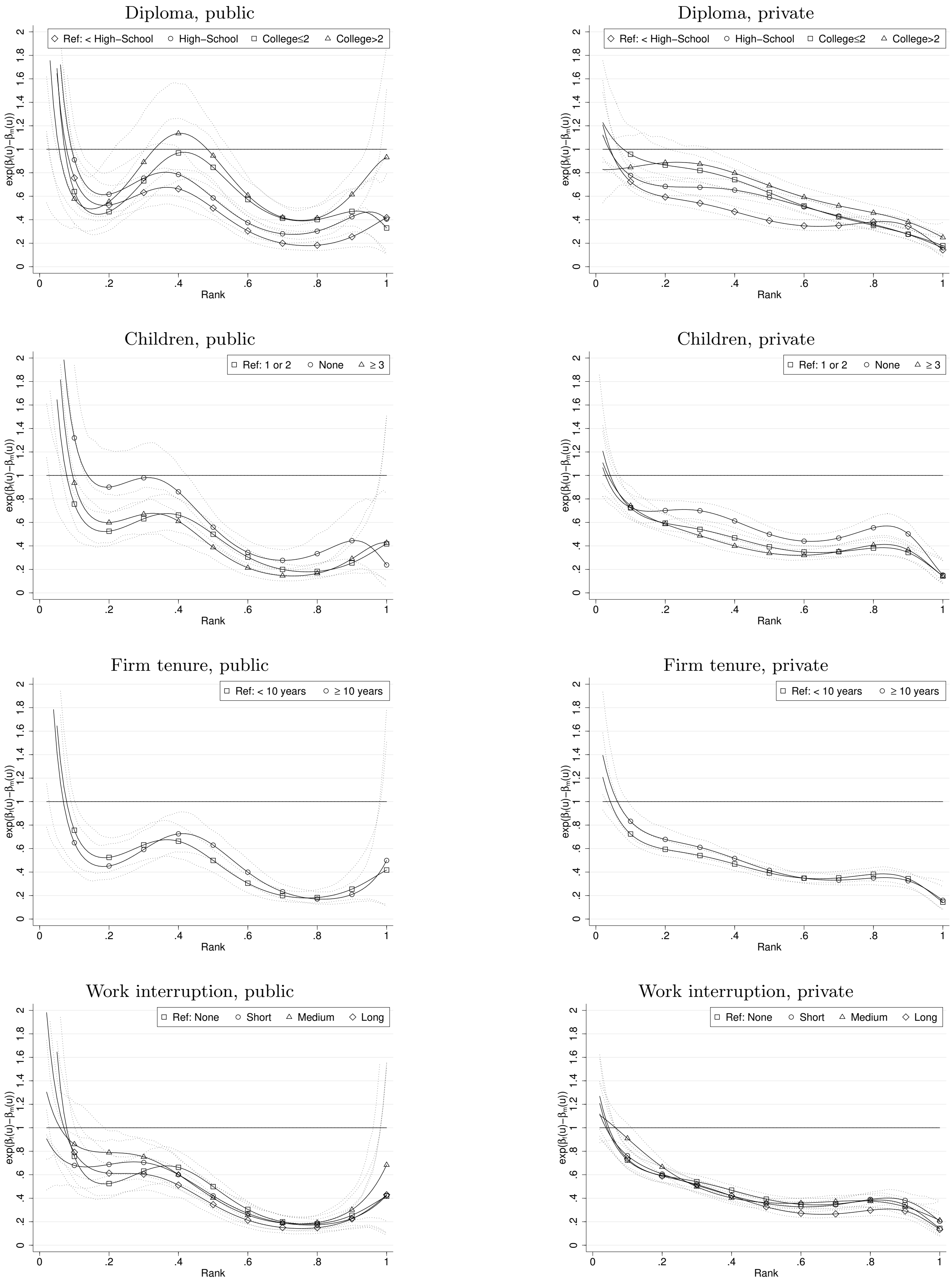
the gender probability ratio of getting a given job position



Note: The non-parametric estimator is obtained by applying the empirical strategy proposed by Gobillon, Meurs and Roux (2015). The semi-parametric estimator is obtained by applying the empirical strategy proposed in the current paper. The confidence interval at the 5% level obtained by bootstrap using 100 replications is reported in dotted lines.

