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

Laurent Gobillon, Dominique Meurs, Sébastien Roux

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2017 – 13

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

Laurent Gobillon
Dominique Meurs
Sébastien Roux

JEL Codes: C51, J31, J45
Keywords: assignment, distributions, counterfactuals, wages, gender, public sector

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE – UNIVERSITE PARIS 1

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

Laurent Gobillon* Dominique Meurs† Sébastien Roux‡

April 16, 2017§

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 outcomes across groups and we show how it can motivate decomposition and counterfactual exercises. It constitutes an alternative to more descriptive methods such as Oaxaca decompositions and quantile counterfactual approaches. In an application, we study gender disparities in the public and private sectors with a French exhaustive administrative dataset. 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. The gender wage gap in the public sector is 13.3 points and it increases by only 0.7 percentage points when workers are assigned to job positions according to the rules of the private sector. Nevertheless, the gender gap at the last decile in the public sector increases by as much as 3.6 percentage points when using the assignment rules of the private sector.

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

JEL Classification: C51, J31, J45

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

†University of Paris Ouest Nanterre (EconomiX) 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. 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, Roland Rathelot, Jean-Marc Robin, Christian Schluter, Antoine Terracol as well as seminar and conference participants at Bristol, Hannover, Timbergen, CREST, Paris 1, Rennes, Strasbourg, World Congress 2015 of the Econometric Society in Montreal, SOLE 2016 in Seattle, AFSE 2014 in Lyon, TEPP 2015 in Aussois, AMSE-Banque de France 2015 conference in Paris, AMSE 2014 Workshop on Gender and Ethnic Differentials in Market Outcomes in Aix-en-Provence, and OFCE. 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 allocated to positions along a hierarchy depending on their characteristics. This model can be brought to the data to study differences in outcomes 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 the outcome distribution. All individuals consider that the position at the highest rank is the most attractive but some of them find it too constraining to apply. Observable characteristics, including the group, can influence both the propensity to apply and applicants' chances of being selected for the position. Individuals not selected for the position consider the position at the second highest rank, and so on, until all positions are filled. Overall, this model can be seen as a way of assigning individuals to positions according to their characteristics. It is an extension of Gobillon, Meurs and Roux (2015) in which chances of being selected vary across groups but individual heterogeneity is not taken into account. The effects of observable characteristics on the propensity to get positions along the hierarchy can be estimated. It is then possible to make decompositions of the difference in outcomes between groups and to construct counterfactuals for each group by fixing parameters underlying the propensity to get positions to alternative values.

As our approach is based on a theoretical framework, it provides an alternative to more descriptive methods involving 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). The latter approaches use conditional outcome distributions from other samples as counterfactuals, such that the overall outcome distribution depends on the counterfactual. Equilibrium effects leading to a reassignment of individuals across positions are not modeled. By contrast, in our setting, counterfactuals are the result of an equilibrium when modifying structural parameters. They are generated by changing the way individuals are assigned to positions and counterfactual assignments affect the outcome distribution conditional on individual characteristics but not the overall outcome distribution which is held fixed. Our approach also differs from traditional ones in the definition of rank when studying quantile differences between groups. Indeed, quantile differences between two groups involve two different ranks which are the ranks in the conditional outcome distributions of the two groups. Our model rathers considers the rank in the outcome distribution of positions, which is a common index for all groups.[1]

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, 2015). 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 particularity of our logit specification for the probability of a position being filled is that the coefficients of explanatory variables are group-specific polynomial series of the rank. We thus allow for group-specific probabilities of being selected that depend on observable characteristics and the rank.[2] Counterfactuals of outcome 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), as 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 recruitments 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 assigment 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

[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.

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

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. 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 average gender 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 explain how counterfactuals can be constructed from an assigment model by changing the propensities to get positions. We detail in Section 3 how the model can be estimated and empirical counterparts of counterfactuals can be obtained. Section 4 presents our data and discusses the results on the estimated parameters underlying the assignment rules and the counterfactuals. Finally, Section 5 concludes.

2 Theoretical framework

We propose an assignment model in which individuals are allocated to positions along a hierarchy depending on their characteristics. 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.

2.1 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 $\{X^k\}_{k=1,...,K}$ 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 such that they can be ranked in a hierarchy.[3] We suppose that two positions cannot be associated with the same rank. 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. 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.[4]

More formally, the process leading to the choice of an applicant can be described in the following way. For a position of rank u, 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 value derived from an applicant i, denoted $V_i(u)$ and labelled "Individual value", is supposed to take the form:

$$V_i(u) = \ln\varphi(u\,|X_i, j(i)\,) + \varepsilon_i(u) \tag{1}$$

[3]For instance, if an outcome is attached to each position and outcomes associated to positions are all different, positions can be ranked in descending order of outcome value. But this is not necessary and differences in allocations and outcomes across groups can also be studied when positions are ranked according to the values of any other variable such that values associated to positions are all different.

[4]The framework needs to be extended when several positions have the same rank in the hierarchy as available individuals are interested in all the positions to the same extent at such ranks. In particular, assumptions must be made on whether positions at those ranks are filled simultaneously or sequentially.

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. This 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:

$$\Omega(u) = \left\{ i \text{ applying for rank-}u \text{ position } \middle| \text{for all } \widetilde{u} > u,\ i \text{ not applying for rank-}\widetilde{u} \text{ position or } V_i(\widetilde{u}) < \max_{k\in\Omega(\widetilde{u})} V_k(\widetilde{u}) \right\} \tag{2}$$

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. Second, 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).[5] This assumption ensures that for any given position, the probability of selecting an individual follows a logit model. Under this assumption, the probability that the individual chosen for the position of rank u is of gender j and has characteristics X verifies:

$$P(j(u) = j, X(u) = X) = n(u|X,j)\,\phi(u|X,j) \tag{3}$$

with:

$$\phi(u|X,j) = \frac{\mu(u|X,j)}{\int n(u|X,f)\,\mu(u|X,f)\,dX + \int n(u|X,m)\,\mu(u|X,m)\,dX} \tag{4}$$

where $\mu(u|X,j) = \upsilon(u|X,j)\,\varphi(u|X,j)$ captures both the propensity to apply and the fixed component entering the value of individuals with characteristics X and gender j. The denominator on the right-hand side 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, $\mu(u|X,j)$, 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 $\phi(u|X,j)$ as the conditional probability of getting the position of rank u, and to $\phi(u|X,m)/\phi(u|X,f) = \mu(u|X,m)/\mu(u|X,f)$ as the gender conditional probability ratio of getting the position

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

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 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:

$$n\left(u-du\left|X,j\right.\right)=n\left(u\left|X,j\right.\right)-n\left(u\left|X,j\right.\right)\phi\left(u\left|X,j\right.\right)du \tag{5}$$

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

$$n'\left(u\left|X,j\right.\right)=\phi\left(u\left|X,j\right.\right)n\left(u\left|X,j\right.\right) \tag{6}$$

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 (6) for all X and j, where $\phi\left(u\left|X,j\right.\right)$ verifies (4), has a unique solution.

2.2 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:

$$\phi\left(u\left|j\right.\right)=\int p\left(u\left|X,j\right.\right)\phi\left(u\left|X,j\right.\right)dX \tag{7}$$

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 $\phi\left(u\left|f\right.\right)/\phi\left(u\left|m\right.\right)$.

Suppose that we are able to construct estimators of conditional individual weights at any given rank, $\mu\left(u\left|X,j\right.\right)$.

[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.

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 of getting the position of rank u derived from (7) and rearranging the terms, we get:

$$\begin{aligned}\log\left[\phi(u|f)/\phi(u|m)\right] &= \int\left[p(u|X,f)-p(u|X,m)\right]\log\mu^r(u|X)\,dX \\ &+\int\left[\log\mu(u|X,f)-\log\mu^r(u|X)\right]p(u|X,f)\,dX \\ &-\int\left[\log\mu(u|X,m)-\log\mu^r(u|X)\right]p(u|X,m)\,dX+r(u)\end{aligned} \tag{8}$$

where:

$$\begin{aligned}r(u) &= \left[\int p(u|X,f)\log\mu(u|X,f)\,dX-\log\left[\int p(u|X,f)\,\mu(u|X,f)\,dX\right]\right] \\ &-\left[\int p(u|X,m)\log\mu(u|X,m)\,dX-\log\left[\int p(u|X,m)\,\mu(u|X,m)\,dX\right]\right]\end{aligned} \tag{9}$$

The first right-hand side term in (8) 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 exercices that involve a reassignment of individuals when changing conditional individual weights.

2.3 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 (6) where the conditional probability of getting a position has been replaced by its expression (4) and conditional individual weights by their counterfactuals. These counterfactual measures verify:

$$n^{*\prime}(u\,|X,j\,)=\frac{n^{*}(u\,|X,j\,)\,\mu^{*}(u\,|X,j\,)}{\int n^{*}(u\,|X,f\,)\,\mu^{*}(u\,|X,f\,)\,dX+\int n^{*}(u\,|X,m\,)\,\mu^{*}(u\,|X,m\,)\,dX} \tag{10}$$

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

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 (7) 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 $p^{*}(u\,|X,j\,)=n^{*}(u\,|X,j\,)\,/n^{*}(u\,|j\,)$ with $n^{*}(u\,|j\,)=\int n^{*}(u\,|X,j\,)\,dX$. The counterfactual gender-j probability of getting the position is given by:

$$\phi^{*}(u\,|j\,)=\int p^{*}(u\,|X,j\,)\,\mu^{*}(u\,|X,j\,)\,dX \tag{11}$$

Now consider the specific case in which an outcome is attached to each position and positions are ranked according to the outcome. In the counterfactual situation, individuals are reallocated across positions according to the alternative assignment rules while holding fixed the outcome distribution of positions. This yields a change in the gender-specific distributions of outcomes. The counterfactual of gender-j outcome cumulative obtained when workers have the alternative individual conditional weights verifies:

$$F_{j}^{*}(w)=\frac{1}{n\,(j)}\int n^{*}(F\,(w)\,|X,j\,)\,dX \tag{12}$$

where $F\,(\cdot)$ is the outcome cumulative. The counterfactual gender-j cumulative is simply the proportion of gender-j individuals still available for positions below a given outcome such that the set of gender-j available individuals was determined according to the alternative assignment rule. Note that the outcome 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 (12) gives the counterfactual of gender-j outcome density:

$$f_{j}^{*}(w)=\frac{f\,(w)}{n\,(j)}\int n^{*\prime}(F\,(w)\,|X,j\,)\,dX \tag{13}$$

where $f\,(\cdot)$ is the outcome density. The derivative of the counterfactual measure of available workers can be replaced by its expression given by (10) to get the following expression:[7]

$$f_j^*(w) = \frac{f(w)}{n(j)} \frac{n^*(F(w)|j)\,\phi^*(F(w)|j)}{n^*(F(w)|f)\,\phi^*(F(w)|f) + n^*(F(w)|m)\,\phi^*(F(w)|m)} \tag{14}$$

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

3 Empirical strategy

3.1 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:

$$\mu(u|X,j) = \exp\left[X\beta_j(u)\right] \tag{15}$$

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 propension 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 (4) 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 (1) 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 (13) using (10) 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 (11), 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 (13).

[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:

$$\begin{aligned} L &= P\left(u_{i_1}=u^1, u_{i_2}=u^2, ..., u_{i_N}=u^N \middle| \overrightarrow{X}_N, \overrightarrow{j}_N\right) && (16)\\ &= P\left(u_{i_N}=u^N \middle| \overrightarrow{X}_N, \overrightarrow{j}_N\right)\prod_{k=1}^{N-1} P\left(u_{i_k}=u^k \middle| u_{i_{k+1}}=u^{k+1}, ..., u_{i_N}=u^N, \overrightarrow{X}_N, \overrightarrow{j}_N\right) && (17)\\ &= \prod_{k=1}^{N} P\left(u_{i_k}=u^k \middle| \{i_1,...,i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right) && (18)\end{aligned}$$

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 (18), $P\left(u_{i_k}=u^k \middle| \{i_1,...,i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right)$ is the empirical counterpart of $\phi\left(u_{i_k}\left|X_{i_k}, j\left(i_k\right)\right.\right)$ and it verifies:

$$P\left(u_{i_k}=u^k \middle| \{i_1,...,i_k\}\in\Omega\left(u^k\right), \overrightarrow{X}_k, \overrightarrow{j}_k\right) = \frac{\mu\left(u_{i_k}\left|X_{i_k}, j\left(i_k\right)\right.\right)}{\sum\limits_{\ell\leqslant k}\mu\left(u_{i_k}\left|X_{i_\ell}, j\left(i_\ell\right)\right.\right)} = \frac{\exp\left[X_{i_k}\beta_{j(i_k)}\left(u_{i_k}\right)\right]}{\sum\limits_{\ell\leqslant k}\exp\left[X_{i_\ell}\beta_{j(i_\ell)}\left(u_{i_k}\right)\right]} \quad (19)$$

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, and the asymptotic distribution of estimated parameters has long been established (see Andersen and Gill, 1982).

3.2 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 (8). Assume that benchmark conditional individual weights $\mu^r\left(u\left|X\right.\right)$ are also of the form (15) with estimable polynomial coefficients $\beta^r(u)$. In that case, the decomposition simplifies to:

$$\begin{aligned} \log\left[\phi\left(u\left|f\right.\right)/\phi\left(u\left|m\right.\right)\right] &= \left[E\left(X\left|f,u\right.\right)-E\left(X\left|m,u\right.\right)\right]\beta^r(u)\\ &\quad +E\left(X\left|f,u\right.\right)\left[\beta_f(u)-\beta^r(u)\right]-E\left(X\left|m,u\right.\right)\left[\beta_m(u)-\beta^r(u)\right]+r(u) \qquad (20)\end{aligned}$$

where $E(X|j,u)$ are the average characteristics of gender-j individuals available for a position at rank u.[10] The left-hand side term involves the gender probability ratio of getting a position of rank u that can be estimated non parametrically following Gobillon, Meurs and Roux (2015). Right-hand side terms can be obtained by replacing average characteristics by 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).

