

IZA DP No. 6928

Estimating Gender Differences in Access to Jobs

Laurent Gobillon
Dominique Meurs
Sébastien Roux

October 2012

Estimating Gender Differences in Access to Jobs

Laurent Gobillon

INED, PSE, CEPR and IZA

Dominique Meurs

University of Paris Ouest Nanterre (EconomiX) and INED

Sébastien Roux

DARES and CREST

Discussion Paper No. 6928
October 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Estimating Gender Differences in Access to Jobs^{*}

This paper proposes a new measure of gender differences in access to jobs based on a job assignment model. This measure is the probability ratio of getting a job for females and males at each rank of the wage ladder. We derive a non-parametric estimator of this access measure and estimate it for French full-time executives aged 40-45 in the private sector. Our results show that the gender difference in the probability of getting a job increases along the wage ladder from 9% to 50%. Females thus have a significantly lower access to high-paid jobs than to low-paid jobs.

JEL Classification: J16, J31, J71

Keywords: gender, discrimination, wages, quantiles, job assignment model, glass ceiling

Corresponding author:

Laurent Gobillon
Institut National d'Etudes Démographiques (INED)
133 Boulevard Davout
75980 Paris Cedex 20
France
E-mail: laurent.gobillon@ined.fr

^{*} We thank Joseph Altonji, Christian Bontemps, Allison Booth, Moshe Buchinski, Pierre Dubois, Francis Kramarz, Thierry Magnac, Ronald Oaxaca, Patrick Puhani, Jean-Marc Robin, Bernard Salanié and two anonymous referees for their useful comments, as well as the participants of seminars and conferences at the University of Hanover, INED (Paris), CREST (Paris), University of Toulouse, Paris School of Economics, University of Paris Ouest Nanterre, World Congress of the Econometric Society (Shanghai), ESEM (Barcelona), EALE (Tallinn). All remaining errors are ours.

1 Introduction

Females are under-represented in high-paid occupations and this fact explains a large part of the gender wage differential. In this paper, we propose a formal way to quantify the gender difference in access to each position along the job ladder and to assess whether females have a lower access to high-paid jobs than to low-paid jobs.

Since Albrecht, Bjorklund and Vroman (2003), quantile regressions have been used to estimate the difference between the conditional wage distributions of males and females. It is said to be a “glass ceiling effect” when this gap gets wider in the upper tail. This widely used approach is fruitful to highlight the gender bias in promotions but it remains purely descriptive.

We propose a more precise measure of this gender bias which can be justified with a microfounded framework. Our work builds on the literature on job assignment models which posits the existence of heterogenous job positions (see Sattinger, 1993; Teulings, 1995; Fortin and Lemieux, 2002; Costrell and Loury, 2004). In our framework, each position is characterized by a specific wage offer to applicants. Some male and female workers apply for the best-paid job. Males and females are characterized by different chances of getting this specific job. The manager of the best-paid job selects a worker among applicants accordingly. The manager of the second best-paid job hires an individual among the remaining workers, and so on. We define an access function as the probability ratio of females and males getting a job of a given rank.

In an empirical section, we estimate this access function on full-time executives aged 40 – 45 working in French private and public firms. We find that at the bottom of the wage distribution (5th percentile), the probability of females getting a job is 9% lower than that of males. The difference in probability is far larger at the top of the wage distribution (95th percentile) and climbs to 50%. Females thus have a significantly lower access to high-paid jobs than to low-paid jobs.

Our analysis is related to a few empirical studies which investigate the gender difference in promotions. Blackaby, Booth and Frank (2005) find a significant gender promotion gap on the UK academic labour market after controlling for productivity differences. Pekkarinen and Vartiainen (2006) show on Finnish data that among blue-collar workers, females have to reach a higher productivity threshold to get promoted than males. This kind of research is limited to case studies as it exploits detailed information on the individual ranking along the job ladder in a specific sector.

Here, we consider the whole private sector and postulate that the wage is a reasonable proxy for the position in the job hierarchy: a higher wage corresponds to a better position. The wage has already been used by Bertrand and Hallock (2001) to define the group of top corporate jobs which is shown to include only 2.5% of women. In fact, results obtained by Baker, Gibbs and Holmstrom (1994) on a single firm support the fact that the rank of a position in the job hierarchy is correlated with the wage, although the wage is not strictly attached to the job.

Killinsworth and Reimers (1983) argue that neither the type nor the rank of a position is perfectly indexed by the wage. This is particularly true for blue collars, for whom wages increase significantly with job tenure. Also, some blue collars occupy jobs which are paid at the minimum wage but do not correspond to the same hierarchical position. Hence, we restrict our attention to executives whose wage reflects more closely the rank along the job ladder. We only analyse full-time workers aged 40 – 45 who should constitute a rather homogenous population. Additionally, the effect of supply side factors for female full-time executives aged 40 – 45 should be rather limited.¹

Our results show that females have a lower access to jobs than males at all ranks in the wage distribution. Moreover, their access to jobs decreases along the job ladder. We also conduct our analysis for specific industries as they constitute more homogenous labour markets. We consider more specifically banking and insurance as they are labour intensive with a large share of females and have contrasting wage policies in France. Banks rely on a rigid job classification inherited from the early eighties when they belonged to the public sector. By contrast, insurance companies propose much more individualized careers. Females are found to have a far better access to high-paid jobs than males in the banking industry than in the insurance industry. Results on pooled industries, insurance and banking are qualitatively in line with the usual interpretation of the gender quantile wage difference increasing with the rank.

We then stratify our estimation of the access function by group defined from age, labour market experience or country of birth (France or foreign country). Results are very similar across groups. This is in line with our use of an homogeneous population. We also make an alternative assumption on the size of the labour market, assuming that the competition of workers for jobs occurs within each firm rather than on the national market. We estimate the average access function across large firms employing more than 150 full-time executives aged 40 – 45. When pooling all industries, results are quite similar to those obtained when competition is supposed to occur on the national market. For the specific insurance industry, results are a bit different, as we find that the gender difference in access to high-paid jobs is slightly less pronounced. This change is generated by some heterogeneity in the level of wages among firms. Finally, we show that our results on the lower access of females to high-paid jobs are robust when workers occupying lower ranking jobs than executive positions are included in the sample.

In the last section, the access function is micro-founded with three theoretical mechanisms that can lead to a lower access of females to high-paid jobs. The first one is based on gender differences in labour supply. We consider that males and females have *ex-ante* the same distribution of productivity but are heterogenous in their labour supply. We show that the access function is decreasing if the difference in job application rates between females and males increases along the job ladder. The second mechanism is based on taste discrimination (Becker, 1971).

¹It can still be argued that, for a given job position, females bargain less for the wage than males and may end up earning less at the same hierarchical level (Babcock and Lashever, 2003). However, the effect on wages of the gender difference in bargaining power should be rather small in our case, as French wages scales are based on a job classification system negotiated at the industry level. The actual wage depends strongly on the ranking of the position in the job classification and is independent of gender.

It is straightforward to show that the access function is decreasing if the preference for females relative to males decreases along the job ladder. Finally, the third mechanism is a version of statistical discrimination where the distribution of skills is the same for males and females but skills are observed with more noise for females than for males (Aigner and Cain, 1977; Coate and Loury, 1993; Cornell and Welch, 1996).² We show that statistical discrimination can then lead to a decreasing access function. This can occur because skilled males are more often hired for high-paid jobs as their skills are observed with less uncertainty. Males looking for less-paid jobs are those not selected for high-paid positions and they are on average less skilled than females. As a consequence, the hiring rate of males for less-paid jobs is lower than for high-paid jobs. The three mechanisms can occur simultaneously and the empirical access function captures the aggregate of the three explanations.^e

The rest of the paper is organized as follows. In section 2, we present our framework. Our econometric strategy to estimate the access function is detailed in section 3. We then describe our dataset and report some stylized facts in section 4. We comment our estimation results in section 5. We conduct some robustness checks when taking into account the individual observed heterogeneity and segmented markets in section 6. Some additional robustness checks such as the inclusion in the sample of workers unable to access an executive position are proposed in section 7. In section 8, we present three theoretical mechanisms that can generate a decreasing access function. Concluding remarks are given in the last section.

2 The framework

2.1 Our model

We propose a job assignment framework that allows to consistently measure the gender differences in access to jobs. Consider a countable number of workers applying for a countable number of job positions. There is a proportion n_m of males in the whole population of workers, which we refer to as the *measure* of males for clarity hereafter, and a measure $n_f = 1 - n_m$ of females. For simplicity, we assume that there is a bijection between workers and job positions. This assumption rules out the existence of workers not being hired and job positions not being filled. It will be relaxed in Section 7.

A worker is primarily interested in the job yielding the highest wage. Job positions are heterogeneous such that each job position is associated with a specific fixed wage through a contract. This wage is not allowed to depend on the gender of the applicant. We suppose that two job positions cannot be associated with the same wage offer so that each job can be uniquely identified by its *rank* in the wage distribution.³ Workers first consider the best

²Another version of statistical discrimination posits a gender difference in average productivity (Arrow, 1971; Phelps, 1972) but we think it is more interesting to show that a decreasing access function can be obtained without differences in the intrinsic characteristics of the male and female subpopulations.

³The wage distribution is supposed to be exogenous. We could introduce some mechanisms on the labour market to endogenize this

ranked position as it offers the highest wage. Those who are not hired turn to the second best ranked position, and so on.