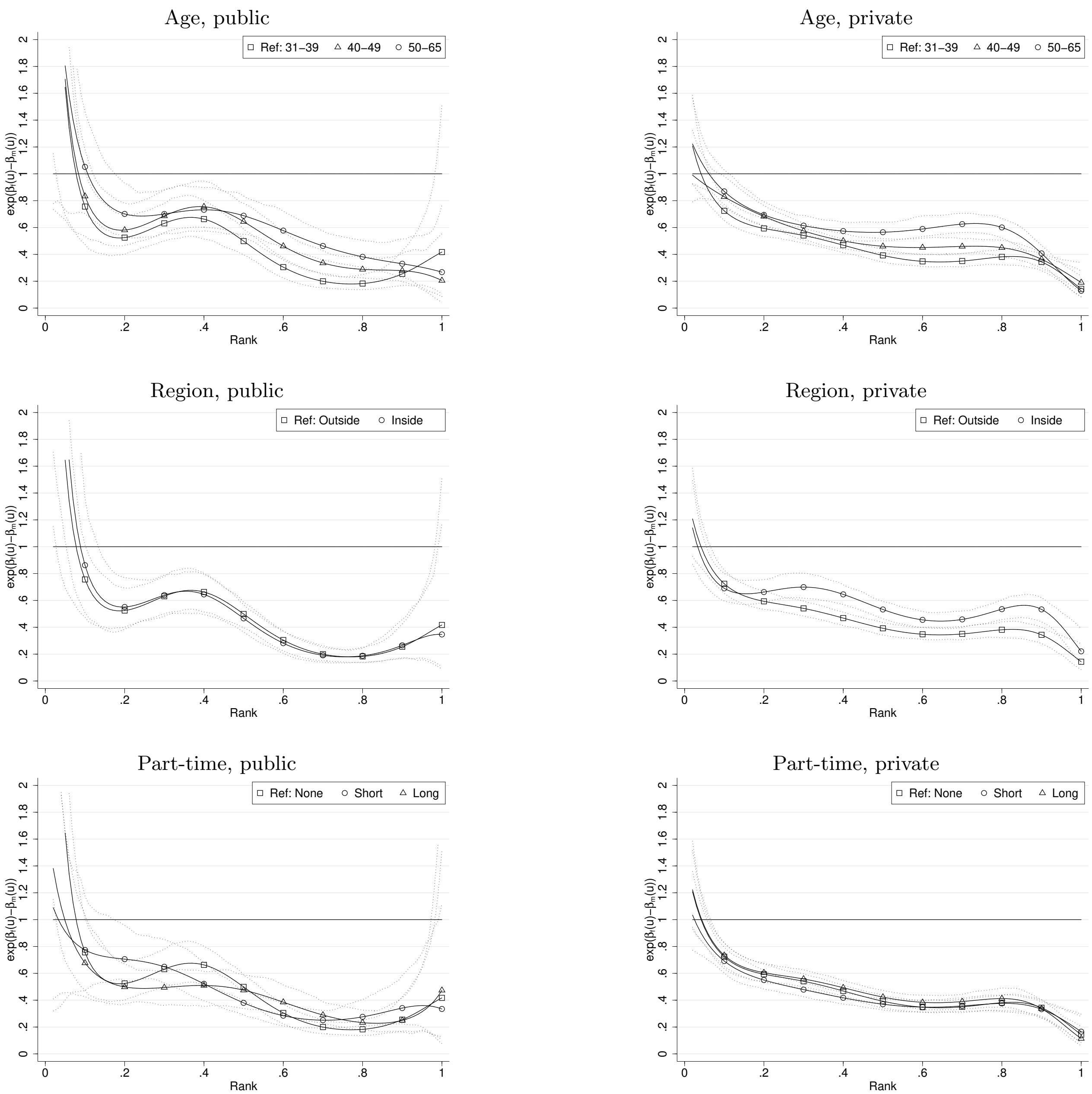
Figure 6: Exponentiated gender difference in the effects of category dummies on

the gender probability ratio of getting a given job position for each individual characteristic