We now turn to the evaluation of the counterfactual gender-j probabilities of getting every position given by (11). For that purpose, we need to recover $n^*(u|X,j)$ when using the counterfactual conditional individual weights $\mu^*(u|X,j)$ which we suppose to be of the form (15) with polynomial coefficients $\beta_j^*(u)$. A direct method would consist in solving the system of non-linear differential equations given by (10). 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 (6) is:

$$N^*\left(u^k|X,j\right) = N^*\left(u^{k+1}|X,j\right) - D_{k+1}(X,j) \tag{21}$$

where $N^*(u|X,j)$ is the counterfactual number of gender-j individuals with characteristics X available at rank u, and $D_k(X,j)$ 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 (4). 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 Appendix B that $E\left[N^*\left(u^{\lfloor vN\rfloor+1}|X,j\right)\right]/N \underset{a.s.}{\rightarrow} n^*(v|X,j)$ for all $v \in]0,1[$ 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 $\exp\left[X\beta_j(u)\right]$ depends on the rank. We show that the proof of the original theorem can be generalized to the case where the

[10] 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:

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

which is the expression reported in the text. The expressions for the other right-hand side terms of (20) corresponding to the unexplained part can be established in the same way.

influence of explanatory variables varies across positions.

The expectation $E\left[N^*\left(u^{\lfloor vN \rfloor+1}\left|X,j\right.\right)/N\right]$ can be estimated using a simulation procedure that involves reassignments of individuals of the original sample to the positions based on the counterfactual assigment rules. Theoretical foundations of the simulation approach are detailed in Appendix C. 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:

$$P\left(u_i=u^k\right|\Omega^s\left(u^k\right),\overrightarrow{X}_k^s,\overrightarrow{j}_k^s\Big)=\frac{\exp\left[X_i\beta^*_{j(i)}\left(u^k\right)\right]}{\sum_{k\in\Omega^s(u^k)}\exp\left[X_k\beta^*_{j(k)}\left(u^k\right)\right]} \tag{22}$$

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 $N^s\left(u^k\left|X,j\right.\right)=Card\ \Omega_j^s\left(u^k,X\right)$ and an estimator of the expected counterfactual number of available individuals with given characteristics and gender is $\widehat{N}^*\left(u^k\left|X,j\right.\right)=\sum_{s=1}^S N^s\left(u^k\left|X,j\right.\right)/S$. We show in Appendix C that when $S\to+\infty$, we have for all $v\in(0,1)$, $\widehat{N}^*\left(u^{\lfloor vN \rfloor+1}\left|X,j\right.\right)\underset{a.s.}{\to}E\left[N^*\left(v\left|X,j\right.\right)\right]$.

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 $p^*\left(u^k\left|X,j\right.\right)$ which enter the counterfactual probability (11) are:

$$\widehat{p}^*\left(u^k\left|X,j\right.\right)=\frac{\widehat{N}^*\left(u^k\left|X,j\right.\right)}{\widehat{N}^*\left(u^k\left|j\right.\right)} \tag{23}$$

where $\widehat{N}^*\left(u\left|j\right.\right)=\sum_\ell\widehat{N}^*\left(u\left|X^\ell,j\right.\right)$ 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:

$$\widehat{\phi}^*\left(u^k\left|j\right.\right)=\sum\nolimits_\ell\widehat{p}^*\left(u^k\left|X^\ell,j\right.\right)\exp\left[X^\ell\widehat{\beta}^*_j\left(u^k\right)\right] \tag{24}$$

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

We can also recover the counterfactuals of gender-j cumulative and density of outcomes when positions are ranked according to the outcome. We first consider the original sample and sort positions in ascending order according to their hierarchy, denoting by w^k the k^{th} outcome. 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 (12) as:

$$\widehat{F}_j^* \left(w^k\right) = \frac{\widehat{N}^* \left(\widehat{F} \left(w^k\right) |j\right)}{N (j)} \tag{25}$$

where $N(j)$ is the number of gender-j individuals and $\widehat{F} \left(w^k\right)$ is an estimator of the outcome cumulative of positions computed at outcome w^k. A counterfactual of gender-j density is obtained from equation (14) replacing right-hand terms by estimators:

$$\widehat{f}_j^* \left(w^k\right) = \frac{\widehat{f} \left(w^k\right)}{N (j) /N} \frac{\widehat{N}^* \left(F \left(w^k\right) |j\right) \widehat{\phi}^* \left(F \left(w^k\right) |j\right)}{\widehat{N}^* \left(F \left(w^k\right) |f\right) \widehat{\phi}^* \left(F \left(w^k\right) |f\right) + \widehat{N}^* \left(F \left(w^k\right) |m\right) \widehat{\phi}^* \left(F \left(w^k\right) |m\right)} \tag{26}$$

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

4 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 D. 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.

4.1 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.[11] 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. We limit our attention to the single year 2011 as our assignment model is cross-section. 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

[11] 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.

children being the reference). Note that we are rather parcimonious 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 outcomes and individual characteristics for the two genders in the two sectors. Figure 1 shows that gender outcome 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, 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]

4.2 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 (15) 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. As we will see, 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.[12]

[12]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.

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 postions 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 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]

4.3 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 (20). 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 E.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 E.2. We also assess whether ignored unobserved individual heterogeneity might bias the estimates and Appendix E.3 suggests that it is not the case as long as this heterogeneity is not exagerately important.

5 Conclusion

In this paper, we show how to quantify outcome 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 outcomes. 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 outcome. 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 outcome 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] Chernozhukov V., Fernández-Val I. and B. Melly (2013), "Inference on counterfactual distributions", *Econometrica*, 81(6), pp. 2205-2268.

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

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] Fack G., Grenet J. and Y. He (2015), "Beyond Truth-Telling: Preference Estimation with Centralized School Choice", PSE Working Paper 2015-35.

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

[13] 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.

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

[15] 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.

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

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

[18] 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.

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

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

[21] 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.

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

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

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

[25] 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.

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

[27] 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 $\left\{n\left(\cdot\left|X^\ell, f\right.\right), n\left(\cdot\left|X^\ell, m\right.\right)\right\}_{\ell=1,\dots,L}$ verifying (6) where $\phi\left(u\left|X, j\right.\right)$ is given by (4).*

Proof. The proof revolves around the application of the Cauchy-Lipschitz theorem. Plugging (4) into (6), we get for any j and k:

$$n'\left(u\left|X^k, j\right.\right) = \frac{n\left(u\left|X^k, j\right.\right)\mu\left(u\left|X^k, j\right.\right)}{\sum\limits_{\ell,g} n\left(u\left|X^\ell, g\right.\right)\mu\left(u\left|X^\ell, g\right.\right)} \tag{27}$$

Introduce the vectors

$$\overline{\mu}(u) = \left[\mu\left(u\left|X^1, f\right.\right), .., \mu\left(u\left|X^L, f\right.\right), \mu\left(u\left|X^1, m\right.\right), ..., \mu\left(u\left|X^L, m\right.\right)\right]' \tag{28}$$

$$\overline{n}(u) = \left[n\left(u\left|X^1, f\right.\right), .., n\left(u\left|X^L, f\right.\right), n\left(u\left|X^1, m\right.\right), ..., n\left(u\left|X^L, m\right.\right)\right]' \tag{29}$$

A stacked version of (27) is given by:

$$\overline{n}'(u) = g\left(u, \overline{n}(u)\right) \tag{30}$$

with:

$$g\left(u, \overline{n}(u)\right) = \frac{\overline{n}(u) .* .\overline{\mu}(u)}{\langle \overline{n}(u), \overline{\mu}(u)\rangle} \tag{31}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclydian 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 (30) 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 (30) has a unique solution for $\overline{n}(\cdot)$ on $[\varepsilon, 1]$. As this is true for ε arbitrarily close to zero, (30) has a unique solution for $\overline{n}(\cdot)$ on $(0,1]$.

B Consistency

In this Appendix, we prove that $N\left(u^{\lfloor vN \rfloor}\left|X\right.\right)/N \underset{a.s.}{\rightarrow} n\left(v\left|X\right.\right)$ for all $v \in (0,1)$ where gender j enters the set of observable charateristics to simplify the notations. The proof of consistency can be decomposed into three stages:

- We first consider a setting where $\beta(.)$ is constant and we establishes consistency using statistical results proposed by Rosén (1972).
- We then extend the consistency result to the case where $\beta(.)$ is piece-wise constant. This is done by proving consistency on each interval.
- We finally extend the consistency result to the case where $\beta(.)$ is a continuous function by using a piece-wise constant approximation of it.

1/ Case of a constant function $\beta(.)$

Lemma B1 (Rosén, 1972). Consider the sampling without replacement of N individuals with weights p_i for $i = 1, ..., N$, such that $\sum_{i=1}^{N} p_i = 1$. Define implicitly the function $t(.)$ with the equation:

$$N - y = \sum_{i=1}^{N} \exp\left(-p_i t\left(y\right)\right) \tag{32}$$

Attribute to each individual i a quantity a_i and define $Z_k = \sum_{i=1}^{N} W_i a_i$ where W_i is a dummy for individual i to be in the sample constituted by the first k draws. We then have:

$$EZ_k = \sum_{i=1}^{N} a_i \left(1 - \exp\left[-p_i t\left(k\right)\right]\right) - r_E\left(N-k\right) N^{\frac{1}{2}} \tag{33}$$

$$VZ_k = \sum_{i=1}^{N} \left[a_i - \frac{\sum_{i=1}^{N} p_i a_i \exp\left[-p_i t\left(k\right)\right]}{\sum_{i=1}^{N} p_i \exp\left[-p_i t\left(k\right)\right]}\right]^2 \left[1 - \exp\left[-p_i t\left(k\right)\right]\right] \exp\left[-p_i t\left(k\right)\right] + r_V\left(k\right) \tag{34}$$

where $\lim_{N\to\infty} \max_{\tau_1 \leq k/N \leq \tau_2} \left|r_E\left(k\right)\right| = 0$ and $\lim_{N\to\infty} \max_{\tau_1 \leq k/N \leq \tau_2} \left|r_V\left(k\right)/N\right| = 0$ for all $0 < \tau_1 < \tau_2 < 1$.

Denote $k = \lfloor uN \rfloor$ with $\lfloor \cdot \rfloor$ denoting the integer part for $u \in (0,1)$ where we omit the dependance of k with respect to u and N for readability in the proofs, and by u^k the k^{th} empirical rank which is given by $u^k = k/N$. We are going to show that:

$$\frac{N\left(u^k\left|X\right.\right)}{N} = n\left(u\left|X\right.\right) + O\left(N^{-1/2}\right) \tag{35}$$

The roadmap of the proof is the following:

1. Using Lemma B1, we will first show that:

$$E\left[N\left(u^k\left|X\right.\right)\right]=N\left(X\right)\exp\left[-p_X^N t_N\left(1-k/N\right)\right]+r_E\left(k\right)N^{\frac{1}{2}} \tag{36}$$

with $\lim_{N\to\infty}\max_{\tau_1 N\leq k\leq\tau_2 N}\left|r_E\left(k\right)\right|=0$ for every $0<\tau_1<\tau_2<1$, $N\left(X\right)$ the number of individuals with characteristics X in the population and $t_N\left(u\right)=t\left(Nu\right)/N$ with $t\left(\bullet\right)$ defined according to (32) with $p_i=p_{X_i}^N$ where:

$$p_X^N=\frac{\exp\left(X\beta\right)}{E_{X,emp}\left[\exp\left(X\beta\right)\right]} \tag{37}$$

with $E_{X,emp}$ the empirical expectation computed using the empirical distribution of X such that:

$$E_{X,emp}\left[\exp\left(X\beta\right)\right]=\sum_\ell N\left(X^\ell\right)\exp\left(X^\ell\beta\right)/N \tag{38}$$

2. We will establish the intermediary result that $t_N\left(u\right)$ converges almost surely to $t_\infty\left(u\right)$ over $[0,1)$, where t_∞ is implicitely defined by the equation:

$$1-u=E_X\left[\exp\left(-p_X t_\infty\left(u\right)\right)\right] \tag{39}$$

where E_X is the expectation computed using the aymptotic distribution of X and:

$$p_X=\frac{\exp\left(X\beta\right)}{E_X\left[\exp\left(X\beta\right)\right]} \tag{40}$$

and show that we have:

$$N^{1/2}\left(t_N\left(u\right)-t_\infty\left(u\right)\right)\xrightarrow{L}N\left(0,\frac{1}{\left(1-u\right)^2}\overrightarrow{g}\left(u\right)'V\overrightarrow{g}\left(u\right)\right) \tag{41}$$

where V is a matrix which terms (l,l') for $l\neq l'$ are given by $-n\left(X^l\right)n\left(X^{l'}\right)$ and for $l=l'$ are given by $n\left(X^l\right)\left[1-n\left(X^l\right)\right]$, and $\overrightarrow{g}\left(u\right)=\left[g\left(X^1,u\right),...,g\left(X^L,u\right)\right]'$ with $g\left(X,u\right)=p_X t_\infty\left(u\right)E_X\left[p_X\exp\left(-p_X t_\infty\left(u\right)\right)\right]+$ $\exp\left[-p_X t_\infty\left(u\right)\right]$.

3. From 1/ and 2/, we will deduce that:

$$\frac{E\left[N\left(u^k\left|X\right.\right)\right]}{N}=n\left(u\left|X\right.\right)+o\left(N^{-1/2}\right) \tag{42}$$

4. Finally, we will get that:

$$\frac{N\left(u^k\left|X\right.\right)}{N}=n\left(u\left|X\right.\right)+O\left(N^{-1/2}\right) \tag{43}$$

Step 1 can be shown by applying Lemma B1 with $p_i = p_X^N$ being the probability that an individual has characteristics X, $a_i = 1_{i\in X}$ being a dummy for individual i having characteristics X and the sum (33) being computed for $N-k$ draws. We then have $Z_{N-k} = \sum_{i=1}^{N} \left(1 - 1_{i\in\Omega(u^k)}\right) 1_{i\in X}$ where $1_{i\in\Omega(u^k)}$ denotes the dummy for still being in the risk set after $N-k$ draws such that:

$$EZ_{N-k} = N\left(X\right)\left[1 - \exp\left[-p_X^N t\left(N-k\right)/N\right]\right] - r_E\left(k\right) N^{\frac{1}{2}} \tag{44}$$

As we have:

$$Z_{N-k} = \sum_{i=1}^{N} \left(1_{i\in X} - 1_{i\in\Omega(u^k|X)}\right) = N\left(X\right) - N\left(u^k\left|X\right.\right) \tag{45}$$

equation (44) can be rewritten as:

$$N\left(X\right) - E\left[N\left(u^k\left|X\right.\right)\right] = N\left(X\right)\left[1 - \exp\left[-p_X^N t_N\left(1-k/N\right)\right]\right] - r_E\left(k\right) N^{\frac{1}{2}} \tag{46}$$

with $t_N\left(u\right) = t\left(Nu\right)/N$. Rearranging the terms, we get expression (36).