For any job position of given rank u , there is a measure denoted $n_j(u)$ of gender- j workers considering the job (ie. all the gender- j workers not hired for jobs of higher rank). We have $n_j(1) = n_j$ as all workers consider the first job and $n_j(0) = 0$ as all workers end up being employed.

We can then determine for each gender j a differential equation which is verified by the measure of available workers at each rank. Consider an arbitrarily small interval du in the unit interval. The proportion of jobs in this small interval is du since ranks are equally spaced (and dense) in the unit interval. The instantaneous probability of a gender- j worker getting a job in this interval is denoted $\phi_j(u)$. This probability can vary with gender for various reasons such as differences in labour supply (a different share of males and females considering that work conditions are too constraining to apply), taste discrimination against females or statistical discrimination against females. We discuss in Section 8 how the instantaneous probability can be micro-founded theoretically to embed these three explanations.

It is possible to relate the instantaneous probabilities of getting a job with the process through which jobs are filled. The measure of jobs occupied by workers of a given gender j is $n_j(u) \phi_j(u) du$. For this gender, the measure of workers available for a job of rank $u - du$ can be deduced from the measure of workers available for a job of rank u subtracting the workers who get the jobs of ranks between $u - du$ and u :

$$n_j(u - du) = n_j(u) - n_j(u) \phi_j(u) du \tag{1}$$

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

$$n'_j(u) = -\phi_j(u) n_j(u) \tag{2}$$

This relationship states that the instantaneous variation in the measure of gender- j workers around rank u depends on the stock of gender- j workers and their chances of getting a job.

We now introduce a measure of relative access to jobs. Consider one male worker and one female worker available for a job position of given rank u . These two workers have different chances of being hired as they are not of the same gender. We can characterize the gender difference in access to jobs with the function defined as the probability ratio of a female and a male being hired for the job position:

$$h(u) \equiv \frac{\phi_f(u)}{\phi_m(u)} \tag{3}$$

This function $h(\cdot)$ measures the gender relative access to jobs and we label it the “access function”. When the access function takes the value one at all ranks, males and females have the same chances of getting each job

distribution but it is beyond the scope of this paper as the wage setting is not relevant to our empirical approach.

position. When the access function takes a value lower than one for a job position of given rank, females have less chances than males of getting the job.

It is then possible to formally define a lower access to high-paid jobs than to low-paid jobs for females:

Definition 1 *Females have a **lower access to high-paid jobs** than to low-paid jobs if there are some ranks u_0 and u_1 such that for any $u \in]u_0, u_1[$ and $v > u_1$, we have $h(u) > h(v)$ and $h(v) < 1$.*

For instance, females have a lower access to high-paid jobs than to low-paid jobs when the access function is continuous, strictly decreasing and takes some values lower than one at the highest ranks. It is also the case when the access function is a two-step function with the second step at a value lower than the first step and lower than one. In particular, an interpretation of the specification where the second step takes a zero value is the existence of a glass ceiling.⁴ In practice, a lower access to high-paid jobs can be contrasted to a uniform access to all jobs when the access function is constant at all ranks.

2.2 Gender quantile differences

The recent empirical literature on the glass ceiling has focused on the difference between the quantiles of the wage distributions of males and females. Typically, when this difference increases with the rank, it is usually said that there is a glass ceiling. We believe that this approach confuses two dimensions, the job position and the associated wage, possibly leading to inaccurate interpretations. Figure 1 presents a simple scheme illustrating this point. Suppose a classic job ladder where the wage increases more than proportionally with the rank in the wage distribution. Positions are occupied alternately by a female and a male. The gender quantile difference for high-paid jobs is larger than for low-paid jobs, which means that the gender wage gap widens along the job ladder. It is tempting to conclude that there is a glass ceiling effect (according to the loose labelling of the literature), but this interpretation is questionable as the odds of a female (or a male) occupying a position are roughly constant along the job ladder.

[Insert Figure 1]

In fact, this interpretation does not rest on any straightforward rationale and has two caveats. First, it does not control for the spacing between wages and thus mixes the ranks of positions on the job ladder with earnings. Second, the rank at which quantiles are computed has a different meaning for the two genders. For males, it corresponds to the rank in the wage distribution of males. For females, it corresponds to the rank in the wage distribution of females.

⁴Very often in the literature, the glass ceiling is more loosely defined. It is considered that there is a glass ceiling effect when females' access to jobs is particularly low for top positions.

We now show that it is possible to generate from our framework a gender quantile difference which increases with the rank even if there is no glass ceiling and the difference in access to jobs between males and females is the same at all ranks. We first solve the model when the access function is constant with $h(u) = .672$ at all ranks and the proportion of females is that of the banking industry (28.7%).⁵ The numerical solution allows to compute $v_j(u) = \frac{n_j(u)}{n_j}$ as well as $u_j = v_j^{-1}$ which gives for a job of given rank in the wage distribution of gender j , its rank in the wage distribution of job positions.⁶ We can then relate the quantile function of gender j denoted $\lambda_j(\cdot)$ to the quantile function of job positions $\lambda(\cdot)$ through the relationship: $\lambda_j(v) = \lambda[u_j(v)]$. The gender quantile difference is given by:

$$(\lambda_m - \lambda_f)(v) = \lambda[u_m(v)] - \lambda[u_f(v)] \quad (4)$$

We can compute the gender quantile difference using the solution $u_j(\cdot)$ of the model and the wage distribution of job positions in the banking industry for $\lambda(\cdot)$. The gender quantile difference represented on Figure 2 is an increasing function above rank .6. Whereas the increase is small just above that rank, the curve becomes very steep above rank .9. The literature would conclude to a glass ceiling whereas there is none.

[*Insert Figure 2*]

Hence, economic interpretations should rather rely on the primitive function of a model which is the access function in our case. We now propose an econometric approach to estimate the access function non parametrically from the data.

3 Estimation strategy

We now show how the access function can be estimated non-parametrically from a cross-section dataset containing for each worker some information on gender and wage. First recall that the access function can be reinterpreted as the unit probability ratio of females and males getting a given job. From equation (2), each unit probability can be rewritten as:

$$\phi_j(u) = \frac{n'_j(u)}{n_j(u)} \quad (5)$$

⁵These choices are made clear in the empirical section. Indeed, we will show that the gender difference in access to jobs is nearly uniform in the banking industry and that the access function takes values close to .672 at all ranks.

⁶We determine numerically for each gender the measure of available workers at each rank at the equilibrium. For females, we use the algorithm proposed by Bulirsch and Stoer (for the implementation, see Press et al., 1992, p. 724-732) to solve the differential equation giving $n_f(\cdot)$. As all jobs are filled, we have: $n_f(u) + n_m(u) = u$. Deriving this equality, we get: $n'_m(u) = 1 - n'_f(u)$. Using (2) for the two genders, we get: $n'_f(u)/n'_m(u) = h(u)n_f(u)/n_m(u)$, which is equivalent to $\frac{n'_f(u)}{1 - n'_f(u)} = h(u)n_f(u) / [u - n_f(u)]$. Rewriting this expression to get an expression of $n'_f(u)$, we obtain: $n'_f(u) = \frac{n_f(u)h(u)}{u - n_f(u) + n_f(u)h(u)}$. This differential equation is solved backward from the highest to the lowest rank using the initial condition $n_f(1) = n_f$. After the differential equation for females has been solved, we deduce the solution for males using the relationship $n_m(u) = u - n_f(u)$. It is then straightforward to compute $v_j(\cdot)$ and its inverse from the solution.

We introduce for gender- j workers, the random variable corresponding to their rank in the wage distribution of job positions, U_j . The cumulative (resp. density) of this variable is denoted F_{U_j} (resp. f_{U_j}). The cumulative verifies the relationship: $F_{U_j}(u) = n_j(u)/n_j$. Hence, each unit probability can be rewritten as:

$$\phi_j(u) = \frac{f_{U_j}(u)}{F_{U_j}(u)} \quad (6)$$

The numerator and denominator of the gender- j unit probability only depend on the distribution of ranks of gender- j workers in the wage distribution of job positions. This means that in practice, the ranks of workers of each gender in the distribution of job positions are sufficient to estimate the unit probabilities, and thus the access function. These ranks can be computed very easily from the data.

For each gender, the numerator and denominator of the unit probability of getting a job can be evaluated using the usual Rosenblatt-Parzen Kernel estimator of the density $f_{U_j}(\cdot)$ which is given by:

$$\tilde{f}_{U_j}(u) = \frac{1}{\omega_j N_j} \sum_{i|j(i)=j} K\left(\frac{u-u_i}{\omega_j}\right) \quad (7)$$

where $K(\cdot)$ is a Kernel, ω_j is the bandwidth, $j(i)$ is the gender of individual i and u_i is his/her rank in the wage distribution of job positions. A standard estimator of the cumulative $F_{U_j}(\cdot)$ is then given by:

$$\tilde{F}_{U_j}(u) = \int_0^u \tilde{f}_{U_j}(u) du = \frac{1}{N_j} \sum_{i|j(i)=j} L\left(\frac{u-u_i}{\omega_j}\right) \quad (8)$$

where $L(u) = \int_{-\infty}^u K(v) dv$. However, we expect the estimators $\tilde{f}_{U_j}(\cdot)$ and $\tilde{F}_{U_j}(\cdot)$ to perform quite poorly for ranks close to zero and one as standard non-parametric estimators are asymptotically biased at the boundaries (see Härdle, 1990).