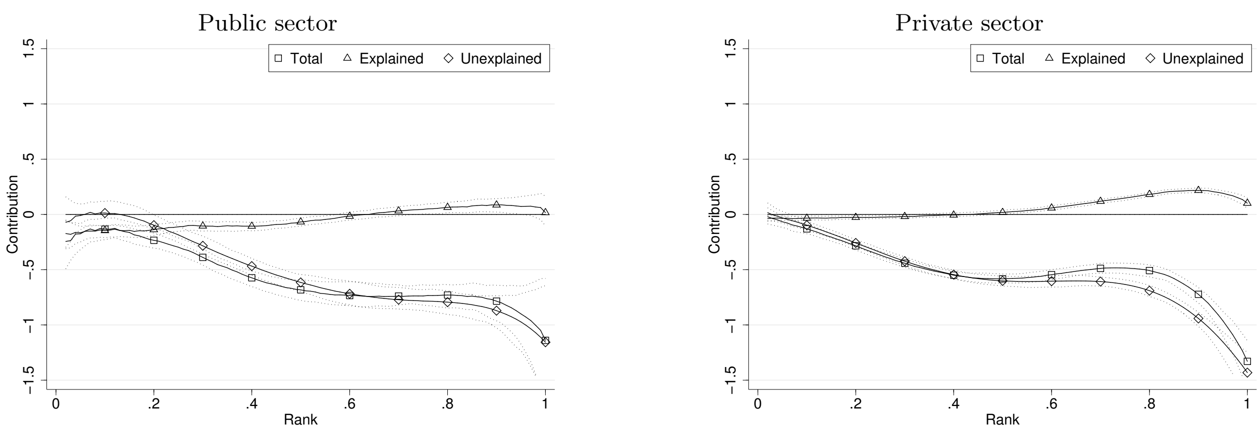
Note: The graph title gives the individual characteristic for which the exponentiated gender differences in the effects of category dummies are graphed, as well as the sector. The category corresponding to the reference is mentioned and the corresponding curve is the same across all graphs for a given sector. Indeed, this curve represents the gender probability ratio of getting a given job position as a function of rank for a worker whose values of all individual characteristics are fixed to the reference.

Figure 6 (Cont.): Exponentiated gender difference in the effects of category dummies on the gender probability ratio of getting a given job position for each individual characteristic



Note: The graph title gives the individual characteristic for which the exponentiated gender differences in the effects of category dummies are graphed, as well as the sector. The category corresponding to the reference is mentioned and the corresponding curve is the same across all graphs for a given sector. Indeed, this curve represents the gender probability ratio of getting a given job position as a function of rank for a worker whose values of all individual characteristics are fixed to the reference.

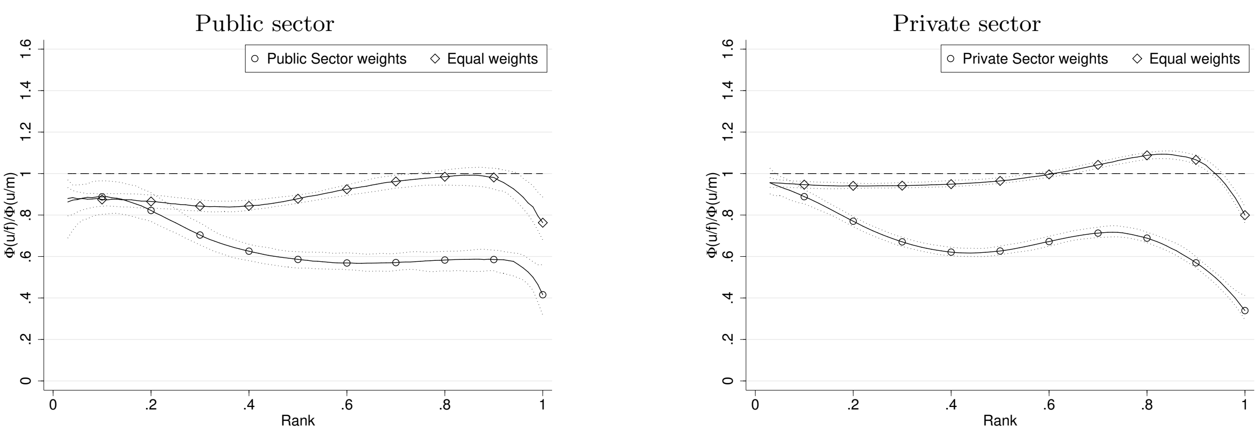
Figure 7: Oaxaca decomposition of the gender probability ratio of getting a given job position



Note: "Total": gender difference in the logarithm of the average probability of getting a given job position; "Explained": part of "Total" that can be attributed to the gender difference in observable characteristics valued using the estimated coefficients obtained for the whole population; "Unexplained": part of "Total" that can be attributed to the deviation of gender coefficients of observable characteristics from the ones of the whole population. Note that "Total" is not exactly equal to the sum of "Explained" and "Unexplained" since it also involves a residual term due to the non-linearity of the logarithm function.