Step 2 can be completed by first showing that $t_N\left(u\right)$ converges almost surely to $t_\infty\left(u\right)$ over $[0,1)$. Rewriting (32) applied to $p_i = p_{X_i}^N$ where X_i are the characteristics of individual i and $y = Nu$, we get:

$$N - Nu = \sum_{i=1}^{N} \exp\left(-p_{X_i}^N t\left(Nu\right)\right) \tag{47}$$

Considering the definition of t_N and the fact that the characteristics of individuals are distributed across the values X^ℓ with $\ell = 1, ..., N$, we get:

$$1 - u = E_{X,emp}\left[\exp\left[-p_X^N t_N\left(u\right)\right]\right] \tag{48}$$

We have $N\left(X^\ell\right)/N \to n\left(X^\ell\right)$ almost surely and thus:

$$\lim_N p_X^N = \frac{\exp\left(X\beta\right)}{\sum_\ell n\left(X^\ell\right)\exp\left(X^\ell\beta\right)} = \frac{\exp\left(X\beta\right)}{E_X\left[\exp\left(X\beta\right)\right]} \equiv p_X \tag{49}$$

Since function $f\left(x\right) = E_X\left[\exp\left(-p_X x\right)\right]$ is continuous and monotonic, we finally have $\lim_{N\to\infty} t_N\left(u\right) = t_\infty\left(u\right)$ for all $u \in [0,1)$ with:

$$1 - u = E_X\left[\exp\left(-p_X t_\infty\left(u\right)\right)\right] \tag{50}$$

We now establish the asymptotic distribution of t_N. For that purpose, we establish the asymptotic distribution of $E_X\left[\exp\left(-p_X t_N(u)\right)\right]$ and apply the delta method to the inverse of the function $f(\bullet)$. We have the decomposition:

$$\begin{aligned} E_X\left[\exp\left(-p_X t_N(u)\right)\right] - E_X\left[\exp\left(-p_X t_\infty(u)\right)\right] &= E_X\left[\exp\left(-p_X t_N(u)\right)\right] - E_X\left[\exp\left(-p_X^N t_N(u)\right)\right] \\ &\quad + E_X\left[\exp\left(-p_X^N t_N(u)\right)\right] - E_X\left[\exp\left(-p_X t_\infty(u)\right)\right] \quad (51) \end{aligned}$$

Rearranging the first right-hand side term in (51), we get:

$$E_X\left[\exp\left(-p_X t_N(u)\right)\right] - E_X\left[\exp\left(-p_X^N t_N(u)\right)\right] = E_X\left[\exp\left(-p_X t_N(u)\right)\left[1 - \exp\left(-\left(p_X^N - p_X\right)t_N(u)\right)\right]\right] \quad (52)$$

To develop further this expression, we first derive useful properties for $p_X^N - p_X$. Using the fact that p_X^N converges to p_X, we have:

$$\begin{aligned} p_X^N - p_X &= \frac{\exp(X\beta)}{E_{X,emp}\left[\exp(X\beta)\right]} - \frac{\exp(X\beta)}{E_X\left[\exp(X\beta)\right]} && (53) \\ &= -\frac{\exp(X\beta)}{E_{X,emp}\left[\exp(X\beta)\right] E_X\left[\exp(X\beta)\right]}\left[E_{X,emp}\left[\exp(X\beta)\right] - E_X\left[\exp(X\beta)\right]\right] && (54) \\ &= -\frac{\exp(X\beta)}{E_{X,emp}\left[\exp(X\beta)\right] E_X\left[\exp(X\beta)\right]}\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right]\exp\left(X^\ell\beta\right) && (55) \\ &= -p_X^N \sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] p_{X^\ell} && (56) \end{aligned}$$

Using the fact that $p_X^N = \left(p_X^N - p_X\right) + p_X$ on the right-hand side and rearranging terms by passing those involving $p_X^N - p_X$ on the left-hand side, we get:

$$p_X^N - p_X = -p_X \frac{\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] p_{X^\ell}}{1 + \sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] p_{X^\ell}} \quad (57)$$

Using the central limit theorem, we have $N\left(X^\ell\right)/N - n\left(X^\ell\right) = O\left(N^{-1/2}\right)$. Using a Taylor expansion of the right-hand side denominator of (57), we then get that:

$$p_X^N - p_X = -p_X \sum_\ell \left[N\left(X_\ell\right)/N - n\left(X_\ell\right)\right] p_{X^\ell} + O\left(N^{-1}\right) \quad (58)$$

Injecting (58) into (52) and making a Taylor expansion, we get:

$$\begin{aligned} & E_X\left[\exp\left(-p_X t_N(u)\right)\right] - E_X\left[\exp\left(-p_X^N t_N(u)\right)\right] \\ = & -E_X\left[\exp\left(-p_X t_N(u)\right)\left[p_X \sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] p_{X^\ell} t_N(u) + O\left(N^{-1}\right)\right]\right] && (59) \\ = & -\left[\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] p_{X^\ell}\right] E_X\left[p_X t_N(u)\exp\left(-p_X t_N(u)\right)\right] + O\left(N^{-1}\right) && (60) \end{aligned}$$

Using the fact that $t_N(u) \to t_\infty(u)$ almost surely, equation (60) becomes:

$$\begin{aligned}
& E_X\left[\exp\left(-p_X t_N(u)\right)\right] - E_X\left[\exp\left(-p_X^N t_N(u)\right)\right] \\
= & -\left[\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] p_{X_\ell}\right] t_\infty(u)\, E_X\left[p_X \exp\left(-p_X t_\infty(u)\right)\right] + o\left(N^{-1/2}\right) \qquad (61)
\end{aligned}$$

The second right-hand side term in (51) can be rewritten considering that by definition:

$$E_X\left[\exp\left(-p_X t_\infty(u)\right)\right] = 1 - u = \sum_\ell N\left(X^\ell\right) \exp\left[-p_{X^\ell}^N t_N(u)\right]/N \qquad (62)$$

This yields:

$$\begin{aligned}
& E_X\left[\exp\left(-p_X^N t_N(u)\right)\right] - E_X\left[\exp\left(-p_X t_\infty(u)\right)\right] \\
= & -\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] \exp\left[-p_{X^\ell}^N t_N(u)\right] \qquad (63) \\
= & -\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] \exp\left[-p_{X^\ell} t_\infty(u)\right] + o\left(N^{-1/2}\right) \qquad (64)
\end{aligned}$$

Substituting (61) and (64) into (51) gives:

$$\begin{aligned}
& E_X\left[\exp\left(-p_X t_N(u)\right) - E_X\left[\exp\left(-p_X t_\infty(u)\right)\right]\right] \\
= & -\sum_\ell \left[N\left(X^\ell\right)/N - n\left(X^\ell\right)\right] g\left(X^\ell, u\right) + o\left(N^{-1/2}\right) \qquad (65)
\end{aligned}$$

where $g(X,u) = p_X t_\infty(u)\, E_X\left[p_X \exp\left(-p_X t_\infty(u)\right)\right] + \exp\left[-p_X t_\infty(u)\right]$.

The vector $\left[N\left(X^1\right)/N, ..., N\left(X^L\right)/N\right]'$ contains the averages of dummies for the categories of a multinomial law giving to an individual i a value of explanatory variables X_i. Its expectation is $\left[n\left(X^1\right), ..., n\left(X^L\right)\right]'$ and its law is asymptotically normal with an asymptotic covariance matrix V which terms (l, l') for $l \neq l'$ are given by $-n\left(X^l\right) n\left(X^{l'}\right)$ and for $l = l'$ are given by $n\left(X^l\right)\left[1 - n\left(X^l\right)\right]$. Defining $\overrightarrow{g}(u) = \left[g\left(X^1, u\right), ..., g\left(X^L, u\right)\right]'$ and applying the delta method, we obtain:

$$N^{1/2}\left(E_X\left[\exp\left(-p_X t_N(u)\right)\right] - E_X\left[\exp\left(-p_X t_\infty(u)\right)\right]\right) \xrightarrow{L} N\left(0, \overrightarrow{g}(u)' V \overrightarrow{g}(u)\right) \qquad (66)$$

Applying the delta method again to the inverse of function $f(x) = E_X\left[\exp\left(-p_X x\right)\right]$ which derivative is given by $f^{-1\prime}(y) = 1/f'\left(f^{-1}(y)\right) = 1/E_X\left[\exp\left(-p_X f^{-1}(y)\right)\right]$, and considering the equality (50), we finally get the expression:

$$N^{1/2}\left(t_N(u) - t_\infty(u)\right) \xrightarrow{L} N\left(0, \frac{1}{(1-u)^2} \overrightarrow{g}(u)' V \overrightarrow{g}(u)\right) \qquad (67)$$

Step 3 can be shown from (36) using Taylor expansions. Indeed, we have by definition that $|u - k/N| \leq 1/N$. We also have from (58) that $p_X^N - p_X = O(N^{-1/2})$ since $N(X)/N - n(X) = O\left(N^{-1/2}\right)$, and from (67) that $t_N(1-u) = t_\infty(1-u) + O(N^{-1/2})$. This yields that:

$$\exp\left[-p_X^N t_N(1 - k/N)\right] = \exp\left[-p_X t_\infty(1-u)\right] + O\left(N^{-1/2}\right) \tag{68}$$

and thus, we get from (36):

$$E\left[N\left(u^k \left|X\right.\right)\right]/N = n(X)\exp\left[-p_X t_\infty(1-u)\right] + O\left(N^{-1/2}\right) \tag{69}$$

Denoting $\overline{n}_X(u) = n(X)\exp(-p_X t_\infty(1-u))$, it is possible to establish that $\overline{n}_X(u) = n(u\left|X\right.)$ which will complete the proof. This can be done by deriving $\overline{n}_X(u)$, as we have:

$$\overline{n}'_X(u) = p_X t'_\infty(1-u)\, n(X)\exp(-p_X t_\infty(1-u)) \tag{70}$$

and deriving equation (39), we get:

$$t'_\infty(1-u)\, E_X\left[p_X \exp(-p_X t_\infty(1-u))\right] = 1 \tag{71}$$

which yields after substituting $p_X t'_\infty(1-u)$ by its expression into (70) that:

$$\begin{aligned}\overline{n}'_X(u) &= \frac{p_X n(X)\exp(-p_X t_\infty(1-u))}{E_X\left[p_X \exp(-p_X t_\infty(1-u))\right]} &(72)\\ &= \frac{p_X \overline{n}_X(u)}{\sum\limits_\ell p_{X^\ell}\overline{n}_{X^\ell}(u)} &(73)\end{aligned}$$

and, as shown in Appendix A, this expression has a unique solution under the initial condition $\overline{n}(1, X) = n(X)$ which is $n(u\left|X\right.)$.

Step 4 can be established showing that the variance of $N\left(u^k \left|X\right.\right)/N$ tends to zero as N tends to infinity. Applying Lemma A1 with a_i a dummy for individual i having characteristics X, we get:

$$\begin{aligned}V\left[N\left(u^k\left|X\right.\right)/N\right] &= \sum_\ell N\left(X^\ell\right)\left[1_{\{X^\ell = X\}} - \frac{N(X)\, p_X^N \exp\left[-p_X^N t(N-k)/N\right]/N}{E_{emp}\left[p_X^N \exp\left[-p_X^N t(N-k)/N\right]\right]}\right]^2 \\ &\quad\times\left[1 - \exp\left[-p_{X^\ell}^N t(N-k)/N\right]\right]\exp\left[-p_{X^\ell}^N t(N-k)/N\right]/N^2 + r_V(N-k)/N^2 \quad (74)\end{aligned}$$

The term in brackets is bounded by one. As function $t(\cdot)$ is positive, $\exp\left[-p_X^N t(N-k)/N\right]$ is also bounded by one. This yields:

$$V\left[N\left(u^k\left|X\right.\right)/N\right]=O\left(1/N\right)+r_V\left(N-k\right)/N^2 \tag{75}$$

and this is enough to get: $V\left[N\left(u^k\left|X\right.\right)/N\right]\to 0$ as N tends to infinity.

2/ Case of a piece-wise constant function $\beta\left(.\right)$

We now consider the case where the coefficients of explanatory variables are piece-wise constant and we denote them as $\beta\left(u\right)$. These coefficients change value at knots υ^z such that $\varepsilon<\upsilon^1<...<\upsilon^Z=1$ with ε positive and small. We consider the asymptotic case where N, Z and N/Z (the average number of individuals by interval) all tend to infinity. We can apply the proof of case 1/ over the interval $\left(\upsilon^{Z-1},1\right]$ and we have:

$$p\lim E\left[N\left(u^k\left|X\right.\right)/N\right]=n\left(u\left|X\right.\right),\ \forall u\in\left(\upsilon_Z,1\right] \tag{76}$$

where:

$$n'\left(u\left|X\right.\right)=\frac{\exp\left[X\beta\left(1\right)\right]n\left(u\left|X\right.\right)}{\sum\limits_{\ell}\exp\left[X^{\ell}\beta\left(1\right)\right]n\left(u\left|X^{\ell}\right.\right)} \tag{77}$$

with $n\left(1\left|X\right.\right)=n\left(X\right)$. This proves that the limit is $n\left(u\left|X\right.\right)$ as defined in the text on the interval $\left(\upsilon^Z,1\right]$. We are going to show recursively that this result holds at lower ranks. Assume that it is true down to rank υ^z and consider that $n\left(\upsilon^z\left|X\right.\right)=\liminf_{\upsilon^z<u}n\left(u\left|X\right.\right)$. Proof of case 1/ can be applied over the interval $\left(\upsilon^{z-1},\upsilon^z\right]$ and we have:

$$p\lim E\left[N\left(u^k\left|X\right.\right)/N\right]=n\left(u\left|X\right.\right),\ \forall u\in\left(\upsilon^{z-1},\upsilon^z\right] \tag{78}$$

where:

$$n'\left(u\left|X\right.\right)=\frac{\exp\left[X\beta\left(\upsilon^z\right)\right]n\left(u\left|X\right.\right)}{\sum\limits_{\ell}\exp\left[X^{\ell}\beta\left(\upsilon^z\right)\right]n\left(u\left|X^{\ell}\right.\right)} \tag{79}$$

This proves that the limit is $n\left(u\left|X\right.\right)$ as defined in the text on the interval $\left(\upsilon^{z-1},\upsilon^z\right]$. After iterations, we get the result for the whole interval $\left[\varepsilon,1\right]$.