Therefore, we rather consider some estimators based on local polynomial approximations in line with Jones (1993) and Lejeune and Sarda (1992). The density and cumulative functions evaluated at a given rank u are obtained by minimizing:

$$\min_{\alpha, \beta} \int_0^1 K\left(\frac{u-v}{\omega_j}\right) [g_j(v) - \alpha - \beta(u-v)]^2 dv \quad (9)$$

where g_j is either the empirical density function defined as $g_j(v) = N_j^{-1} \sum_{i|j(i)=j} \delta_{u_i}(v)$ where $\delta_u(v)$ is the Dirac delta function associated to v , or the empirical cumulative function defined as $g_j(v) = N_j^{-1} \sum_{i|j(i)=j} 1_{\{u_i \leq v\}}$. The estimators of the density and cumulative functions at rank u , denoted $\hat{f}_{U_j}(u)$ and $\hat{F}_{U_j}(u)$, are given by the value of α derived from their respective minimization program. Here, the boundary problem is taken into account by simply fixing the bounds of the integral in (9) to zero and one.⁷ The bias at these boundaries is of order $O(\omega_j^2)$ and tends to zero when the bandwidth tends to zero.

⁷If, instead, the bounds of the integral in (9) were fixed to $-\infty$ and $+\infty$, the estimators of the density and cumulative would turn out to be the usual kernel estimators given by (7) and (8).

In our application, the Kernel is chosen to be Epanechnikov and we experiment with values of bandwidth which are multiples of the value given by the rule of thumb (Silverman, 1986). For a bandwidth taking the value given by the rule of thumb, there are not enough observations at small ranks for the density and cumulative to be estimated properly. We thus decided to use a bandwidth which is equal to two times the value given by the rule of thumb.⁸ Note that the Epanechnikov kernel takes the value zero outside the interval $[-\omega_j, \omega_j]$. It is possible to show that because of this property, the estimator of the density obtained from (9) takes exactly the same value as the estimator given by (7) for ranks in the interval $[\omega_j, 1 - \omega_j]$. In fact, the two estimators for the density computed at points in this interval coincide because there is no boundary problem.⁹ Technical details to obtain some analytical formulas for the estimators are given in Appendix A.

For gender j , an estimator of the unit probability of getting a job can then be computed as $\hat{\phi}_j(u) = \hat{f}_{U_j}(u) / \hat{F}_{U_j}(u)$. We finally obtain an estimator of the access function:

$$\hat{h}(u) = \frac{\hat{\phi}_f(u)}{\hat{\phi}_m(u)} \quad (10)$$

This estimator is computed for a grid of 1000 ranks in $[0, 1]$ which are equally spaced. The confidence interval of the access function at each rank is computed by bootstrap with replacement (100 replications).

It is also possible to adopt a parametric specification of the access function and estimate the parameters from a minimization program as shown in Appendix B. More specifically, we consider the linear specification $h(u) = a - b.u$ where a is the access to jobs at the lowest rank and b captures a decrease or an increase in access to jobs with the rank. A test of the linear specification using the Cramer Van-Mises statistic is also provided.

4 Descriptive statistics

4.1 The data

The wage distribution of job positions, as well as the males' and females' wage distributions, are constructed from the *Déclarations Annuelles de Données Sociales* (DADS) or Annual Declarations of Social Data database. These data have been collected from employers for tax purposes every year since 1994. They are *exhaustive* for all private and public firms. For each job, the data contain some information on the industry, contract type (full-time/part-time), daily wage, socio-occupational category, age, sex and country of birth (France/foreign country¹⁰) of the employee. A limitation is that the education level of employees is not reported.

⁸Results obtained for different values of the bandwidth are available upon request.

⁹Note that it is not exactly the case for the cumulative as it is computed using values close to boundary 0.

¹⁰For workers born in a foreign country, our data do not allow us to distinguish which country it is.

As our model is static, we consider the single year 2003. For that year, there are 20,599,456 jobs in 1,599,865 firms. We want to restrict our attention to a subpopulation of workers for whom the assumptions of our model are more likely to be verified. Because of the minimum wage, some blue collars and clerks may be paid the same wage although they are ranked differently along the job ladder. Also, the length of tenure has an important effect on the wage of blue collars even if they do not move to another job position. We discard low-skilled workers from our analysis to avoid these issues and rather focus on workers with an executive job position (business managers, top executives, engineers and marketing staff). There are 2,173,975 executive job positions in 318,852 firms.

We want to study similar subpopulations of males and females competing for the same positions. For that purpose, we restrict our sample to executives aged 40 – 45 working full time. Executive females still on the market at those ages usually have not experienced career interruptions, are more career-oriented and compete for jobs with males. Labour supply factors such as family duties should thus play a limited role in the analysis. Having a range of only six years for age limits the cohort effects.

Table 1 shows that for the 40 – 45 age bracket, there are 354,968 executive job positions in 86,989 firms. 22.4% of the executives are females. As reported in Table A.1, the average age is the same for males and females (42.4 years) and the proportion of workers born in a foreign country is very close at 11% for males and 10% for females.¹¹ The wage distribution is skewed to the right and the mean daily wage (139 euros) is higher than the median daily wage (109 euros). The dispersion is very large and the standard error of wages stands at 602 euros.

[*Insert Table 1*]

Figure 3 represents the log-wage densities of male and female full-time executives aged 40 – 45 working in a private or public firm. It shows that wages of males are more dispersed and that high wages are more frequent than for females. The gender difference in median wage is as large as 17 euros. In line with the literature on quantile regressions, Figure 4 represents the quantile wage difference as a function of rank. Males have a higher wage than females at every rank and the gap widens as the rank increases. Using the gender quantile values reported in table 1, we obtain a wage difference of 14% at the bottom of the distribution (5th percentile) which increases to 26% at the top of the distribution (95th percentile).

[*Insert Figures 3 and 4*]

We will take a closer look at two industries: banking and insurance. These industries share some similarities as they are both labor intensive and both employ a high proportion of executives. Also, the proportion of females is above the average, which is quite usual in the service industries: 28.7% (resp. 36.9%) of executives are females in banking (resp. insurance). This ensures that there is a large pool of female executives competing for promotion with their

¹¹As shown in Appendix D with the 2003 French Labour Survey, the distributions of diplomas for our subpopulations of males and females are rather close.

male counterparts. The common organizational features of the two industries contrast with the differences in their wage structure. The wage dispersion is far larger in the banking industry than in the insurance industry. There is also a far larger gender gap in median wage for insurance (21 euros) than for banking (13 euros).

It is of particular interest to study the females' access to high-paid jobs in these two industries as economic performance in these industries relies heavily on the quality of human resource management (Bartel, 2004). We represented the gender quantile difference as a function of rank for the two industries in Figures 5 and 6. Whereas the gender quantile difference starts increasing for insurance at low ranks, it increases only above rank .8 for banking. Using the gender quantile values reported in Table 1, we obtain a wage difference for insurance (resp. banking) of 13% (resp.11%) at the bottom of the distribution (5th percentile) which increases to 53% (resp. 33%) at the top of the distribution (95th percentile).

[*Insert Figures 5 and 6*]

5 Results

Figure 7 represents the estimator of the access function \hat{h} and the confidence interval at each rank of the wage distribution of job positions. Recall that $\hat{h}(u)$ can be interpreted as the gender probability ratio of getting a job at rank u . When $\hat{h}(u) > 1$, females have a better access to the job than males. When $\hat{h}(u) < 1$, males have a better access to the job than females. As the access function takes values which are always lower than one, the probability of getting a job at any rank is lower for females than for males. However, the values are close to one for the first ranks, indicating that females and males have almost the same access to lower-paid jobs. For instance, the probability of females getting a job at rank .05 is only 9% lower than the probability of males as shown in Table 2. Between ranks .2 and .8, the access to job slightly decreases for females compared to males. After rank .8, the access function decreases more sharply, pointing to a far lower female access to the best-paid jobs. The probability of females getting a job at rank .95 is 50% lower than the probability of males.

[*Insert Figure 7*]

[*Insert Table 2*]

We now look at banking and insurance, which are closely related as shown by the recent take-overs across these two industries. These industries have different wage policies. Banks rely on a job classification and on regulations which are quite rigid as they are inherited from the period when most banks belonged to the public sector. Insurance companies give more weight to the individualization of careers (Dejonghes and Gasnier, 1990). We find that there is a sharp contrast in the access function between the two industries. For insurance, it decreases sharply from rank .6 to rank 1, pointing to a lower female access to high-paid jobs than to low-paid jobs (Figure 8). For banking, the

access function decreases only at the highest rank and the overall pattern points to a uniform lower access to jobs for females (Figure 9).

We can assess more accurately the extent to which females have a lower access to high-paid jobs than to low-paid jobs from the slope of the access function. Indeed, the steeper the slope, the larger the difference in female access to high-paid jobs and to low-paid jobs. We thus estimate a linear specification of the access function, $h(u) = a - b.u$, and compare the value of the slope parameter b for the pooled industries, banking and insurance (see Appendix B for details on the procedure). For the pooled industries, an increase of one decile in the wage distribution of job positions ($\Delta u = .1$) yields a decrease in access to jobs of females relative to males of 2.8% ($b = .28$) as shown in Table 3. Whereas the decrease is smaller in the banking industry, at .7%, it is more than two times larger in the insurance industry, at 6.0%. Interestingly, a statistical test shows that the linearity of the access function is not rejected at the five percent level for the pooled industries, nor for banking and insurance. As the slope of the access function is small in the banking industry for ranks between .1 and .9, we tried to approximate the access function of that sector with a constant specification: $h(u) = \gamma$. The constant is estimated to be .672 and the specification is not rejected at the five percent level. Hence, the access function in the banking sector is nearly constant except at the extremes.