Figure 8: Counterfactual gender probability ratio of getting a given job position

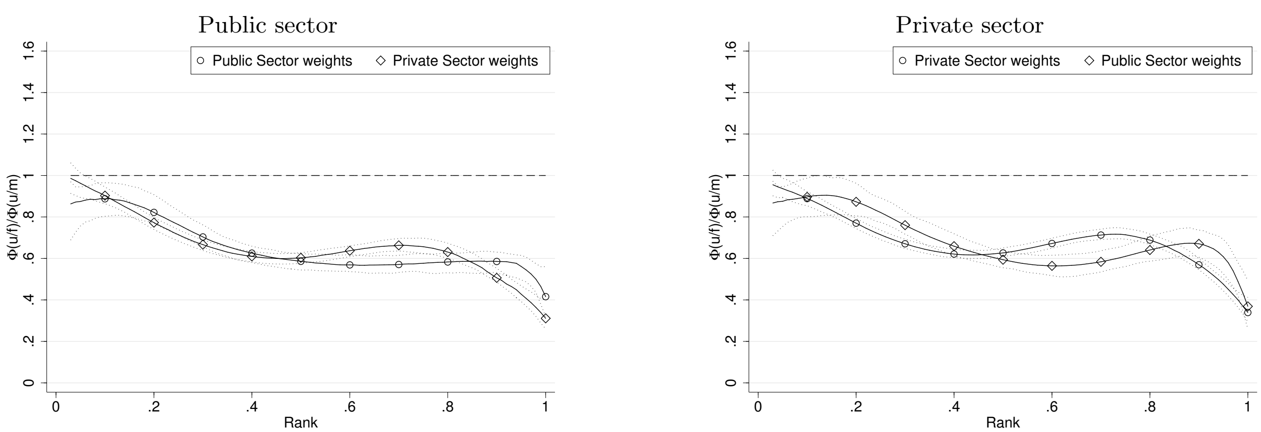
when conditional individual weights are equal for the two genders



Note: The curve corresponding to "Equal weights" is obtained by fixing conditional individual weights to the same values for the two genders, these common values being obtained by estimating parameters for the whole population. The curve corresponding to "Private Sector weights" (resp. "Public Sector weights") is obtained by fixing conditional individual weights to those of the private (resp. public) sector. The confidence interval at the 5% level obtained by bootstrap using 100 replications is reported in dotted lines.

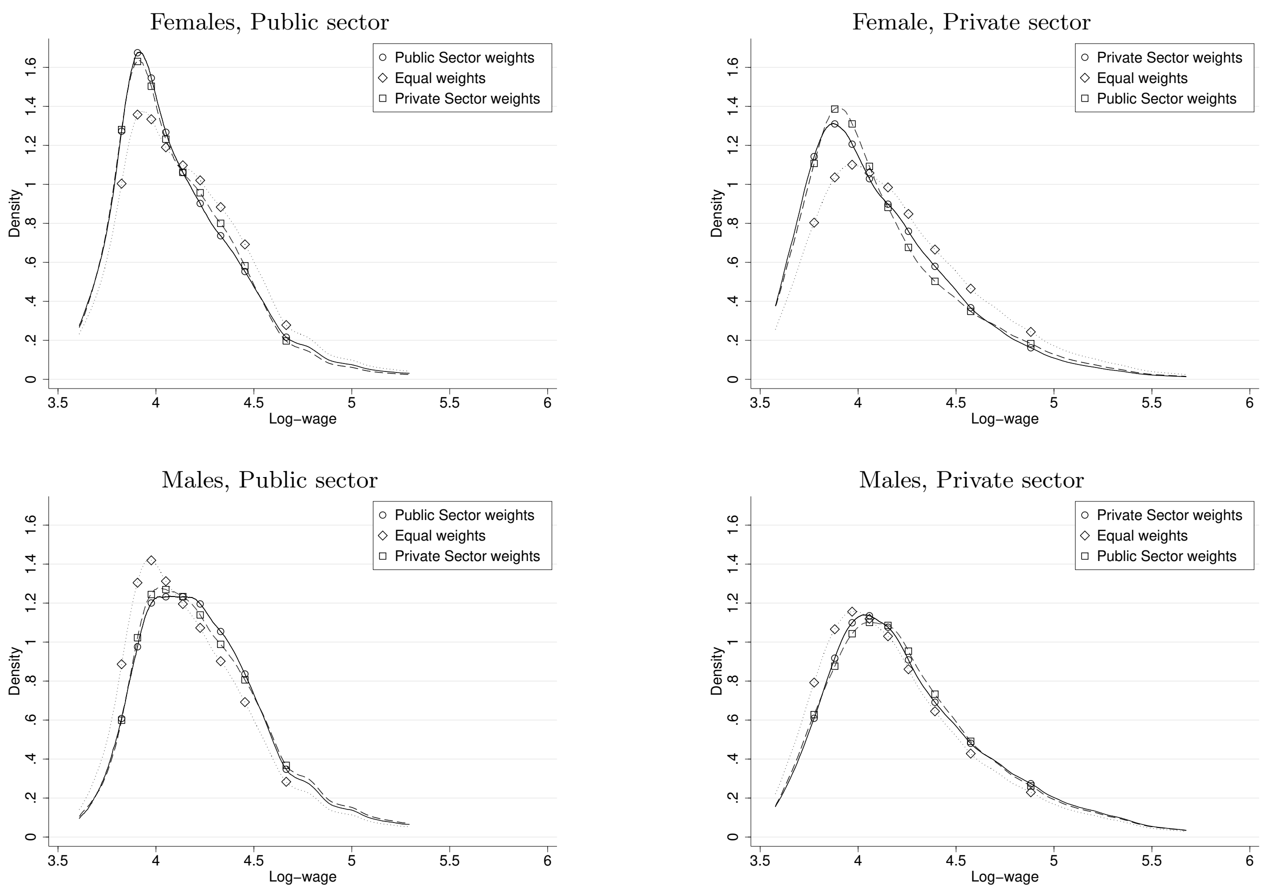
Figure 9: Counterfactual gender probability ratio of getting a given job position