3/ Case of a continuous function $\beta\left(.\right)$

We now consider the case where the coefficients $\beta\left(u\right)$ are continuous and we use a step-wise approximation of $\beta\left(u\right)$ given by:

$$\beta_Z\left(u\right)=\sum_{z=1}^{Z}\beta\left(\frac{z}{Z}\right)1\left\{z/Z-1<u\leq z/Z\right\} \tag{80}$$

We define by $n_Z\left(u\left|X\right.\right)$ the measure of individuals with characteristics X still available for the job of rank u when the assignment mechanism is defined by $\beta_Z\left(u\right)$ instead of $\beta\left(u\right)$, and by $N_Z\left(u\left|X\right.\right)$ the random variable that corresponds to the number of individuals with characteristics X still available at rank u under this assignment rule. The proportion of individuals in the population with characteristics X still available at rank u under this assignment rule is given by $P_Z(u|X) = N_Z(u|X)/N$. We have shown in case 2/ that $P_Z(u|X)$ tends to $n_Z(u|X)$ when N, Z and N/Z tend to infinity. The corresponding proportion when using the assignment rule defined by $\beta(u)$ is $P(u|X) = N(u|X)/N$, and we denote by $\Delta P(u|X) = P(u|X) - P_Z(u|X)$ the difference in the proportion of individuals under the two assignment rules.

We are going to show that for any $z \leq Z$ and rank $\upsilon^z = \lfloor N.z/Z \rfloor /N$, we have that:

$$\Delta P(\upsilon^z|X) = O\left(\frac{1}{Z^{1/2}}\right) + O\left(\frac{1}{N^{1/2}}\right) \tag{81}$$

uniformely over z in the sense that $O(\cdot)$ involves a bound that does not depend on z. The idea of the proof is to consider that:

$$\Delta P\left(\upsilon^{z-1}|X\right) = E_{\Omega(\upsilon^z)}\left[\Delta P\left(\upsilon^{z-1}|X\right)\right] + \varepsilon_{z,X} \tag{82}$$

where $\varepsilon_{z,X}$ is a random component depending on X and $E_{\Omega(\upsilon^z)}$ is the expectation operator conditional on the information available at rank υ^z, and then bound the right-hand side terms. Equation (81) will be enough together with case 2/ to get our main result. In more details, the proof can be decomposed into the following steps:

- 1. We first show that the random component verifies:

$$\varepsilon_{z,X} = O\left(Z^{-1/2}\right) + O\left(N^{-1/2}\right) \tag{83}$$

  2. We will then show that we have:

$$E_{\Omega(\upsilon^z)}\left[\Delta P\left(\upsilon^{z-1}|X\right)\right] = \Delta P\left(\Delta\upsilon^z|X\right) + O\left(\frac{1}{Z^2}\right) + O\left(\frac{1}{Z}\right)\max_{\ell}\left[\Delta P\left(\upsilon^z\left|X^{\ell}\right.\right)\right] \tag{84}$$

  We then obtain (81) by recurrence. Consider first the case $z = Z$. The relationship (81) holds trivially since $N_Z\left(\upsilon^Z|X\right) = N\left(\upsilon^Z|X\right) = N(X)$ as $\upsilon^Z = 1$. Then, supposing that (81) is true for a given z, it is straightfoward from (82), (83) and (84) to show that it is also true for $z-1$.

  3. Finally, the relationship (81) is enough to show that $E\left[N\left(u^{\lfloor vN \rfloor+1}|X\right)\right]/N \underset{a.s.}{\rightarrow} n(v|X)$ when N tends to infinity and Z is chosen such that Z and N/Z tend to infinity (this is for instance the case when $Z = N^{1/2}$). This can be done by applying the results of case 2/ to the piece-wise constant function $\beta_Z(.)$ since there is then an infinity of positions in each of the intervals defined by the nodes.

Step 1 can be completed considering that the processes underlying $P\left(\upsilon^{z-1}|X\right)$ et $P_Z\left(\upsilon^{z-1}|X\right)$ are independent, as we have:

$$V_{\Omega(\upsilon^z)}\left(\varepsilon_{z,X}\right) = V_{\Omega(\upsilon^z)}\left[P\left(\upsilon^z\left|X\right.\right) - P\left(\upsilon^{z-1}\left|X\right.\right)\right] + V_{\Omega(\upsilon^z)}\left[P_Z\left(\upsilon^z\left|X\right.\right) - P_Z\left(\upsilon^{z-1}\left|X\right.\right)\right] \tag{85}$$

where $V_{\Omega(\upsilon^z)}$ is the variance operator conditional on the information available at rank υ^z. By construction,

$$\begin{aligned} 0 &\leq P\left(\upsilon^z\left|X\right.\right) - P\left(\upsilon^{z-1}\left|X\right.\right) \\ &\leq \frac{\left\lfloor N\frac{z}{Z}\right\rfloor - \left\lfloor N\frac{z-1}{Z}\right\rfloor}{N} &(86)\\ &\leq \frac{1}{Z} + \frac{1}{N} = O\left(\frac{1}{Z}\right) + O\left(\frac{1}{N}\right) &(87)\end{aligned}$$

The same inequality holds for $P_Z\left(\upsilon^z\left|X\right.\right) - P_Z\left(\upsilon^{z-1}\left|X\right.\right)$ and it yields:

$$V_{\Omega(\upsilon^z)}\left(\varepsilon_{z,X}\right) = O\left(\frac{1}{Z}\right) + O\left(\frac{1}{N}\right) \tag{88}$$

which leads to equation (83).

Step 2 can be shown considering that we have:

$$E_{\Omega(\upsilon^z)}\left[P\left(\upsilon^z\left|X\right.\right) - P\left(\upsilon^{z-1}\left|X\right.\right)\right] = \frac{1}{N}\sum_{k=\left\lfloor N\frac{z-1}{Z}\right\rfloor+1}^{\left\lfloor N\frac{z}{Z}\right\rfloor} E_{\Omega(\upsilon^z)}\left[\frac{P\left(u^k\left|X\right.\right)\exp\left(X\beta\left(u^k\right)\right)}{\sum_{\ell} P\left(u^k\left|X^\ell\right.\right)\exp\left(X^\ell\beta\left(u^k\right)\right)}\right] \tag{89}$$

To ease the notations in the sequel, we denote the length of an interval as:

$$\Delta_z = \left(\left\lfloor N\frac{z}{Z}\right\rfloor - \left\lfloor N\frac{z-1}{Z}\right\rfloor\right)/N \tag{90}$$

where N and Z are not introduced as subscripts to ease the readability. For all values of X and all values of k such that $\left\lfloor N\frac{z-1}{Z}\right\rfloor + 1 \leq k \leq \left\lfloor N\frac{z}{Z}\right\rfloor$, we have:

$$P\left(\upsilon^{z-1}\left|X\right.\right) \leq P\left(u^k\left|X\right.\right) \leq P\left(\upsilon^z\left|X\right.\right) \tag{91}$$

A lower bound for $P\left(\upsilon^{z-1}\left|X\right.\right)$ is the decrease of the proportion $P\left(\upsilon^z\left|X\right.\right)$ when all the jobs with ranks between υ^{z-1} and υ^z are filled by individuals with characteristics X, and thus we have the following inequalities:

$$\begin{aligned} P\left(\upsilon^z\left|X\right.\right) - \Delta_z &\leq P\left(u^k\left|X\right.\right) \leq P\left(\upsilon^z\left|X\right.\right) &(92)\\ \min_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X\beta\left(u^k\right)\right) &\leq \exp\left(X\beta\left(u^k\right)\right) \leq \max_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X\beta\left(u^k\right)\right) &(93)\end{aligned}$$

Using these inequalities, we obtain that:

$$\Delta_z \frac{\left(P\left(\upsilon^z\left|X\right.\right)-\Delta_z\right)\min\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X\beta\left(u^k\right)\right)}{\sum\limits_{\ell}P\left(\upsilon^z\left|X^\ell\right.\right)\max\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X^\ell\beta\left(u^k\right)\right)}$$

$$\leq \frac{1}{N}\sum_{k=\left\lfloor N\frac{z-1}{Z}\right\rfloor+1}^{\left\lfloor N\frac{z}{Z}\right\rfloor} E_{\Omega(\upsilon^z)}\left[\frac{P\left(u^k\left|X\right.\right)\exp\left(X\beta\left(u^k\right)\right)}{\sum\limits_{\ell}P\left(u^k\left|X^\ell\right.\right)\exp\left(X^\ell\beta\left(u^k\right)\right)}\right]\leq \tag{94}$$

$$\Delta_z \frac{P\left(\upsilon^z\left|X\right.\right)\max\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X\beta\left(u^k\right)\right)}{\sum\limits_{\ell}\left[P\left(\upsilon^z\left|X^\ell\right.\right)-\Delta_z\right]\min\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X^\ell\beta\left(u^k\right)\right)} \tag{95}$$

and from (89), we get:

$$\frac{\left(P\left(\upsilon^z\left|X\right.\right)-\Delta_z\right)\min\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X\beta\left(u^k\right)\right)}{\sum\limits_{\ell}P\left(\upsilon^z\left|X^\ell\right.\right)\max\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X^\ell\beta\left(u^k\right)\right)}$$

$$\leq \frac{E_{\Omega(\upsilon^z)}\left[P\left(\upsilon^z\left|X\right.\right)-P\left(\upsilon^{z-1}\left|X\right.\right)\right]}{\Delta_z}\leq \tag{96}$$

$$\frac{P\left(\upsilon^z\left|X\right.\right)\max\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X\beta\left(u^k\right)\right)}{\sum\limits_{\ell}\left[P\left(\upsilon^z\left|X^\ell\right.\right)-\Delta_z\right]\min\limits_{\frac{z-1}{Z}\leq k\leq\frac{z}{Z}}\exp\left(X^\ell\beta\left(u^k\right)\right)} \tag{97}$$

We can get another inequality using the fact that $\beta_Z(u)$ converges uniformly to $\beta(u)$, as there exists a constant denoted C such that for all X and k such that $\left\lfloor N\frac{z-1}{Z}\right\rfloor+1\leq k\leq\left\lfloor N\frac{z}{Z}\right\rfloor$, we have:

$$\left|\exp X\beta\left(u^k\right)-\exp X\beta_Z\left(\frac{z}{Z}\right)\right|<\frac{C}{Z} \tag{98}$$

Using this inequality and the fact that $\beta_Z\left(\frac{z}{Z}\right)=\beta\left(\frac{z}{Z}\right)$, we deduce from (97) that:

$$\frac{\left(P\left(\upsilon^z\left|X\right.\right)-\Delta_z\right)\left[\exp\left(\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]}{\sum\limits_{\ell}P\left(\upsilon^z\left|X^\ell\right.\right)\left[\exp\left(\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]}$$

$$\leq \frac{E_{\Omega(\upsilon^z)}\left[P\left(\upsilon^z\left|X\right.\right)-P\left(\upsilon^{z-1}\left|X\right.\right)\right]}{\Delta_z}\leq \tag{99}$$

$$\frac{P\left(\upsilon^z\left|X\right.\right)\left[\exp\left(\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]}{\sum\limits_{\ell}\left[P\left(\upsilon^z\left|X^\ell\right.\right)-\Delta_z\right]\left[\exp\left(\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]} \tag{100}$$

An inequality similar to (97) holds when considering the assignment rule given by $\beta^{Z}(u)$, and we have:

$$\begin{aligned} & \frac{\left[P_Z\left(\upsilon^z\left|X\right.\right)-\Delta_z\right]\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell} P_Z\left(u^k\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)} \\ \leq \quad & \frac{E_{\Omega(\upsilon^z)}\left[P\left(\upsilon^z\left|X\right.\right)-P\left(\upsilon^{z-1}\left|X\right.\right)\right]}{\Delta_z} \leq \qquad (101) \\ & \frac{P_Z\left(\upsilon^z\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell}\left[P_Z\left(\upsilon^z\left|X^{\ell}\right.\right)-\Delta_z\right]\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)} \qquad (102) \end{aligned}$$

Computing the difference between (100) and (102), we get the inequality:

$$\begin{aligned} & E_{\Omega(\upsilon^z)}\left[\frac{\Delta P\left(\upsilon^z\left|X\right.\right)}{\Delta_z}\right]-E_{\Omega(\upsilon^z)}\left[\frac{\Delta P\left(\upsilon^{z-1}\left|X\right.\right)}{\Delta_z}\right] \\ \geq \quad & \frac{\left(P\left(\upsilon^z\left|X\right.\right)-\Delta_z\right)\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]}{\sum_{\ell} P\left(\upsilon^z\left|X^{\ell}\right.\right)\left[\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]}-\frac{P_Z\left(\upsilon^z\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell}\left[P_Z\left(\upsilon^z\left|X^{\ell}\right.\right)-\Delta_z\right]\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)} \end{aligned} \qquad (103)$$

We are going to make Taylor expansions of the right-hand side terms of (103). For the denominator of the first term, we have:

$$\begin{aligned} \sum_{\ell} P\left(\upsilon^z\left|X^{\ell}\right.\right)\left[\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right] \quad = \quad & \sum_{\ell} P_Z\left(\upsilon^z\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\sum_{\ell} P_Z\left(\upsilon^z\left|X^{\ell}\right.\right) \\ & +\sum_{\ell}\Delta P\left(\upsilon^z\left|X^{\ell}\right.\right)\left[\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right] \end{aligned} \qquad (104)$$