[*Insert Figures 8 and 9*]

[*Insert Table 3*]

Overall, our results on pooled industries, banking and insurance suggest a lower female access to high-paid jobs than to low-paid jobs.

6 Individual and market heterogeneity

So far, we have considered that workers are heterogenous only in the gender dimension, in the sense that all workers of a given gender have *ex-ante* the same chances of getting a job of a given rank. However, in our data, workers also differ in age and country of birth, and there is no reason why their access to jobs cannot be influenced by these factors.¹² As a consequence, we estimate access functions for different subgroups of the population and assess whether our results are robust across subgroups.¹³

We represent the access functions for the age-40 group and the age-45 group in Figure 10. The two access functions have a very similar profile although the gender gap in access to jobs is slightly lower for workers aged 40

¹²Workers also differ in their qualifications and labour market experience, which are likely to affect their chances of getting hired but are not available in our dataset. We will assess the effect of labour market experience in Section 7 for a subsample of workers for whom it is possible to construct a measure of labour market experience.

¹³This amounts to assessing the effect of individual variables using a non-parametric approach. A semi-parametric approach is developed in Gobillon, Meurs and Roux (2011).

than for workers aged 45. We also represent the access functions for workers born in France and workers born in a foreign country in Figure 11. The two access functions exhibit a very similar profile although the gender gap in access to jobs is slightly lower for workers born in France than for workers born in a foreign country.

[*Insert Figures 10 and 11*]

We also implicitly assumed that all workers compete on the national market. This is arguable as some individuals spend their whole career in a large firm which can be considered as an internal market. We can redefine the access function under the alternative assumption that each firm is a separate market and workers within each firm compete with each other but not with outsiders. We first compute non parametrically an access function for each firm using the estimation procedure developed in Section 3. An overall access function is then defined as a weighted average of the access functions of firms where the weight is the number of workers in the firm.¹⁴ Alternatively, we adopt a parametric approach and consider that the access function of each firm follows a linear specification where the parameters are common to all firms. An estimation procedure and a test of the linear specification are provided in Appendix B.

For this robustness check, we limit our sample to large firms employing at least 150 full-time executives aged 40 – 45. Indeed, many workers getting their first job in a large firm spend their entire career in that firm. As shown in Table 1, only $175/86989 * 100 = .5\%$ of firms in our sample are large, but they employ 33% of workers. The median wage in large firms is 114 euros, which is a bit higher than for the whole sample (109 euros). By contrast, the wages are far less dispersed, with a standard deviation of 132 euros compared to 602 euros for the whole sample.

Figure 12 shows that the average access function when workers compete on each submarket has a profile quite similar to the access function when all workers of large firms compete on a common national market. The similarity between the two curves is confirmed when evaluating some linear specifications of the access function in the two cases. The estimated specifications are respectively $h(u) = .74 - .09u$ and $h(u) = .69 - .05u$ which are very close. Interestingly, our linear specification is rejected only when competition occurs on the national market and not when it occurs on segmented submarkets. This difference arises because females are often employed by firms offering low wages and we condition out the heterogeneity among firms only when workers compete on segmented submarkets. Indeed, we conduct a within-firm estimation in the spirit of what is done for linear panel data models.

[*Insert Figure 12*]

We performed the same exercise for the insurance and banking industries. For banking, the average access function when competition occurs on each segmented submarket has a profile similar to the access function when competition occurs on the national market. Curves seem to differ for insurance. We estimated a linear specification of the two access functions to facilitate the comparison. We obtained for insurance, respectively, $h(u) = .93 - .66u$ when

¹⁴See Gobillon, Meurs and Roux (2011) for more details.

competition occurs on the national market and $h(u) = .74 - .41u$ when it occurs on each segmented submarket. Hence, the access function begins at a lower level when competition occurs on each segmented submarket but its slope is less steep, suggesting that the lower female access to high-paid jobs is less pronounced. Once again, the difference between the two access functions can be explained by some heterogeneity in the level of wages among firms and the sorting of females in firms offering low wages. We control for this heterogeneity only when competition is supposed to occur within firms. In any case, the gender difference in access to high-paid jobs is larger in the insurance industry than in the banking industry and for pooled industries. Our results are thus qualitatively robust to the assumption on the extent of the market where workers compete for jobs.

7 Robustness checks

We now assess the robustness of our results to some additional changes in our assumptions.

We argued that our population of interest is homogeneous but this could be questioned as the labour market experience is missing from our analysis. Females may have accumulated less experience on the labour market than males because of career interruptions to raise children. For this reason, firms could prefer to select males rather than females for the best positions. We can evaluate to what extent this type of mechanism plays a role for 4% of workers for whom we have a panel over the 1984-2003 period.¹⁵ A comparison of Table A.2 with Tables 1 and A.1. shows that descriptive statistics by gender for the whole dataset and for the 4% subsample are quite close.¹⁶

We use the panel information to construct a measure of experience defined as the number of days spent in a firm since 1984. The median experience for our whole sample is 16.5 years. We define two subgroups depending on whether workers have experience below or above the median. We assess whether our results are robust for each experience group. This would point to a limited effect of experience on gender differences in access to high-paid jobs. Figure 13 shows that the profile of the access function is very similar to the one obtained for the whole population.

[*Insert Figures 13*]

¹⁵We use the panel data of the *Déclaration des Données Sociales* (DADS) which contains some panel information on all workers born in October of an even year over the 1976-2003 period (see Abowd, Kramarz and Roux, 2006, for more details). There is no information for the years 1981 and 1983 because all the resources of the French Institute of Statistics were devoted to the 1982 census. We thus only consider the information from 1984 onwards. The information is also missing for the census year 1990 and we did not take into account that year in our computation of the labour market experience.

¹⁶The average age, the proportion of workers born in a foreign country and the median wage are slightly higher for males and females in the subsample than in the whole sample.

Overall, our findings support our claim that the population is rather homogenous. In particular, this result is driven by the restriction of the analysis to full-time workers. When part-time workers are included in the analysis, controlling for experience lowers the gender difference in access to jobs.

Another limit to our analysis is that in our model, all workers get an executive position. This assumption usually does not hold as some workers may decide not to apply for an executive position because of work conditions and some applicants may not be selected for any executive position. These workers are employed in intermediate positions.¹⁷ It is possible to extend the model to include males and females who do not get an executive job. For a given job of rank u in the wage distribution, it is possible to decompose the measure of gender- j workers available for the job $n_j(u)$ into the measure of workers not occupying any executive job n_{0j} and the measure of workers not yet hired for an executive job but who will be hired $\tilde{n}_j(u)$: $n_j(u) = n_{0j} + \tilde{n}_j(u)$. The cumulative of gender- j ranks now verifies: $F_{U_j}(u) = \tilde{n}_j(u) / \tilde{n}_j$ where \tilde{n}_j is the measure of workers occupying an executive job. Using equation (5), the gender- j unit probability of getting the rank- u job can be rewritten as:

$$\phi_j(u) = \frac{f_{U_j}(u)}{\omega_j + F_{U_j}(u)}$$

where $\omega_j = n_{0j} / \tilde{n}_j$ is the ratio between the measure of workers not occupying an executive job and the measure of workers occupying an executive job. The access function is given again as the ratio between the unit probabilities of females and males.

In the computation of the ratios ω_j , we use the whole population of workers holding an intermediate position as a proxy for the population of workers not hired for an executive position. Indeed, these workers belong to the socio-occupational category just below executive. As one can argue that some of these workers should not be considered as they do not have the skills for becoming executives, we also considered alternatively the subpopulation of workers holding intermediate positions for at least ten years during their career. We expect these workers to have accumulated enough experience to apply for an executive position. In this category, we miss, however, the workers with less experience but enough skills to become executives. The true ratio ω_j probably lies between the bounds obtained using our two alternative definitions of workers not being selected for executive positions. The ratios are found to be $\omega_f = 1.651$ and $\omega_m = 1.334$ when resorting to the whole population of workers occupying an intermediate position as a measure of workers not hired for executive positions. As the ratios are larger than one, there are more workers occupying an intermediate position than workers occupying an executive position. The ratios drop to $\omega_f = .735$ and $\omega_m = .607$ when restricting the sample to workers having at least ten years of experience in intermediate positions. Nevertheless, even in this case, the ratios are large and always higher for females than for males. This means that, as expected, a larger proportion of females occupy an intermediate position.

¹⁷Intermediate positions include: intermediate positions in the health sector or social work sector, intermediate administrative and marketing positions in firms, technician positions and supervisor positions.

Interestingly, the access function at rank zero takes the value $(\omega_m/\omega_f) / (f_{U_m}(0)/f_{U_f}(0))$ and is thus proportional to ω_m/ω_f . Hence, the access function at rank zero will be very slightly lower when using all workers occupying an intermediate position as a measure of workers without an executive position ($\omega_m/\omega_f = .808$) than when using only the workers with at least ten years of experience in intermediate positions ($\omega_m/\omega_f = .826$).