when conditional individual weights are those in the other sector



Note: The curve corresponding to "Private Sector weights" (resp. "Public Sector weights") is obtained by fixing conditional individual weights to those of the private (resp. public) sector. The confidence interval at the 5% level obtained by bootstrap using 100 replications is reported in dotted lines.

Figure 10: Counterfactual log-wage densities in different scenarios



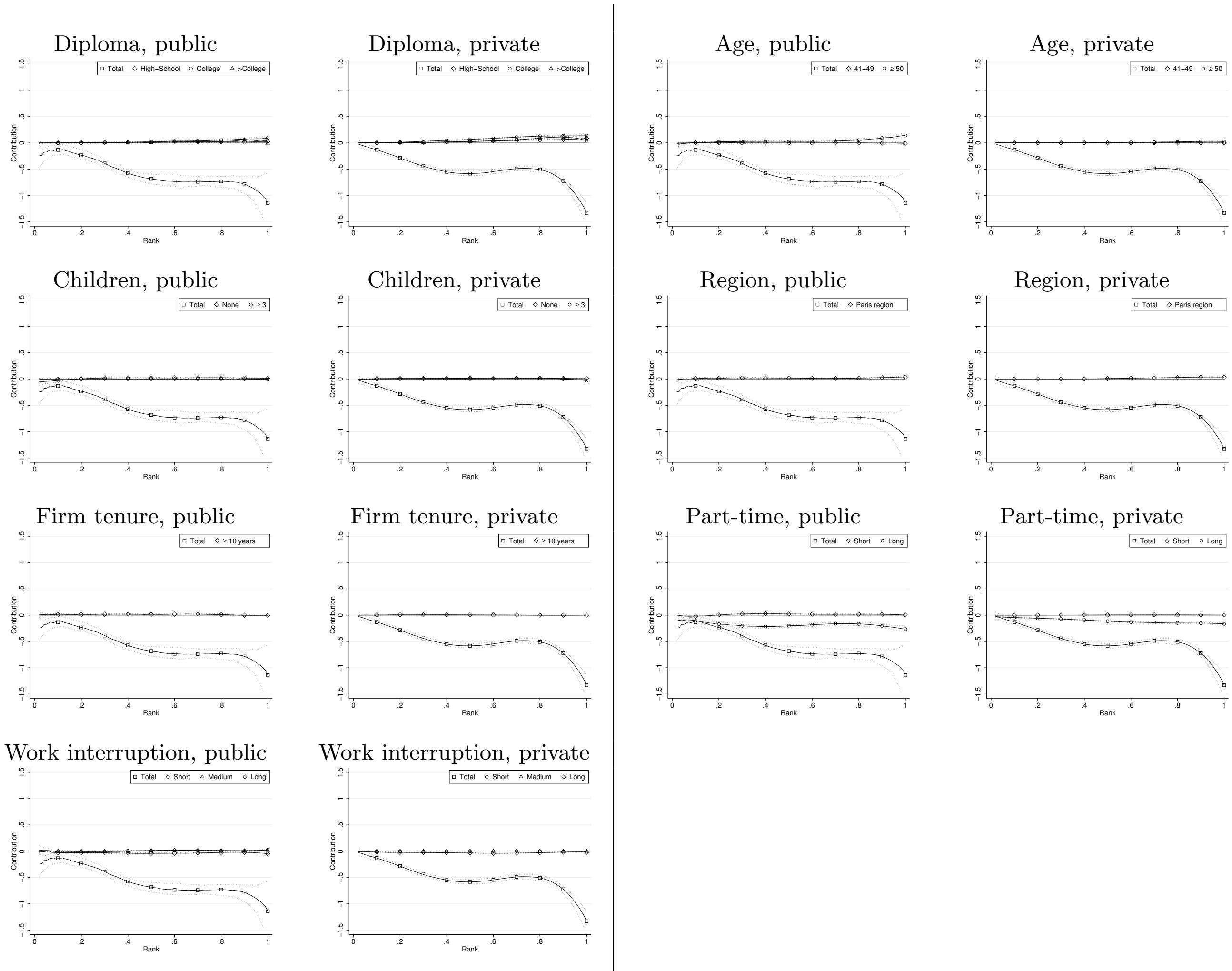
Note: Densities are computed using the logarithm of daily wages generated by the model when reassigning workers to job positions. "Private Sector weights": reassignment using conditional individual weights computed for the private sector; "Public Sector weights": reassignment using conditional individual weights computed for the public sector; "Equal weights": reassignment using the same conditional individual weights for the two genders, these common weights being obtained by estimating parameters for the whole population.