Denoting $\Psi_z=\max_{\ell}\left[\Delta P\left(\upsilon^z\left|X^{\ell}\right.\right)\right]$, we get from (104) that:

$$\sum_{\ell} P\left(\upsilon^z\left|X^{\ell}\right.\right)\left[\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]=\sum_{\ell} P_Z\left(\upsilon^z\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right)+O\left(\Psi_z\right) \qquad (105)$$

As we also have $\Delta_z=O\left(Z^{-1}\right)$, the numerator of the first right-hand side term of (103) can be rewritten such that:

$$\left(P\left(\upsilon^z\left|X\right.\right)-\Delta_z\right)\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]=P\left(\upsilon^z\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right) \qquad (106)$$

Using (105) and (106), and making an additional Taylor expansion, we get that:

$$
\begin{aligned}
&\frac{\left(P\left(\upsilon^z\,|X\right)-\Delta_z\right)\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]}{\sum\limits_{\ell}P\left(\upsilon^z\,|X^\ell\right)\left[\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]} \\
=\;& \frac{P\left(\upsilon^z\,|X\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right)}{\sum\limits_{\ell}P\left(\upsilon^z\,|X^\ell\right)\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)}\left[1+O\left(\frac{1}{Z}\right)+O\left(\Psi_z\right)\right] && (107)\\
=\;& \frac{P\left(\upsilon^z\,|X\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum\limits_{\ell}P_Z\left(\upsilon^z\,|X^\ell\right)\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)}+O\left(\frac{1}{Z}\right)+O\left(\Psi_z\right) && (108)
\end{aligned}
$$

Similarly, we can rewrite the second right-hand side term of (103) as:

$$
\begin{aligned}
\frac{P_Z\left(\upsilon^z\,|X\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum\limits_{\ell}\left[P_Z\left(\upsilon^z\,|X^\ell\right)-\Delta_z\right]\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)} &= \frac{P_Z\left(\upsilon^z\,|X\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum\limits_{\ell}P_Z\left(\upsilon^z\,|X^\ell\right)\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)}\left[1+O\left(\frac{1}{Z}\right)\right] && (109)\\
&= \frac{P_Z\left(\upsilon^z\,|X\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum\limits_{\ell}P_Z\left(\upsilon^z\,|X^\ell\right)\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)}+O\left(\frac{1}{Z}\right) && (110)
\end{aligned}
$$

This yields after an additional Taylor expansion that:

$$
\begin{aligned}
&E_{\Omega(\upsilon^z)}\left[\frac{\Delta P\left(\upsilon^z\,|X\right)}{\Delta_z}\right]-E_{\Omega(\upsilon^z)}\left[\frac{\Delta P\left(\upsilon^{z-1}\,|X\right)}{\Delta_z}\right] \\
\geq\;& \frac{\Delta P\left(\upsilon^z\,|X\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum\limits_{\ell}P_Z\left(\upsilon^z\,|X^\ell\right)\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)}+O\left(\frac{1}{Z}\right)+O\left(\Psi_z\right) && (111)\\
\geq\;& O\left(\frac{1}{Z}\right)+O\left(\Psi_z\right) && (112)
\end{aligned}
$$

In a similar way, we have:

$$
\begin{aligned}
&E_{\Omega(\upsilon^z)}\left[\frac{\Delta P\left(\upsilon^z\,|X\right)}{\Delta_z}\right]-E_{\Omega(\upsilon^z)}\left[\frac{\Delta P\left(\upsilon^{z-1}\,|X\right)}{\Delta_z}\right] \\
\leq\;& \frac{P\left(\upsilon^z\,|X\right)\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]}{\sum\limits_{\ell}\left[P\left(\upsilon^z\,|X^\ell\right)-\Delta_z\right]\left[\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]}-\frac{\left[P_Z\left(\upsilon^z\,|X\right)-\Delta_z\right]\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum\limits_{\ell}P_Z\left(u^k\,|X^\ell\right)\exp\left(X^\ell\beta\left(\frac{z}{Z}\right)\right)} && (113)
\end{aligned}
$$

Using the same line of arguments as before, we get:

$$\sum_{\ell}\left[P\left(\upsilon^{z}\left|X^{\ell}\right.\right)-\Delta_{z}\right]\left[\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right] = \sum_{\ell}P_{Z}\left(\upsilon^{z}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right)+O\left(\Psi_{z}\right) \tag{114}$$

$$P\left(\upsilon^{z}\left|X\right.\right)\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right] = P\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right) \tag{115}$$

$$\left[P_{Z}\left(\upsilon^{z}\left|X\right.\right)-\Delta_{z}\right]\exp\left(X\beta\left(\frac{z}{Z}\right)\right) = P_{Z}\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right) \tag{116}$$

Hence:

$$\begin{aligned}
&\frac{P\left(\upsilon^{z}\left|X\right.\right)\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+\frac{C}{Z}\right]}{\sum_{\ell}\left[P\left(\upsilon^{z}\left|X^{\ell}\right.\right)-\Delta_{z}\right]\left[\exp\left(X\beta\left(\frac{z}{Z}\right)\right)-\frac{C}{Z}\right]}\\
=\; & \frac{P\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right)}{\sum_{\ell}P_{Z}\left(\upsilon^{z}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)}\left[1+O\left(\frac{1}{Z}\right)+O\left(\Psi_{z}\right)\right] && (117)\\
=\; & \frac{P\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell}P_{Z}\left(\upsilon^{z}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)}+O\left(\frac{1}{Z}\right)+O\left(\Psi_{z}\right) && (118)
\end{aligned}$$

Similarly:

$$\begin{aligned}
\frac{\left[P_{Z}\left(\upsilon^{z}\left|X\right.\right)-\Delta_{z}\right]\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell}P_{Z}\left(u^{k}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)} &= \frac{P_{Z}\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)+O\left(\frac{1}{Z}\right)}{\sum_{\ell}P_{Z}\left(u^{k}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)} && (119)\\
&= \frac{P_{Z}\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell}P_{Z}\left(\upsilon^{z}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)}+O\left(\frac{1}{Z}\right) && (120)
\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}
&E_{\Omega(\upsilon^{z})}\left[\frac{\Delta P\left(\upsilon^{z}\left|X\right.\right)}{\Delta_{z}}\right]-E_{\Omega(\upsilon^{z})}\left[\frac{\Delta P\left(\upsilon^{z-1}\left|X\right.\right)}{\Delta_{z}}\right]\\
\leq\; & \frac{\Delta P\left(\upsilon^{z}\left|X\right.\right)\exp\left(X\beta\left(\frac{z}{Z}\right)\right)}{\sum_{\ell}P_{Z}\left(\upsilon^{z}\left|X^{\ell}\right.\right)\exp\left(X^{\ell}\beta\left(\frac{z}{Z}\right)\right)}+O\left(\frac{1}{Z}\right)+O\left(\Psi_{z}\right) && (121)\\
\leq\; & O\left(\frac{1}{Z}\right)+O\left(\Psi_{z}\right) && (122)
\end{aligned}$$

From (112) and (122), we get a lower bound and an upper bound for our quantity of interest that are associated to the same convergence speed. As a consequence, we obtain:

$$E_{\Omega(\upsilon^{z})}\left[\frac{\Delta P\left(\upsilon^{z}\left|X\right.\right)}{\Delta_{z}}\right]-E_{\Omega(\upsilon^{z})}\left[\frac{\Delta P\left(\upsilon^{z-1}\left|X\right.\right)}{\Delta_{z}}\right]=O\left(\frac{1}{Z}\right)+O\left(\Psi_{z}\right) \tag{123}$$

which can be rewritten as:

$$E_{\Omega(v^z)}\left[\Delta P\left(v^{z-1}\left|X\right.\right)\right]=E_{\Omega(v^z)}\left[\Delta P\left(v^{z}\left|X\right.\right)\right]+O\left(\frac{\Delta_z}{Z}\right)+O\left(\Delta_z\Psi_z\right) \tag{124}$$

Using the fact that $E_{\Omega(v^z)}\left[\frac{\Delta P(v^z|X)}{\Delta_z}\right]=\frac{\Delta P(v^z|X)}{\Delta_z}$ and $\Delta_z=O\left(Z^{-1}\right)$, we finally obtain expression (84).

Step 3 can then be completed to obtain our main result. First note that, for any $v\in(0,1]$, we have that $v^{z^N-1}\leq\lfloor vN\rfloor/(N-1)\leq v^{z^N}$ for some z^N and this yields:

$$P\left(v^{z^N-1}\left|X\right.\right)\leq N\left(u^{\lfloor vN\rfloor}\left|X\right.\right)/N\leq P\left(v^{z^N}\left|X\right.\right) \tag{125}$$

It is thus enough to show that both $P\left(v^{z^N-1}\left|X\right.\right)$ and $P\left(v^{z^N}\left|X\right.\right)$ tend to $n\left(v\left|X\right.\right)$ to prove our result. We have the following decomposition:

$$\begin{aligned}P\left(v^{z^N}\left|X\right.\right)-n\left(v\left|X\right.\right) &= \left[P\left(v^{z^N}\left|X\right.\right)-P_Z\left(v^{z^N}\left|X\right.\right)\right]+\left[P_Z\left(v^{z^N}\left|X\right.\right)-n_Z\left(v^{z^N}\left|X\right.\right)\right]\\ &\quad+\left[n_Z\left(v^{z^N}\left|X\right.\right)-n\left(v\left|X\right.\right)\right]\end{aligned} \tag{126}$$

where z^N is such that $v^{z^N-1}\leq\lfloor vN\rfloor/(N-1)\leq v^{z^N}$.

It can be shown that the first right hand-side term converges to zero as N and Z tend to infinity using result (81). We have already shown that the second right-hand side term converges to zero provided that the number of individuals in each interval, which is approximately N/Z tends to infinity. Finally, the third right-hand side term converges to zero between v^{z^N} converges to v and $\beta_Z(u)$ converges to $\beta(u)$ uniformely over the $[0,1]$ interval. Indeed, $n_Z\left(v\left|X\right.\right)$ and $n\left(v\left|X\right.\right)$ verify the same differential equation at infinity, which is given by (6), and this equation has a unique solution for the initial conditions $n(X)$. This yields that $P\left(v^{z^N}\left|X\right.\right)$ converges to $n\left(v\left|X\right.\right)$. A similar proof can be applied to show that $P\left(v^{z^N-1}\left|X\right.\right)$ converges to $n\left(v\left|X\right.\right)$, and we finally get that $N\left(u^{\lfloor vN\rfloor}\left|X\right.\right)/N$ converges to $n\left(v\left|X\right.\right)$.

C Theoretical foundations of the simulation approach

The finite discrete counterpart of the differential equation verified by the measures of available individuals (6) can be rewritten in vector form piling up (21) in the (X,j) dimension as:

$$\overrightarrow{N}^*\left(u^k\right)=\overrightarrow{N}^*\left(u^{k+1}\right)-\overrightarrow{D}_{k+1} \tag{127}$$

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:

$$E\left[\overrightarrow{N}^*\left(u^k\right)\right] = \overrightarrow{N} - E\left(\overrightarrow{M}_{k+1}\right) \tag{128}$$

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:

$$E\left(\overrightarrow{M}_k\right) = E_{\overrightarrow{M}_k, \overrightarrow{M}_{k+1}, ..., \overrightarrow{M}_N}\left(\overrightarrow{M}_k\right) \tag{129}$$

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:

$$P\left(\overrightarrow{M}_k = \overrightarrow{m}_k, \overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}, ..., \overrightarrow{M}_N = \overrightarrow{m}_N\right) \tag{130}$$

$$= \left[\prod_{\ell=k+1}^{N-1} P\left(\overrightarrow{M}_k = \overrightarrow{m}_k \left|\overrightarrow{M}_{k+1} = \overrightarrow{m}_{k+1}\right.\right)\right] P\left(\overrightarrow{M}_N = \overrightarrow{m}_N\right) \tag{131}$$

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:

$$P\left(1\left|X, j\right.\right) = \frac{N(X, j)\exp\left[X\beta_j^*(1)\right]}{\sum_{\ell,g} N\left(X^\ell, g\right)\exp\left[X^\ell\beta_g^*(1)\right]} \tag{132}$$

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 $\overrightarrow{M}_k\left|\overrightarrow{M}_{k+1} = \overrightarrow{m}^s_{k+1}\right.$ is simple because $\overrightarrow{m}^s_{k+1}$ contains all the information necessary to determine the realization of $\overrightarrow{\Omega}^*(u)$ that we denote $\overrightarrow{\Omega}^s(u)$. We have:

$$P\left(\overrightarrow{M}_k = \overrightarrow{m}_k\left|\overrightarrow{M}_{k+1} = \overrightarrow{m}^s_{k+1}\right.\right) = P\left[\overrightarrow{D}_\ell = \overrightarrow{m}_k - \overrightarrow{m}^s_{k+1}\left|\overrightarrow{\Omega}^*\left(u^k\right) = \overrightarrow{\Omega}^s\left(u^k\right)\right.\right] \tag{133}$$

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:

$$P^s\left(u^k\left|X,j\right.\right)=\frac{N^s\left(u^k\left|X,j\right.\right)\exp\left[X\beta_j^*\left(u^k\right)\right]}{\sum_{\ell,g}N^s\left(u^k\left|X^\ell,g\right.\right)\exp\left[X\beta_g^*\left(u^k\right)\right]} \tag{134}$$

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 (132), 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 (134). 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 $\overrightarrow{N}-\sum_{s=1}^S\overrightarrow{M}_{k+1}^s/S$.

D 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 occuring 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.

E Additional results

E.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 $[E(D|f,u) - E(D|m,u)]\,\beta^r_D(u)$, where $\beta^r_D(u)$ is the polynomial coefficient of the category dummy. Figure E.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 $E(D|f,u)\left[\beta_{fD}(u) - \beta^r_D(u)\right] - E(D|m,u)\left[\beta_{mD}(u) - \beta^r_D(u)\right]$, where $\beta_{jD}(u)$ is the gender-j polynomial coefficient of the category dummy. Figure E.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.

E.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 E.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).