The access functions obtained when considering only workers occupying an executive position and when also considering the two alternative groups of workers not occupying an executive position are represented in Figure 14. The two access functions obtained when also considering workers not occupying an executive position are below one and decreasing. They remain very close to the access function when considering only workers occupying an executive position for ranks above .5. Results on a lower female access to high-paid jobs would thus be robust to alternative assumptions on the extent to which all workers occupy an executive job.

[*Insert Figures 14*]

8 Theoretical foundation of the approach

We now show how the gender access function can be founded theoretically with some simple models of gender differences in labour supply, taste discrimination of managers against females, and statistical discrimination against females. In particular, all these mechanisms can generate a decreasing access function in line with our results. For simplicity, we present these three mechanisms separately, but they can easily be embedded in one single model as explained at the end of the section. We do not argue that their relative role can be identified separately from the data, but rather that they can all contribute to explaining a decreasing access function.

When considering gender differences in labour supply and taste discrimination, we assume that the productivity of all workers whatever their gender is the same. Indeed, no productivity heterogeneity is needed for these mechanisms as they are not related to productivity. By contrast, we suppose that there is some productivity heterogeneity when considering statistical discrimination as, in that case, the choice of workers by firms is based on the observation of the productivity with error. We show that even when the distribution of skills is the same for males and females, it is possible to obtain a decreasing access function. This can happen when the measurement error on female productivity has a larger variance than the measurement error on male productivity.

The recruitment process is similar in the three cases. For a job of rank u , a manager screens the applicants. He derives a utility $V_i(u)$ from an applicant i whose form depends on the mechanism. The manager hires the applicant yielding the highest utility. The set of applicants to the job is the set of workers not hired for a job of higher rank and interested in the job. This set can be defined recursively as:

$$\Omega(u) = \left\{ i \text{ applying for the rank-}u \text{ job} \mid \text{for all } \tilde{u} > u, i \notin \Omega(\tilde{u}) \text{ or } V_i(\tilde{u}) < \max_{k \in \Omega(\tilde{u})} V_k(\tilde{u}) \right\} \quad (11)$$

8.1 Participation

We consider that males and females have *ex-ante* the same productivity. However, males and females are heterogeneous in their labour supply. Some females may find the work conditions for some jobs too constraining more often than males, especially for well-paid jobs which usually involve long working hours and investment not compatible with family constraints.

We thus consider that, for a given job of rank u , there is an exogenous share $1 - v_j(u)$ of available gender- j workers who decide not to apply for the job.¹⁸ The measure of applicants for the job is then $v_j(u)n_j(u)$. We suppose that $v_j(u) > 0$: there are always some workers interested in the position. The utility derived by a manager from an applicant i is a random match quality $V_i(u) = \varepsilon_i(u)$ corresponding to the profit associated with the job. The manager is supposed to observe the match quality so that he can evaluate exactly how much profit he can make from the job if he hires the applicant. The manager hires the applicant yielding the highest profit. The set of applicants for the job $\Omega(u)$ contains all the workers who did not apply for the jobs above rank u because of work constraints or did not draw a match quality high enough to get selected for those jobs.

The resulting allocation of workers is a Nash equilibrium. Workers have no incentive to move from their position. This is because the worker occupying the best position has no incentive to move to a less-paid job. The worker occupying the second best position cannot move to the best position as it is already occupied. Hence, he has no incentive to move, and so on. Also, managers have no incentive to fire an employee as they cannot find a better worker on the market. We assume that at the equilibrium, there is a bijection between workers and job positions so that all job positions are filled and all workers are employed.

For a given job, it is possible to determine a closed formula for the probability that the selected worker is of gender j under some additional assumptions. The maximization program of the manager is a multinomial model with two specificities. First, the choice set consists in all applicants still available after better ranked job positions have been filled. There would be a selection process based on match qualities if the match quality of the workers available for the job was correlated with their match quality for better ranked jobs. We suppose that the match qualities are drawn independently across jobs to avoid this kind of selection mechanism.¹⁹ Second, the choice set contains an infinite but countable number of workers. We extend the extreme value assumption on the law of residuals that is associated with a logit specification to an infinite countable number of jobs following Dagsvik (1994).²⁰ In particular, this assumption ensures that for any given job, the probability of selecting a worker in any

¹⁸In particular, this setting corresponds to the case where the decision to apply for the job is drawn randomly and independently across workers in the same binomial law of parameter $v_j(u)$.

¹⁹This assumption is extreme as it yields that profits made from a given worker are not correlated across jobs. However, it is easy to generate a correlation with some unobserved heterogeneity in productivity among workers as shown later when dealing with statistical discrimination.

²⁰For any job of given rank u , the share of applicants of gender j is given by $\frac{v_j(u)n_j(u)}{v_f(u)n_f(u)+v_m(u)n_m(u)}$. We assume that the points of

given finite subgroup of available applicants follows a logit model. Under this assumption, the following formula is verified by the probability that the worker chosen for the job of rank u is of gender j :

$$P(j(u) = j) = n_j(u) \phi_j(u) \quad (12)$$

with

$$\phi_j(u) = \frac{v_j(u)}{n_f(u)v_f(u) + n_m(u)v_m(u)} \quad (13)$$

The denominator of this probability is a competition term which depends on the measures of available workers of the two genders as well as labour supply effects (preferences for work conditions). Competing workers are available workers of each gender who decide to apply. Some sufficient conditions under which (2) has a solution when instantaneous probabilities of getting a job verify (13) are given in Appendix C.²¹ From this formula, we can rewrite the access function as:

$$h(u) = \frac{v_f(u)}{v_m(u)}$$

The access function is simply the ratio of the application rates of females and males for the job positions along the wage distribution. The access function is decreasing if the application rate of females relative to males decreases as the rank of the job position in the wage distribution increases. By contrast, it is constant if the gender ratio of application rates is constant along the wage distribution.

8.2 Taste discrimination

We now assume that all males and females available for a job of a given rank decide to apply for the job (ie. the application rate is equal to one for each gender), but that there can be some gender-specific taste discrimination by the manager when screening the applicants. We now suppose that the manager chooses an applicant based not only on profit but also on the effect of the applicant's gender on his well-being. The manager hiring a worker for a job of rank u derives a utility from an applicant i which is:

$$V_i(u) = \ln \theta_{j(i)}(u) + \varepsilon_i(u) \quad (14)$$

where $\theta_j(u)$ captures the taste of the manager for gender j . Note that the tastes are allowed to vary along the job ladder. The manager chooses the applicant who grants him the highest level of utility in the set of applicants for the job, $\Omega(u)$, which contains all the workers who did not draw a match quality high enough to get selected for jobs above rank u .

As previously, the resulting allocation of workers is a Nash equilibrium. Also, it is possible to determine for a given job, a closed formula for the probability that the selected worker is of gender j under the same assumption

the sequence $\{j(i), \varepsilon_i(u)\}$, $i \in \Omega(u)$ are the points of a Poisson process with intensity measure $\frac{v_j(u)n_j(u)}{v_f(u)n_f(u) + v_m(u)n_m(u)} \exp(-\varepsilon)d\varepsilon$.

²¹In fact, conditions are the same for the three mechanisms separately or taken together, and we will only present the existence theorem for the full model presented in Subsection 8.4.

as in the case of gender differences in labour supply. The probability that the worker chosen for the job of rank u is of gender j verifies (12) where the unit probability of a gender- j worker getting the job is now given by:

$$\phi_j(u) = \frac{\theta_j(u)}{n_f(u)\theta_f(u) + n_m(u)\theta_m(u)} \quad (15)$$

The denominator of this probability is a competition term which depends on the measures of available workers of the two genders as well as taste discrimination which can make workers of a given gender more likely to be hired than workers of the other gender. From this formula, we can rewrite the access function as:

$$h(u) = \frac{\theta_f(u)}{\theta_m(u)}$$

The access function is simply the preference of the managers for females relative to males for the job positions along the wage distribution. The access function is decreasing if the preference of the manager for males relative to females increases as the rank of the job position in the wage distribution increases. By contrast, it is constant if the preferences of managers are the same whatever the position of the jobs in the wage distribution.

8.3 Statistical discrimination

We now assume that all males and females available for a job of a given rank decide to apply for that job (ie. the application rate is equal to one for each gender) and there is no taste discrimination. However, each worker now has some specific skills. We also consider that the skills of applicants are only observed with some noise by managers in line with Aigner and Cain (1977).