Table D.1: Observed and counterfactual log-wage gaps obtained in different scenarios, hourly wages, part-time workers included

	Public sector				Private sector			
	Observed	Public weights	Equal weights	Private weights	Observed	Private weights	Equal weights	Public weights
Mean								
Males	2.6197	2.6197	2.5620	2.6358	2.6345	2.6339	2.5680	2.6248
Females	2.4811	2.4811	2.5230	2.4694	2.4536	2.4545	2.5446	2.4669
F-M Gap	-0.1385	-0.1386	-0.0391	-0.1664	-0.1809	-0.1794	-0.0233	-0.1579
		(0.0059)	(0.0042)	(0.0060)		(0.0047)	(0.0035)	(0.0070)
Standard deviation								
Males	0.3603	0.3570	0.3487	0.3673	0.4560	0.4545	0.4364	0.4495
Females	0.3132	0.3159	0.3339	0.3014	0.3769	0.3797	0.4296	0.3931
F-M Gap	-0.0471	-0.0410	-0.0149	-0.0659	-0.0791	-0.0748	-0.0067	-0.0564
		(0.0064)	(0.0038)	(0.0060)		(0.0047)	(0.0033)	(0.0076)
First decile								
Males	2.2263	2.2299	2.1949	2.2442	2.1529	2.1532	2.1106	2.1391
Females	2.1497	2.1493	2.1603	2.1435	2.0513	2.0526	2.0848	2.0667
F-M Gap	-0.0767	-0.0807	-0.0346	-0.1007	-0.1016	-0.1005	-0.0259	-0.0724
		(0.0061)	(0.0036)	(0.0047)		(0.0030)	(0.0027)	(0.0068)
First quartile								
Males	2.3621	2.3621	2.3103	2.3655	2.3042	2.3043	2.2485	2.3018
Females	2.2571	2.2569	2.2782	2.2534	2.1700	2.1711	2.2251	2.1802
F-M Gap	-0.1050	-0.1053	-0.0321	-0.1122	-0.1342	-0.1333	-0.0234	-0.1216
		(0.0053)	(0.0028)	(0.0050)		(0.0039)	(0.0027)	(0.0060)
Median								
Males	2.5619	2.5614	2.4919	2.5661	2.5265	2.5295	2 .4662	2.5321
Females	2.4065	2.4046	2.4567	2.4021	2.3595	2.3614	2.4449	2.3616
F-M Gap	-0.1553	-0.1567	-0.0352	-0.1640	-0.1670	-0.1681	-0.0213	-0.1705
		(0.0068)	(0.0056)	(0.0060)		(0.0056)	(0.0036)	(0.0078)
Last quartile								
Males	2.8053	2.8066	2.7437	2.8304	2.8804	2.8825	2.7911	2.8683
Females	2.6530	2.6529	2.7136	2.6465	2.6563	2.6570	2.7751	2.6592
F-M Gap	-0.1523	-0.1537	-0.0300	-0.1839	-0.2241	-0.2256	-0.0160	-0.2091
		(0.0097)	(0.0061)	(0.0093)		(0.0090)	(0.0061)	(0.0133)
Last decile								
Males	3.0905	3.0873	3.0104	3.1337	3.2511	3.2504	3.1631	3.2265
Females	2.9012	2.9015	2.9531	2.8659	2.9916	2.9968	3.1371	3.0369
F-M Gap	-0.1893	-0.1857	-0.0573	-0.2678	-0.2595	-0.2535	-0.0260	-0.1896
		(0.0177)	(0.0109)	(0.0141)		(0.0104)	(0.0072)	(0.0163)