E.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 E.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%

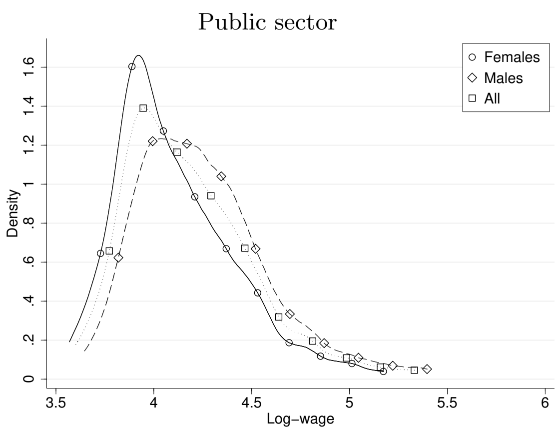
Note: Figures in the table correspond to proportions except % difference which is the female value minus the male value divided by the male value.

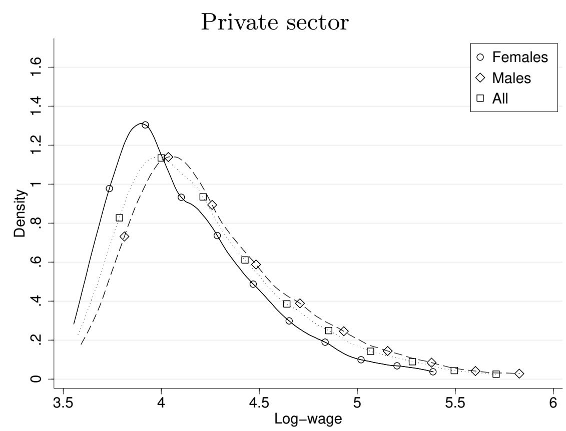
Table 3: Observed and counterfactual log-wage gaps obtained in different scenarios

	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.

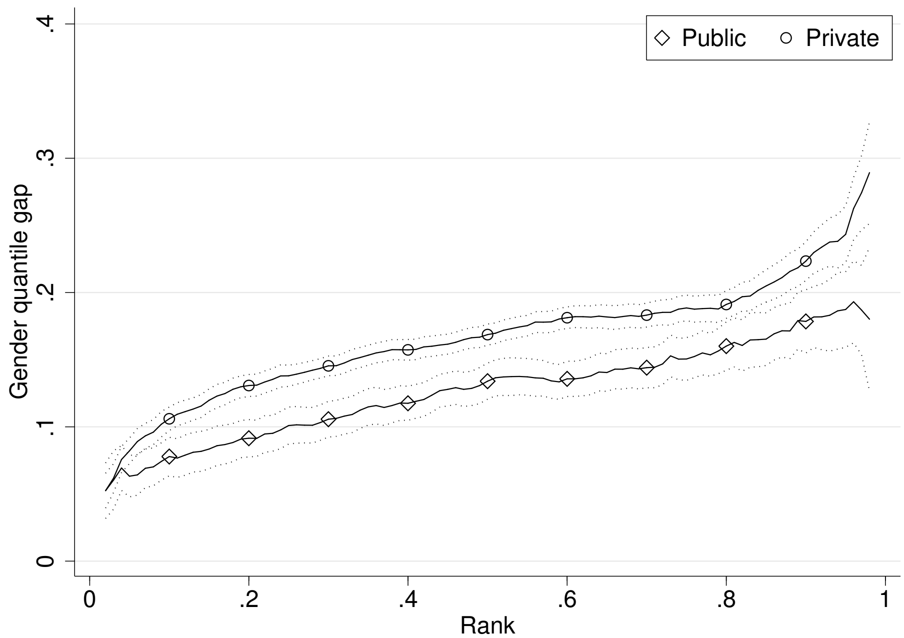
Figure 1: Gender log-wage distributions in the public and private sectors





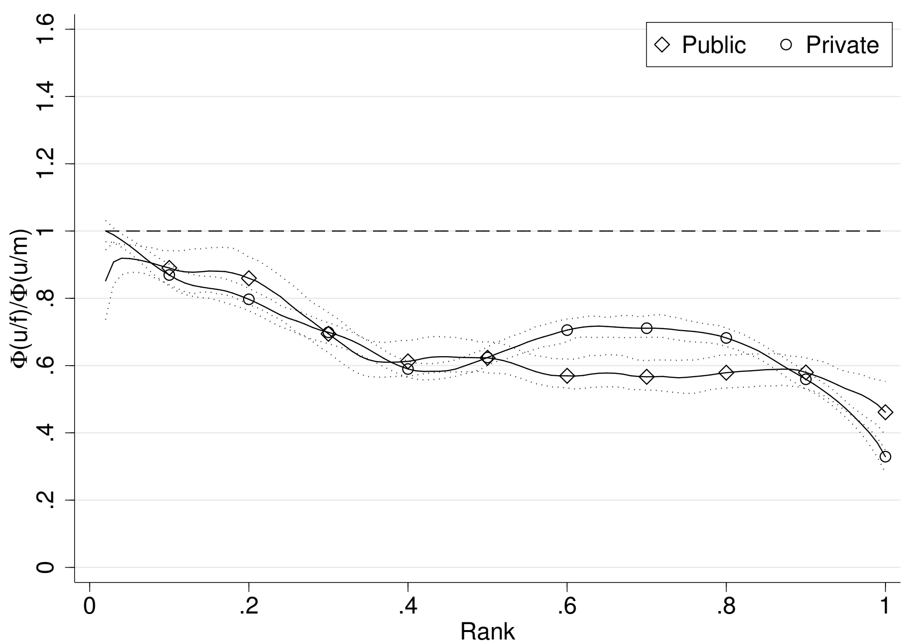
Note: Densities are computed for the logarithm of daily wage.

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.

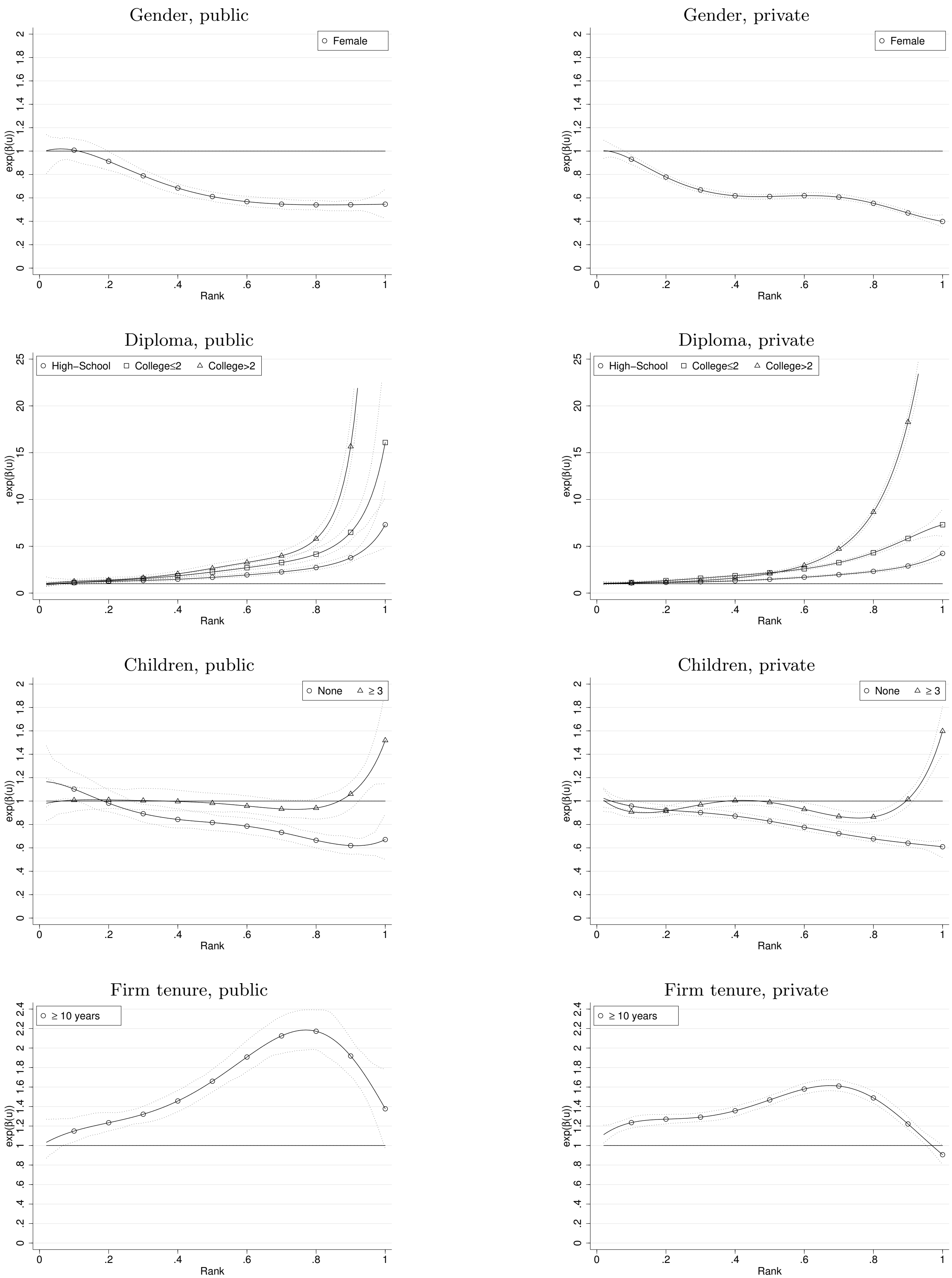
Figure 3: Gender probability ratio of getting a given job position in the public and private sectors



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

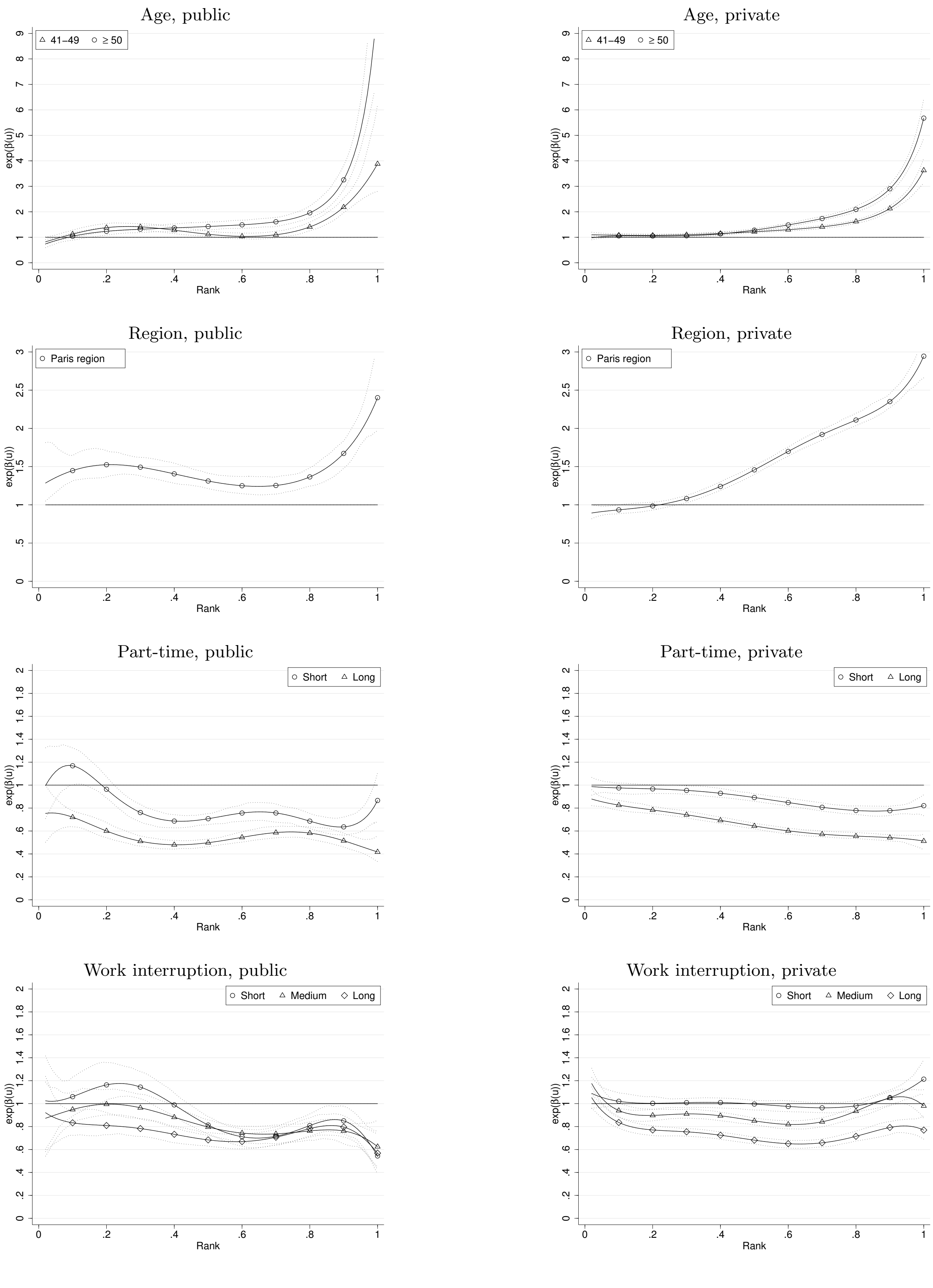
Figure 4: Exponentiated effects of category dummies on individual values

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.

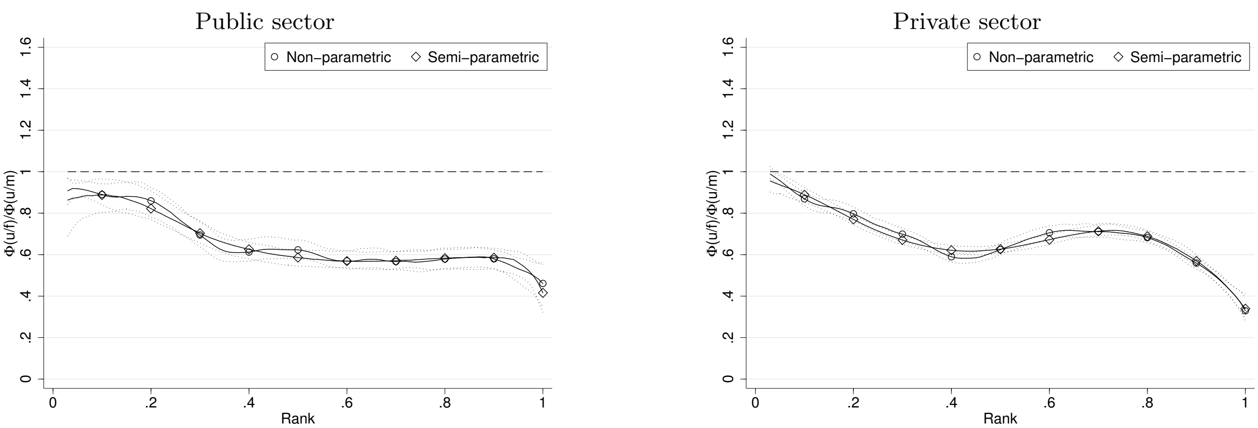
Figure 4 (Cont.): Exponentiated effects of category dummies on individual values for each individual characteristic



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.