Each individual i is now characterized by skills s_i affecting his productivity and the profit of job u occupied by an individual i is now given by:

$$\Pi_i(u) = s_i + \varepsilon_i(u) \quad (16)$$

When a worker applies to the job, the manager only observes the worker's skills with some noise such that the skills perceived can be written as $y_i(u) = s_i + \eta_i(u)$ where $\eta_i(u)$ is a random error term specific to the job of rank u . Note that this error term is not equivalent to the match-specific quality $\varepsilon_i(u)$. Indeed, whereas the manager observes the match quality, he does not observe the error term. We consider that the manager is risk neutral such that his utility is the expected profit. As the manager does not observe the true value of skills, he computes his expected profit conditionally on the information he has on skills. This information is the observed productivity and the fact that the applicant is still in the set of available workers of gender $j(i)$. As the match quality is also observed, the expected profit (ie. utility of the manager) is:

$$V_i(u) \equiv E[\Pi_i(u) | y_i(u), \varepsilon_i(u), i \in \Omega_{j(i)}(u)] = E[s_i | y_i(u), i \in \Omega_{j(i)}(u)] + \varepsilon_i(u) \quad (17)$$

This formula embeds the case where there is some skill heterogeneity perfectly observed by the manager, in which case $y_i(u) = s_i$ and thus $E[s_i | y_i(u), i \in \Omega_{j(i)}(u)] = s_i$. The conditional expected profit for a given individual is

then positively correlated across jobs with the covariance being equal to $V(s_i)$. When there is some noise in the skills observed by the managers, there is still a positive covariance due to skills, but the formula is more intricate because of the filtering process when the rank decreases and the noise which varies across jobs. We make the same additional assumption on the discrete choice model for the selection of an applicant as in the case of taste discrimination such that for any given job, the probability of selecting a worker in any given finite subgroup of available applicants follows a logit model. Equation (12) is verified with:

$$\phi_j(u) = \frac{\Psi_j(u)}{n_f(u)\Psi_f(u) + n_m(u)\Psi_m(u)} \text{ with } \Psi_j(u) = E_{y_i(u)}[\exp E[s_i | y_i(u), i \in \Omega_j(u)]] \quad (18)$$

Here, $\Psi_j(u)$ is the average (exponentiated) expected skills for gender- j workers conditional on the available information at rank u . We finally obtain:

$$h(u) = \frac{\Psi_f(u)}{\Psi_m(u)} \quad (19)$$

The gender with the highest average expected skills has a better access to jobs.

Suppose that the distribution of true skills is the same for males and females, and follows a normal law with mean ω and variance σ^2 for each gender.²² Suppose also that the noise on skills has the same distribution at all ranks for each gender but has a higher dispersion for females. More specifically, the noise is assumed to be normal with mean zero and gender-specific variance σ_j^2 where $\sigma_f^2 > \sigma_m^2$. In that case, it is possible to obtain an access function which is decreasing even if males and females have the same distribution of skills. We have for a given individual i applying for the job at rank $u = 1$ (see Aigner and Cain, 1977):

$$E[s_i | y_i(1), i \in \Omega_{j(i)}(1)] = E[s_i | y_i(1)] = (1 - \gamma_{j(i)})\omega + \gamma_{j(i)}y_i(1) \quad (20)$$

where $\gamma_j = \sigma^2 / (\sigma^2 + \sigma_j^2)$. The formula of the expected skills differs for the two genders only because the variance of the noise differs as shown by the term γ_j . Importantly, expected skills are higher for males than for females for any given level of observed skills $y_i(1) > \omega$. As the manager chooses an applicant depending on the expected skills, the selection process translates into some statistical discrimination in favor of males for observed skills above the average of true skills. Indeed, for a high level of observed skills, expected skills are higher for males than for females because there is a less uncertainty on their skills. This makes males with a high level of observed skills more likely to be hired and there is a filtering process with some transformations of the gender skill distribution as the rank decreases along the wage distribution. At high ranks, males are more likely to be hired, but as their average skills decrease faster than females as the rank decreases, the access to jobs at lower ranks gets better for

²²Alternatively, the case where there is no noise on the skills observed by the managers can be considered. We then have $\Psi_j(u) = E_{s_i}[\exp(s_i) | i \in \Omega_j(u)]$ which is the average exponentiated true skills for gender j . The access function compares the true skills of the available workers between the two genders at each rank in the wage distribution of job positions. If the exponentiated values of true skills are on average larger for males than for females, the access function at the highest rank is lower than one. The access function is decreasing when, as ranks get below one, the best males are on average more often selected for the jobs and leave the population of available workers. The same mechanisms and conclusions apply when there is a noise following the same law for the two genders.

females as they have on average better skills than males. However, there is also a compensating mechanism that goes in the opposite direction. Consider one male and one female having the same value of true skills. As females have on average a higher variance of the noise, the female may draw a much higher value of the observed skills $y_i(1)$ than the male. In that case, it is the female who is hired rather than the male. As the model is not analytically tractable even under our parametric assumptions, we show with some simulations that it is possible to generate a decreasing access function.²³ The implementation of the simulations is detailed in Appendix E. For skills following a standard normal law ($\omega = 0$ and $\sigma^2 = 1$) and noise of variance $\sigma_f^2 = 2$ for females and $\sigma_m^2 = 1$ for males, Figure 15 shows that the access function is decreasing for ranks above .2. This suggests that statistical discrimination could contribute to the decrease in the access function estimated from the data.²⁴

[*Insert Figures 15*]

The assignment of workers is an equilibrium under the extreme assumption that managers do not learn more about the true skills of workers once they are hired. Was it not the case, managers of good jobs would find that the worker they hired had skills which were not that high but drew a value of the noise which was high. In that case, he would want to fire him and hire another worker. Some workers at a lower rank would apply.

8.4 Full model

Of course, the three mechanisms presented here can occur simultaneously. It is easy to show that when including the three mechanisms in the model, the access function is given by:

$$h(u) = \frac{\mu_f(u)}{\mu_m(u)} \text{ with } \mu_j(u) = v_j(u) \theta_j(u) \Psi_j(u)$$

As the three mechanisms can all predict a decreasing access function, this formula confirms that they cannot be identified separately from the empirical access function obtained from the data. In fact, the access function rather captures the aggregate of the three mechanisms.

9 Conclusion

In this paper, we propose a measure of gender difference in access to jobs based on a job assignment framework. Workers compete for job positions which are ranked in decreasing order of their wage. Males and females have

²³In fact, the model is tractable only at rank $u = 1$ and, according to (20), we have for $y_i(1) > \omega$ that the expected true skills are higher for males than for females for some given observed skills. This translates into a higher propensity of managers to hire males rather than females for some given observed skills, males with high observed skills (and thus on average high true skills) being more often hired.

²⁴Note that the simulated access function takes values higher than the estimated access function and even higher than one for ranks below .5. However, this is not contradictory since statistical discrimination can combine with other mechanisms that lower the female access to jobs at all ranks such as taste discrimination.

different chances of getting each job. The manager of the best-paid job selects a worker among applicants. The manager of the second best-paid job hires an individual among the remaining workers, and so on. We define an access function as the probability ratio of females and males getting a job of a given rank.

We propose a non-parametric approach to estimate the access function from cross-section data. We apply our approach to full-time workers aged 40 – 45 using the 2003 *Déclarations Annuelles des Données Sociales* (DADS) which is *exhaustive* for all public and private firms. We find that females have a lower access to jobs at all ranks in the wage distribution of job positions and that the access function is decreasing with the rank. At the lowest ranks, the probability of females getting a given job is 9% lower than the probability of males. The difference between these probabilities is far larger at the highest ranks and climbs to 50%. These results are in line with a lower access to high-paid jobs for females.

We show that a decreasing access function can be theoretically micro-founded with three mechanisms. First, females may apply less often for high-paid jobs because working hours are less compatible with family constraints. Second, there can be taste discrimination against females which increases with the rank. Third, there can be statistical discrimination such that the skill distribution is the same for males and females, but skills are observed with more uncertainty for females than for males by managers. Indeed, highly skilled males are more often hired than highly skilled females for high-paid jobs because managers observe their skills with less noise. Moreover, remaining males who apply to less-paid jobs are on average less skilled than females and the chances of females being hired for less-paid jobs is higher.

Our model is initially designed to interpret cross-section gender wage differences in terms of gender differences in access to jobs. Alternatively, it could be applied to other subgroups of the population such as natives and immigrants. Also, it could be interesting to extend our model to a dynamic setting to study changes in the ranks of males and females in the wage distribution of job positions through job changes, promotions and lay-offs.

10 APPENDIX

10.1 Appendix A: Kernel estimator taking boundaries into account

In this Appendix, we explain our approach to estimate the density and cumulative of ranks of gender- j workers in the wage distribution of job positions. We consider the following generic minimization program:

$$\min_{\alpha, \beta} \int_0^1 K_\omega(u-v) [g(v) - \alpha - \beta(u-v)]^2 dv$$

where $K_\omega(u) = \frac{1}{\omega} K\left(\frac{u}{\omega}\right)$ with $K(u)$ a kernel with ω the bandwidth, $g(v)$ can be either the empirical cdf $F_n(v) = n^{-1} \sum_{i=1}^n 1_{\{u_i \leq v\}}$ to get an estimate of $\widehat{F}(v)$, or the empirical pdf $f_n(v) = n^{-1} \sum_{i=1}^n \delta_{u_i}(v)$, where δ_u is the Dirac function associated to u defined such that $\int_0^1 t(v) \delta_u(v) dv = t(u)$ for any function t of support $[0, 1]$, to get an estimate of $\widehat{f}(v)$.