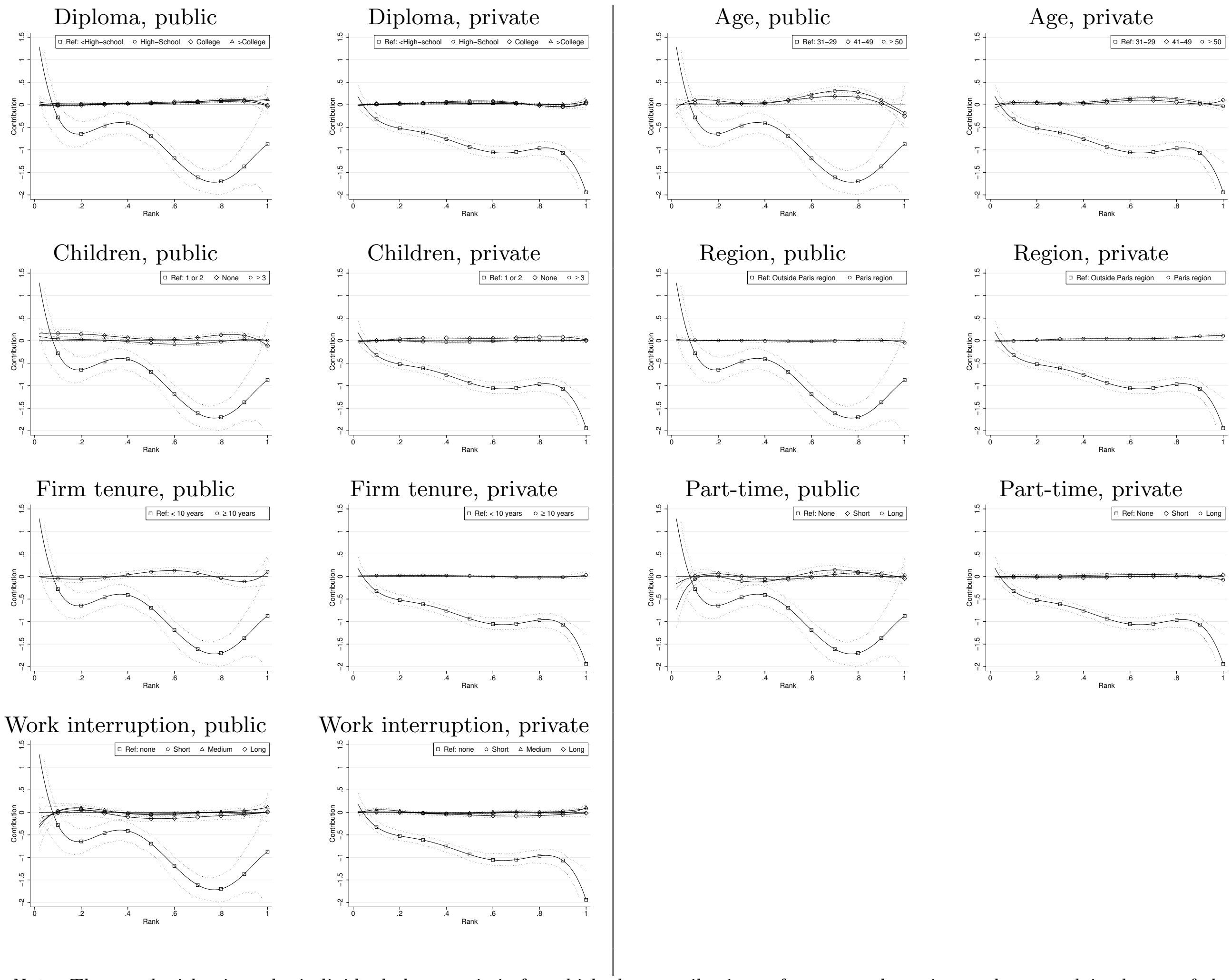
Note: Statistics are computed on the logarithm of hourly wage. Column headings mention either that statistics are computed directly from the data (label “Observed”) or that they are derived from counterfactuals using the conditional individual weights of the public sector, the private sector, or the same conditional individual weights for the two genders in the sector that is considered (respective labels “Private weights”, “Public weights”, ‘’Equal weights”). For counterfactual exercises, the standard deviation of the gender gap obtained by bootstrap is reported in parentheses.