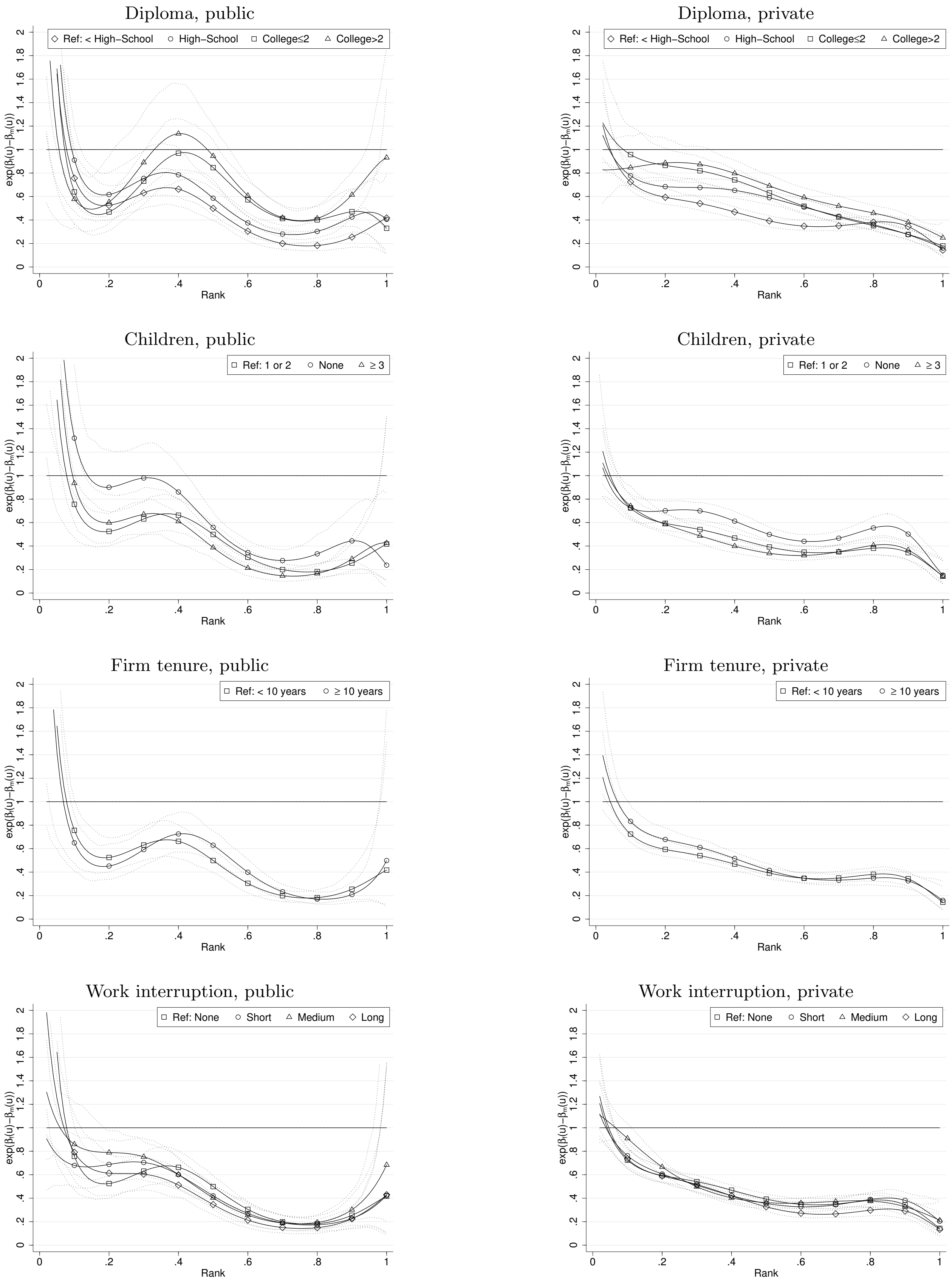
Figure 5: Non-parametric and semi-parametric estimators of

the gender probability ratio of getting a given job position



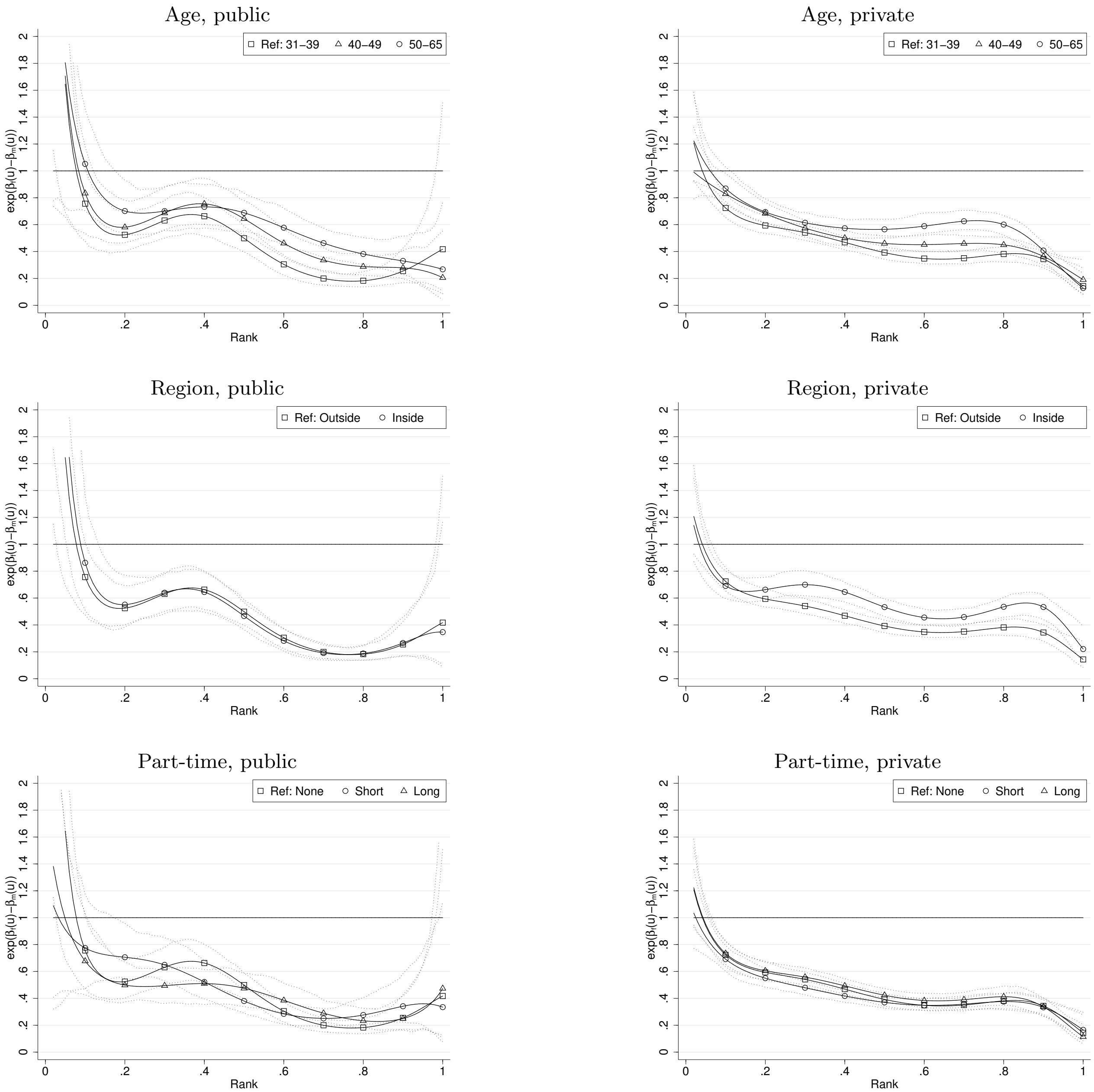
Note: The non-parametric estimator is obtained by applying the empirical strategy proposed by Gobillon, Meurs and Roux (2015). The semi-parametric estimator is obtained by applying the empirical strategy proposed in the current paper. The confidence interval at the 5% level obtained by bootstrap using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



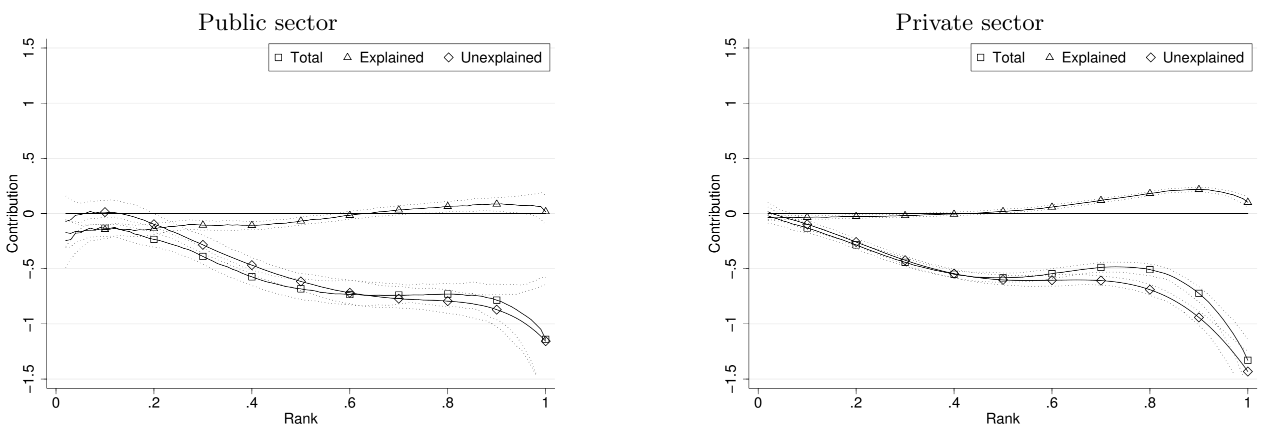
Note: The graph title gives the individual characteristic for which the exponentiated gender differences in the effects of category dummies are graphed, as well as the sector. The category corresponding to the reference is mentioned and the corresponding curve is the same across all graphs for a given 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



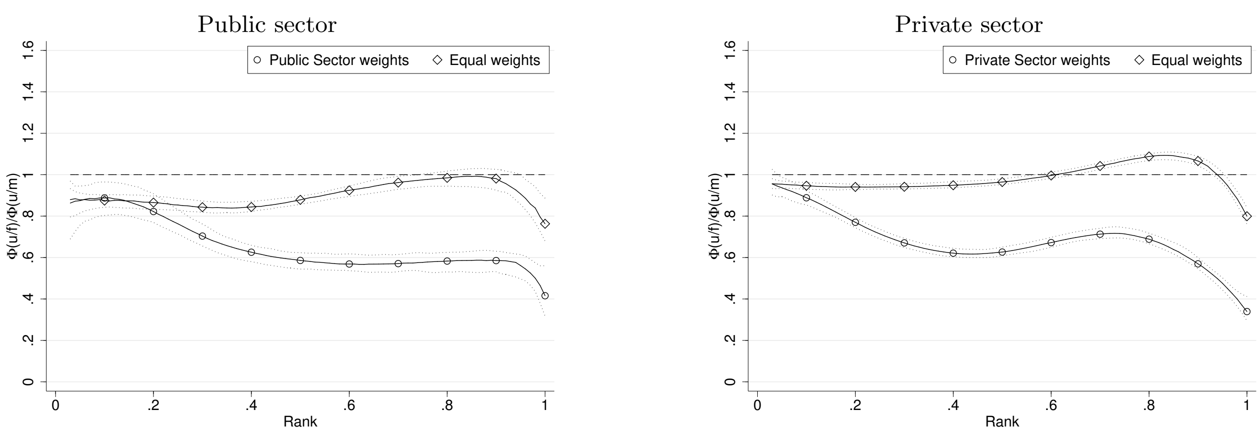
Note: The graph title gives the individual characteristic for which the exponentiated gender differences in the effects of category dummies are graphed, as well as the sector. The category corresponding to the reference is mentioned and the corresponding curve is the same across all graphs for a given 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



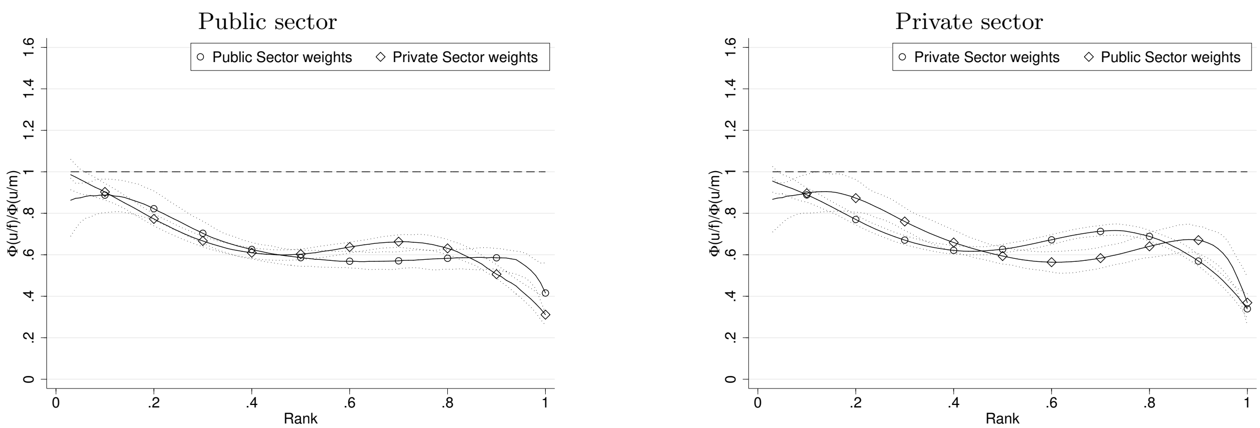
Note: "Total": gender difference in the logarithm of the average probability of getting a given job position; "Explained": part of "Total" that can be attributed to the gender difference in observable characteristics valued using the estimated coefficients obtained for the whole population; "Unexplained": 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