It is possible to solve the minimization program from the two first-order conditions obtained when deriving the minimized expression with respect to α and β :

$$\begin{aligned} \int_0^1 K_\omega(u-v) g(v) dv &= \widehat{\alpha} \int_0^1 K_\omega(u-v) dv + \widehat{\beta} \int_0^1 K_\omega(u-v) (u-v) dv \\ \int_0^1 K_\omega(u-v) g(v) (u-v) dv &= \widehat{\alpha} \int_0^1 K_\omega(u-v) (u-v) dv + \widehat{\beta} \int_0^1 K_\omega(u-v) (u-v)^2 dv \end{aligned}$$

The solution of this system for α is either the estimated density when g is the empirical density or the estimated cumulative when g is the empirical cumulative. Denote:

$$a_\ell(u, \omega, g) = \int_0^1 K_\omega(u-v) t(v) (u-v)^\ell dv$$

where t is any function, we get:

$$\begin{aligned} \widehat{f}(u) &= \frac{a_2(u, \omega, 1) a_0(u, \omega, f_n) - a_1(u, \omega, 1) a_1(u, \omega, f_n)}{a_2(u, \omega, 1) a_0(u, \omega, 1) - a_1(u, \omega, 1)^2} \\ \widehat{F}(u) &= \frac{a_2(u, \omega, 1) a_0(u, \omega, F_n) - a_1(u, \omega, 1) a_1(u, \omega, F_n)}{a_2(u, \omega, 1) a_0(u, \omega, 1) - a_1(u, \omega, 1)^2} \end{aligned}$$

with:

$$\begin{aligned} a_\ell(u, \omega, f_n) &= \int_0^1 K_\omega(u-v) f_n(v) (u-v)^\ell dv \\ &= \int_0^1 \frac{1}{\omega} K\left(\frac{u-v}{\omega}\right) \frac{1}{n} \sum_{i=1}^n \delta_{u_i}(v) (u-v)^\ell dv \\ &= \frac{1}{n\omega} \sum_{i=1}^n \int_0^1 K\left(\frac{u-v}{\omega}\right) \delta_{u_i}(v) (u-v)^\ell dv \\ &= \frac{1}{n\omega} K\left(\frac{u-u_i}{\omega}\right) (u-u_i)^\ell \end{aligned}$$

and:

$$\begin{aligned}
a_\ell(u, \omega, 1) &= \omega^\ell \int_{\frac{u-1}{\omega}}^{\frac{u}{\omega}} K(v) v^\ell dv \\
a_\ell(u, \omega, F_n) &= \omega^\ell \int_{\frac{u-1}{\omega}}^{\frac{u}{\omega}} K(v) v^\ell F_n(u - \omega v) dv \\
&= \frac{\omega^\ell}{n} \sum_{i=1}^n \int_{\frac{u-1}{\omega}}^{\frac{u}{\omega}} K(v) v^\ell 1_{u_i \leq u - \omega v} dv
\end{aligned}$$

Define:

$$I_\ell(v, u) = \int_u^v K(x) x^\ell dx$$

We have:

$$\begin{aligned}
a_\ell(u, \omega, 1) &= \omega^\ell I_\ell\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) \\
a_\ell(u, \omega, F_n) &= \frac{\omega^\ell}{n} \sum_{i=1}^n I_\ell\left(\frac{u-u_i}{\omega}, \frac{u-1}{\omega}\right)
\end{aligned}$$

From this, we get the expressions:

$$\begin{aligned}
\hat{f}(x) &= \frac{1}{n\omega} \sum_{i=1}^n K\left(\frac{u-u_i}{\omega}\right) \left[\frac{I_2\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) - \left(\frac{u-u_i}{\omega}\right) I_1\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right)}{I_2\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) I_0\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) - I_1\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right)^2} \right] \\
\hat{F}(u) &= \frac{1}{n} \sum_{i=1}^n \left[\frac{I_2\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) I_0\left(\frac{u-u_i}{\omega}, \frac{u-1}{\omega}\right) - I_1\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) I_1\left(\frac{u-u_i}{\omega}, \frac{u-1}{\omega}\right)}{I_2\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) I_0\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right) - I_1\left(\frac{u}{\omega}, \frac{u-1}{\omega}\right)^2} \right]
\end{aligned}$$

We can compute these expressions for an Epanechnikov Kernel $K(u) = \frac{3}{4}(1-u^2)1_{|u| \leq 1}$ using the fact that we have:

$$\begin{aligned}
I_\ell(u, v) &= \int_u^v \frac{3}{4}(1-x^2) x^\ell 1_{\{-1 \leq x \leq 1\}} dx \\
&= \int_{u \vee -1}^{v \wedge 1} \frac{3}{4}(1-x^2) x^\ell 1_{\{-1 \leq x \leq 1\}} 1_{\{-1 \leq v; u \leq 1; u \leq v\}} dx \\
&= 1_{\{-1 \leq v; u \leq 1; u \leq v\}} \frac{3}{4} \left[\frac{(v \wedge 1)^{\ell+1} - (u \vee -1)^{\ell+1}}{\ell+1} - \frac{(v \wedge 1)^{\ell+3} - (u \vee -1)^{\ell+3}}{\ell+3} \right]
\end{aligned}$$

where \wedge (resp. \vee) is the operator taking the minimum (resp. maximum) of two arguments.

10.2 Appendix B: estimating a linear access function

In this appendix, we explain how to approximate the access function by a linear function of ranks. We estimate a specification of the form: $h(u) = a - b.u$ and test whether this specification fits the data. This is done in the case of a national job market and some separate firm job submarkets.

10.2.1 National market

The random variable corresponding to the rank of a gender- j worker in the wage distribution of job positions is denoted U_j . Its cumulative function is given by: $F_{U_j}(u) = \frac{n_j(u)}{n_j}$. Its quantile function is $u_j(\cdot)$. We have by definition:

$$v = F_{U_j}[u_j(v)]$$

From this equation, we get:

$$F_{U_f}[u_f(v)] = F_{U_m}[u_m(v)] \quad (21)$$

We use $\frac{N_j}{N}$ where $N = N_f + N_m$ as an estimator of n_j . For a given linear specification of $h(u)$, we can solve the model for $n_f(u)$ and deduce $n_m(u)$ from the equality $n_m(u) = u - n_f(u)$. We can then deduce the quantile function of U_j as it can be written as: $u_j(v) = n_j^{-1}(n_j.v)$. The parameters a and b are estimated minimizing the distance between the left and right-hand sides of (21) after replacing F_{U_j} , $j \in \{m, f\}$ by their empirical counterparts. Denoting $\theta = (a, b)$, the minimization program is:

$$\min_{\theta} C(\theta) \text{ with } C(\theta) = \int_0^1 \left[\widehat{F}_{U_f}[u_f(v)] - \widehat{F}_{U_m}[u_m(v)] \right]^2 du \quad (22)$$

Details on how to evaluate the minimization criterion are given in Combes et al. (2012).

The minimization criterion can be used to conduct a specification test. We have $\frac{N_j}{N} \xrightarrow{P} p_j$ (the proportion of gender- j workers in the population) where $N = N_f + N_m$ and $p_f + p_m = 1$. Using Donsker's theorem and Slutsky's lemma (see Van der Vaart, 1998, example 20.11), we get:

$$N^{1/2} \begin{pmatrix} \widehat{F}_{U_f}[u_f(v)] - v \\ \widehat{F}_{U_m}[u_m(v)] - v \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{1}{\sqrt{p_f}} B_f(v) \\ \frac{1}{\sqrt{p_m}} B_m(v) \end{pmatrix} \quad (23)$$

where $B_f(\cdot)$ and $B_m(\cdot)$ are some independent Brownian bridges.

Applying the continuous function $\Psi(x_1, x_2) = (x_1 - x_2)^2$ to (23), we get:

$$N \left[\widehat{F}_{U_f}[u_f(v)] - \widehat{F}_{U_m}[u_m(v)] \right]^2 \Rightarrow \left(\frac{1}{p_f} + \frac{1}{p_m} \right) B(v)^2 \quad (24)$$

where $B(v) = \left(\frac{1}{p_f} + \frac{1}{p_m} \right)^{-1/2} \left[\frac{1}{\sqrt{p_f}} B_f(v) - \frac{1}{\sqrt{p_m}} B_m(v) \right]$. It is easy to show that $B(\cdot)$ is a Brownian bridge. Indeed, $B(\cdot)$ is Gaussian by construction and we have:

$$\begin{aligned} \text{cov}[B(u), B(v)] &= \left(\frac{1}{p_f} + \frac{1}{p_m} \right)^{-1} \left[\frac{1}{p_f} \text{cov}[B_f(u), B_f(v)] + \frac{1}{p_m} \text{cov}[B_m(u), B_m(v)] \right] \\ &= \left(\frac{1}{p_f} + \frac{1}{p_m} \right)^{-1} \left[\frac{1}{p_f} (u \wedge v - uv) + \frac{1}{p_m} (u \wedge v - uv) \right] \\ &= u \wedge v - uv \end{aligned}$$

Integrating (24) over the $[0, 1]$ interval, we obtain:

$$N \left(\frac{1}{p_f} + \frac{1}{p_m} \right)^{-1} C(\theta) \implies \int_0^1 B(v)^2 dv$$

and the right-hand side follows a Cramer Van-Mises statistic whose threshold at the 5% level is .46136 (see Knott, 1974). We can approximate the left-hand side replacing p_j by $\frac{N_j}{N}$ and θ by $\hat{\theta}$, and then conduct a specification test where the hypothesis we test is the equality (21).