Figure D.1: Decomposition of the explained part of the gender probability ratio of getting a given job into the contribution of every category dummy for each individual characteristic



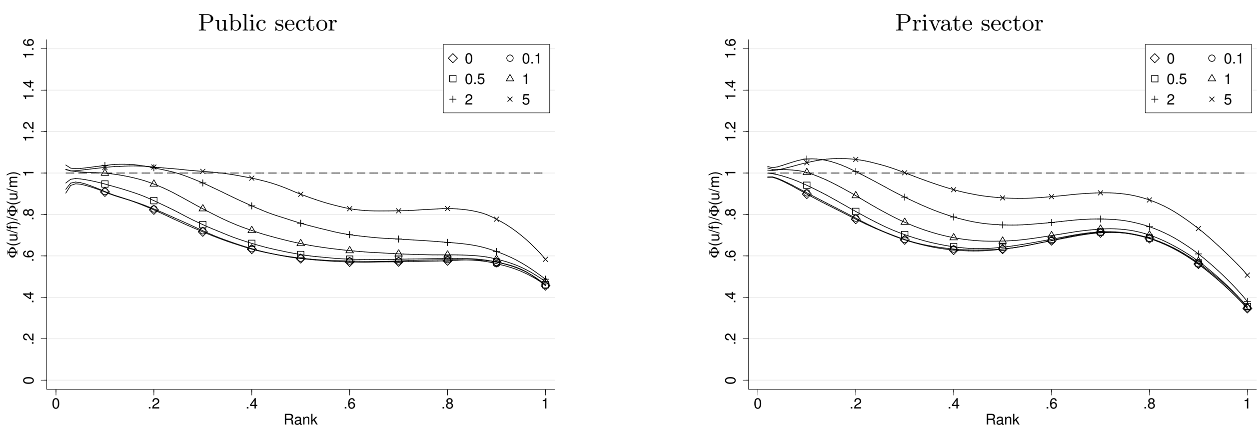
Note: The graph title gives the individual characteristic for which the contributions of category dummies to the explained part of the gender probability ratio are graphed, as well as the sector. "Total": gender difference in the logarithm of the average probability of getting a given job; other label in the legend: category of the individual characteristic for which the contribution is graphed. This contribution is the gender difference in the average of the category dummy valued using the estimated coefficient for the whole population (see Appendix E.1 for more details).

Figure D.2: Decomposition of the unexplained part into the contribution of every category dummy for each individual characteristic



Note: The graph title gives the individual characteristic for which the contributions of category dummies to the unexplained part of the gender probability ratio are graphed, as well as the sector. Label in the legend: category of the individual characteristic for which the contribution is graphed. This contribution is the sum of the gender averages of the category dummy valued using the difference between the gender coefficient of the category dummy and the coefficient for the whole population (see Appendix E.1 for more details).

Figure D.3: Counterfactual gender probability ratio of getting a given job position when adding unobserved individual heterogeneity terms to individual values and reassigning workers



Note: We report the non-parametric estimator of the gender probability ratio of getting a given job position after reassigning workers to positions according to conditional individual weights obtained from our semi-parametric approach once unobserved individual heterogeneity terms have been added to individual values. These terms are drawn in a centered normal law with variance equal to the gender-specific variance of the total effect of observed individual characteristics multiplied by a parameter which is made to vary and is reported in the legend of the graphs (see text for more details).