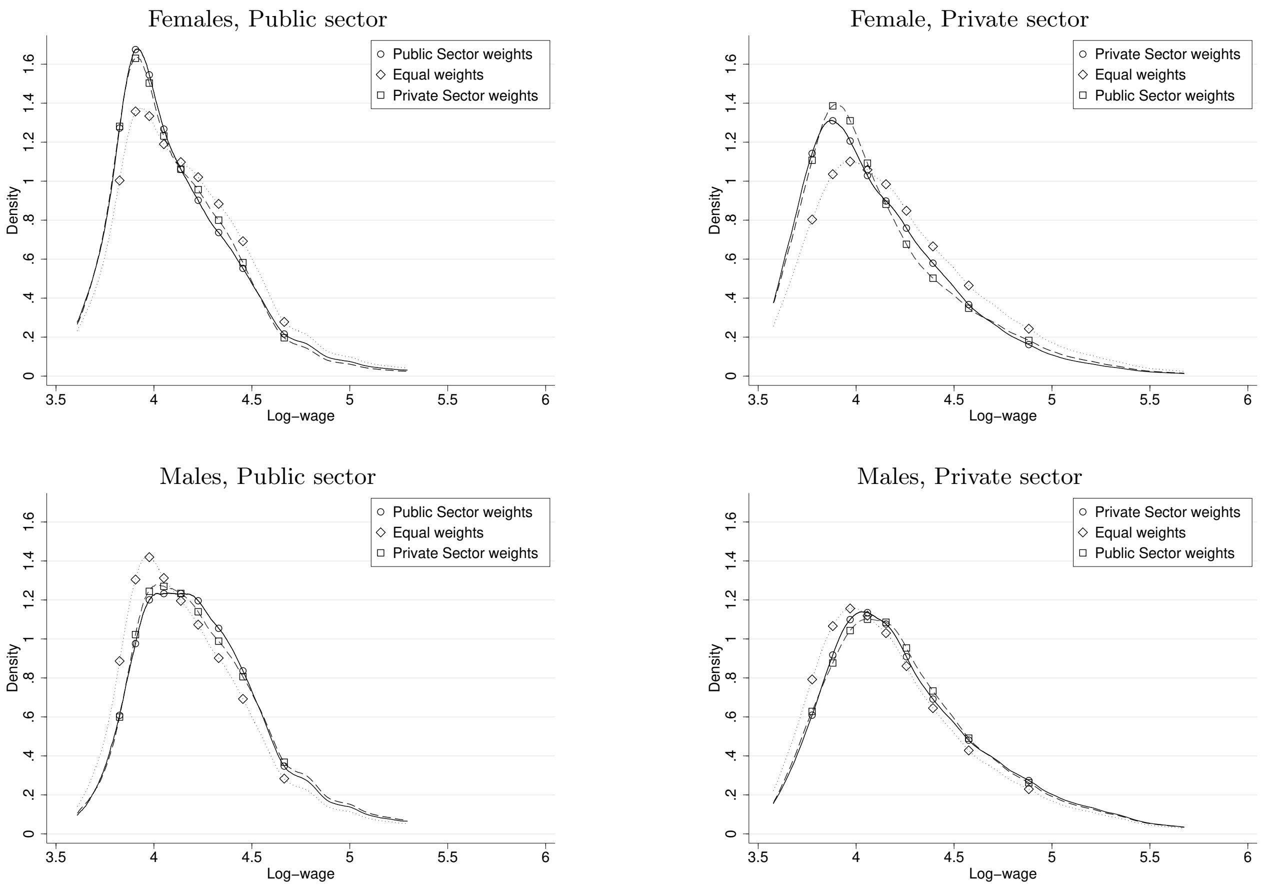
Note: The curve corresponding to "Equal weights" is obtained by fixing conditional individual weights to the same values for the two genders, these common values being obtained by estimating parameters for the whole population. The curve corresponding to "Private Sector weights" (resp. "Public Sector 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



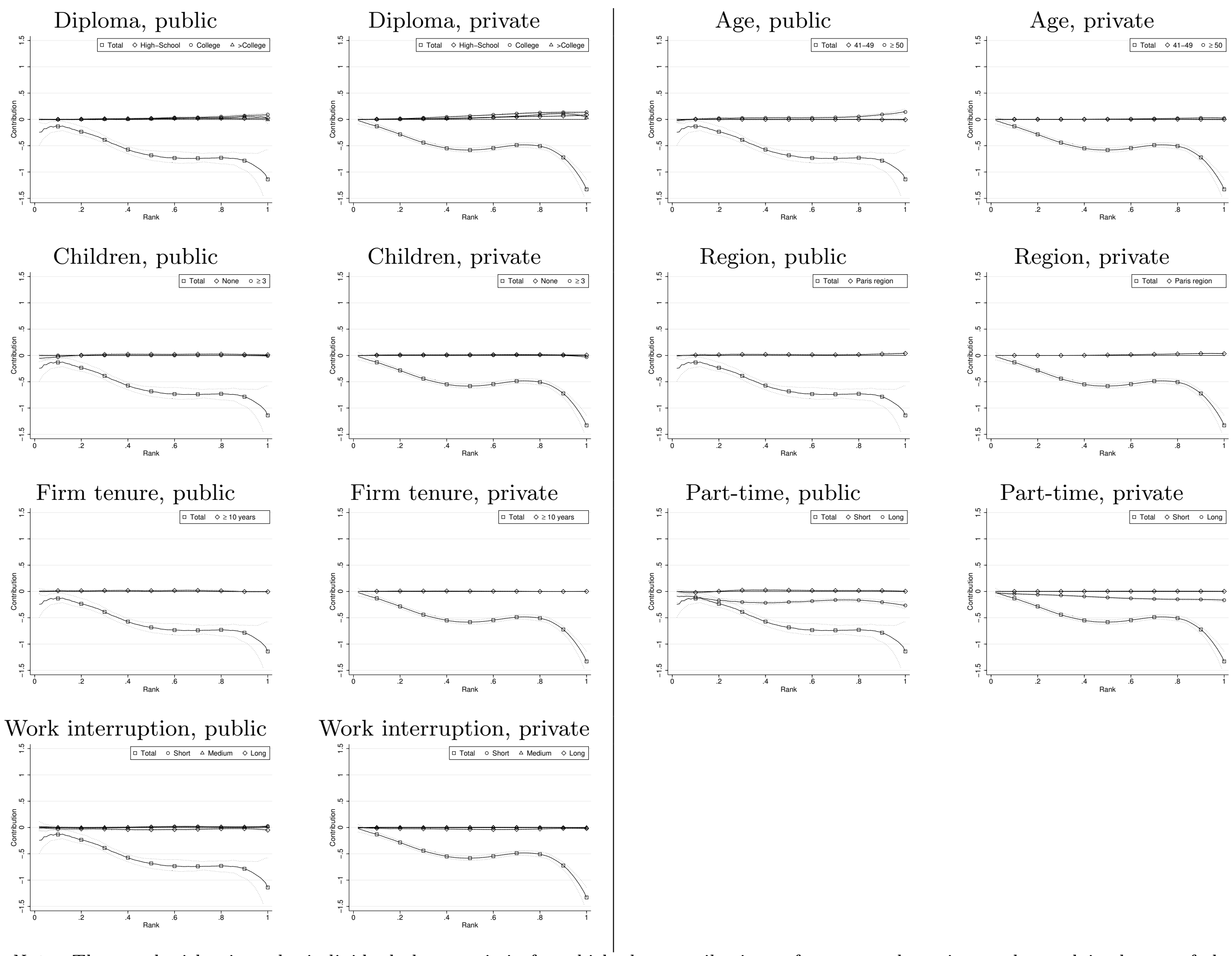
Note: Densities are computed using the logarithm of daily wages generated by the model when reassigning workers to job positions. "Private Sector weights": reassignment using conditional individual weights computed for the private sector; "Public Sector weights": reassignment using conditional individual 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 E.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)

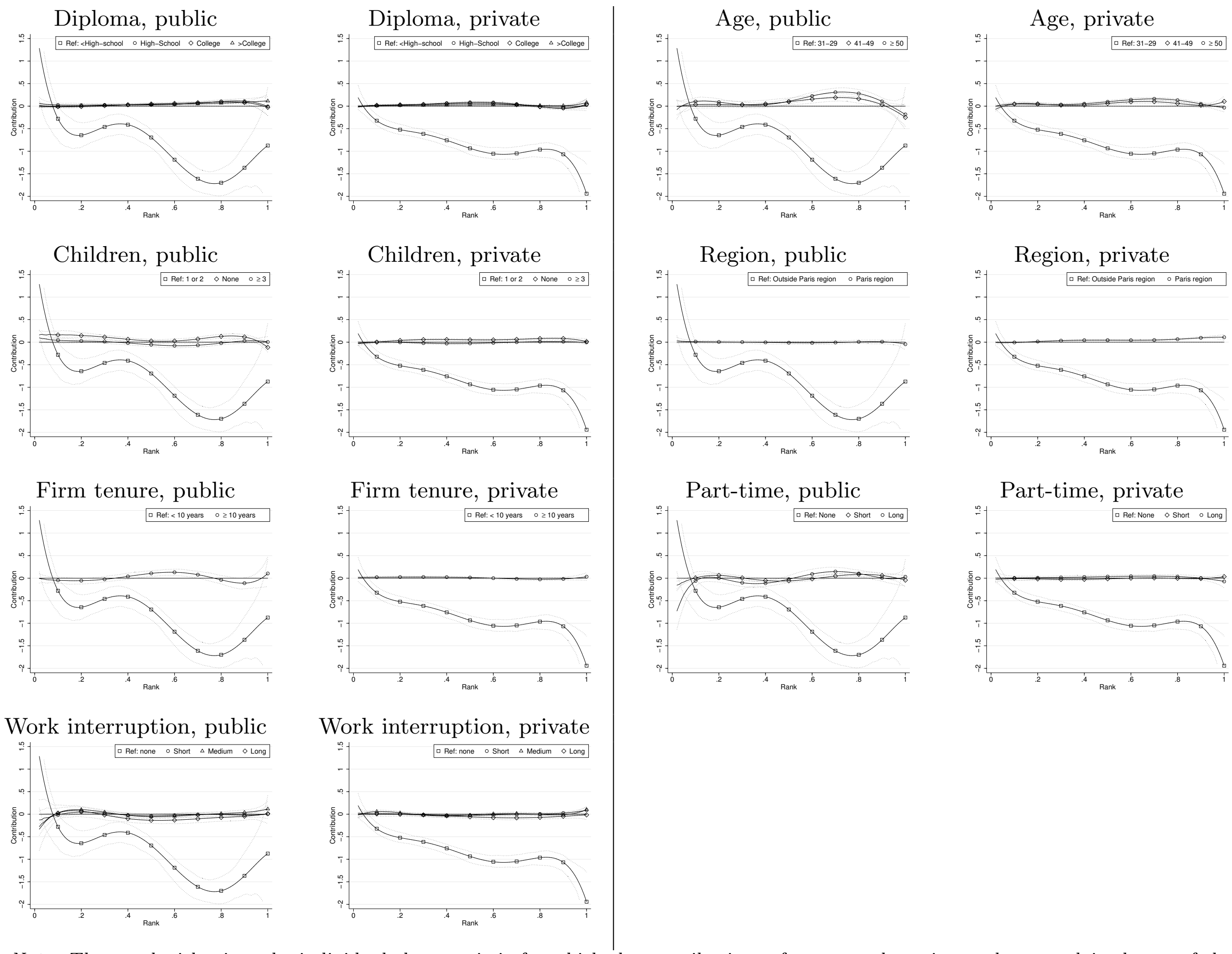
Note: Statistics are computed on the logarithm of hourly wage. Column headings mention either that statistics are computed directly from the data (label "Observed") or that they are derived from counterfactuals using the conditional individual weights of the public sector, the private sector, or the 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 E.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



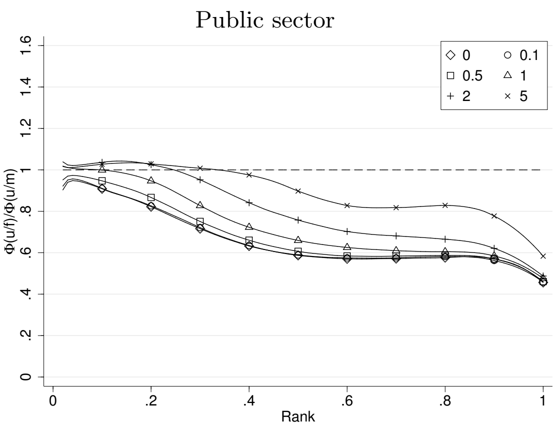
Note: The graph title gives the individual characteristic for which the contributions of category dummies to the explained part of the gender probability ratio are graphed, as well as the sector. "Total": gender difference in the logarithm of the average probability of getting a given job; other label 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).

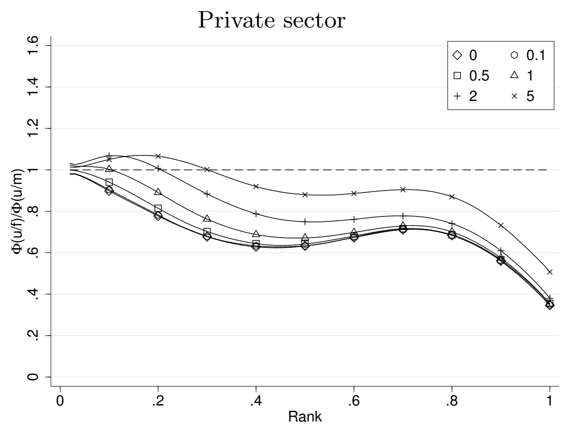
Figure E.2: Decomposition of the unexplained part into the contribution of every category dummy for each individual characteristic



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 E.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).