10.2.2 Separate market for each firm

For any firm z , we consider that the access function takes the linear form $h^z(u) = a - bu$. We then have:

$$d^z(v) = 0 \text{ with } d^z(v) = F_{U_f^z}^z [u_f^z(v)] - F_{U_m^z}^z [u_m^z(v)] \quad (25)$$

where $F_{U_j^z}^z(\cdot)$ is the cumulative function of the random variable U_j^z corresponding to the rank of a gender- j worker in the wage distribution of job positions in firm z , and $u_j^z(\cdot)$ is the corresponding quantile function. Denote:

$$\hat{d}^z(u) = \widehat{F}_{U_f^z}^z [u_f^z(v)] - \widehat{F}_{U_m^z}^z [u_m^z(v)]$$

where $\widehat{F}_{U_j^z}^z(\cdot)$ is the empirical counterpart of $F_{U_j^z}^z(\cdot)$. We can recover an estimator of the parameters θ using the minimization program:

$$\min_{\theta} C^Z(\theta) \text{ with } C^Z(\theta) = \sum_z P^z \int_0^1 \hat{d}^z(u)^2 du$$

where P^z is a weight (in practice, it is chosen to be the proportion of workers in firm z). The minimization criterion can be computed in a way similar to the one given by (22).

Once again, the minimization criterion can be used to conduct a specification test. We have $\frac{N_j^z}{N} \xrightarrow{P} p_j^z$ (the proportion of gender- j workers in firm z) where $N = \sum_{z=1}^Z N_f^z + \sum_{z=1}^Z N_m^z$ and $\sum_{z=1}^Z p_f^z + \sum_{z=1}^Z p_m^z = 1$. Using again Donsker's theorem and Slutsky's Lemma, we get:

$$N^{1/2} \begin{pmatrix} \widehat{F}_{U_f^1}^1 [u_f^1(v)] - v \\ \widehat{F}_{U_m^1}^1 [u_m^1(v)] - v \\ \dots \\ \widehat{F}_{U_f^Z}^Z [u_f^Z(v)] - v \\ \widehat{F}_{U_m^Z}^Z [u_m^Z(v)] - v \end{pmatrix} \implies \begin{pmatrix} \frac{1}{\sqrt{p_f^1}} B_f^1(v) \\ \frac{1}{\sqrt{p_m^1}} B_m^1(v) \\ \dots \\ \frac{1}{\sqrt{p_f^Z}} B_f^Z(v) \\ \frac{1}{\sqrt{p_m^Z}} B_m^Z(v) \end{pmatrix} \quad (26)$$

where $B_j^z(\cdot)$, with $j \in \{f, m\}$ and $z \in \{1, \dots, Z\}$ are some independent Brownian Bridges.

Applying the continuous function $\Psi(x_1, \dots, x_{2Z}) = \sum_{z=1}^Z P^z \cdot (x_{2z} - x_{2z-1})^2$ to (26), we get:

$$N \sum_{z=1}^Z P^z \cdot \widehat{d}^z(v)^2 \implies \sum_{z=1}^Z P^z \cdot \left(\frac{1}{p_f^z} + \frac{1}{p_m^z} \right) B^z(v)^2$$

where $B^z(\cdot)$, $z \in \{1, \dots, Z\}$ are some independent Brownian bridges. Integrating this equation over the $[0, 1]$ interval, we obtain:

$$N \left[\sum_{z=1}^Z P^z \cdot \left(\frac{1}{p_f^z} + \frac{1}{p_m^z} \right) \right]^{-1} C^Z(\theta) \implies \int_0^1 B(v)^2 dv$$

where the right-hand side follows a Cramer Van-Mises statistic. We can approximate the left-hand side replacing p_j^z by $\frac{N_j^z}{N}$ and θ by $\widehat{\theta}$, and then conduct a specification test where the hypothesis we test is the set of equalities (25).

10.3 Appendix C: solution of the model

10.3.1 Theorem

We have the following existence theorem for the full model (the proof is given in the next subsection):

Theorem 2 *Suppose that $\mu_m(\cdot)$ and $\mu_f(\cdot)$ are C^1 on $(0, 1]$ and there is a constant $c > 0$ such that $\phi_m(u) > c$ and $\mu_f(u) > c$ for all $u \in (0, 1]$, then there is a unique two-uplet $\{n_f(\cdot), n_m(\cdot)\}$ verifying (2) where $\phi_j(\cdot)$ is given by:*

$$\phi_j(u) = \frac{\mu_j(u)}{n_f(u)\mu_f(u) + n_m(u)\mu_m(u)}$$

We assume in our theorem that the gender-specific exogenous components must take their value above a strictly positive threshold, such that males and females can access all jobs. This assumption is made in order to ensure that the unit probabilities are always well-defined as the denominator in their formula then cannot be zero. We can extend the model to some cases where workers of a given gender have no access to some jobs and show that the model still has a solution. Consider, for instance, the case where females cannot access the best-paid jobs of ranks above a given threshold \widetilde{u} because of a glass ceiling effect but have access to all jobs of ranks below this threshold. In that case, all the jobs of ranks above the threshold are occupied by males. For jobs of rank below the threshold, there is then a measure n_f of available females competing with a measure $n_m - (1 - \widetilde{u})$ of available males (provided that not all males have been hired for the best-paid jobs). It is possible to apply our existence theorem on the subset of ranks below the threshold and get a global solution on the whole set of ranks using a continuity argument.

Also note that the theorem can be extended to the case where the gender-value functions are not continuous, but rather discontinuous at a finite number of ranks. First consider the case where there is only one point of discontinuity. It is possible to apply the existence theorem separately for the subset of ranks below that point, and

the subset of ranks above that point. The solution on the whole set of ranks can be recovered from the solutions on the two subsets of ranks using again a continuity argument. This procedure can easily be extended to the case where there are more points of discontinuity.

10.3.2 Proof

The proof revolves around the application of the Cauchy-Lipschitz theorem. For the full model, we have:

$$\phi_f(u) = \frac{\mu_f(u)}{n_f(u)\mu_f(u) + n_m(u)\mu_m(u)} \quad (27)$$

Plugging (27) into (2) for females, we get:

$$n'_f(u) = \frac{n_f(u)\mu_f(u)}{n_m(u)\mu_m(u) + n_f(u)\mu_f(u)} \quad (28)$$

Using the equality $n_f(u) + n_m(u) = u$, we obtain:

$$n'_f(u) = g(u, n_f(u)) \quad (29)$$

where

$$g(u, n) = \frac{n\mu_f(u)}{n\mu_f(u) + (u - n)\mu_m(u)}$$

We can state that the denominator of $g(\cdot, \cdot)$ is strictly positive for sure only on $\tilde{\Phi} = \{(u, n) \in (0, 1] \times [0, n_f] \mid n \leq u\}$. As $\mu_f(\cdot)$ and $\mu_m(\cdot)$ are C^1 on $\tilde{\Phi}$, it is straightforward to show that $g(\cdot, \cdot)$ is C^1 on $\tilde{\Phi}$. On any compact set $\{(u, n) \in [\varepsilon, 1] \times [0, n_f] \mid n \leq u\}$ with $\varepsilon \in (0, 1)$, $g(\cdot, \cdot)$ is Lipschitzienne and (29) has a solution for $n_f(\cdot)$ on $[\varepsilon, 1]$. As this is true for ε arbitrarily close to zero, (29) has a unique solution for $n_f(\cdot)$ on $(0, 1]$. There is then a unique solution for $n_m(\cdot)$ on $(0, 1]$ which is given by $n_m(u) = u - n_f(u)$.

10.4 Appendix D: characteristics of male and female subpopulations

In this Appendix, we show that the distributions of highest qualifications and age at the end of initial schooling are close for our subpopulations of males and females using the 2003 French Labour Survey.

We only select executives aged 40-45 working full-time and obtain a subsample of 1,685 individuals including 24.0% of females (the corresponding figure is 22.4% for our DADS sample). As shown in Table A.3, when comparing the distributions of highest qualifications for the two genders, we find that the proportions of graduates and undergraduates are a bit higher for females than for males. Overall, the distributions of highest qualifications of the two genders are not statistically different at the 5% threshold according to a Chi-square test (but they are statistically significant at the 10% threshold).

As shown in Table A.4, we also find that the average age of initial schooling is slightly higher for males (21.5 years) than for females (21.2 years). The difference can be explained by a higher average age of initial schooling

for male graduates than for female graduates (although a slightly lower proportion of males becomes graduates). It is not significant even at the 10% threshold when conducting a T-student test.

Even though it would be better to control for education in our analysis, it is likely that omitting variables related to education does not noticeably bias the results as gender differences in education are rather small.

10.5 Appendix E: Simulation procedure for statistical discrimination

In this appendix, we explain how we conduct the simulations to generate the access function when there is some statistical discrimination. We generate $N = 17750$ observations among which 3976 are considered to be females. These number are fixed such that the sample size is 5% of workers in the private sector and the proportion of females is that observed in the data (22.4%).²⁵ For each observation, we draw the skills in a normal law with mean $\omega = 0$ and variance $\sigma^2 = 1$. There are also N jobs ranked in decreasing order of attractiveness for workers.

We want to compute the access function (19) where $\Psi_j(u)$ is given in (18). We compute the access function backwards, evaluating it first for the most attractive job at rank $u = 1$. For an observation which corresponds to a male (resp. female), we draw a noise in a normal law with mean zero and variance $\sigma_m^2 = 1$ (resp. $\sigma_f^2 = 2$). Summing the simulated skills and noise, we generate the observed skills $y_i(1)$.

We then compute $E[s_i | y_i(1), i \in \Omega_j(1)]$ for all $i \in \Omega_j(1)$, where $\Omega_j(1)$ is the whole sample of gender- j workers. This is done conducting a non-parametric regression of s_i on $y_i(1)$ such that:

$$\widehat{E}[s_i | y_i(1), i \in \Omega_j(1)] = \left[\sum_{k|j(k)=j} K\left(\frac{y_k(1) - y_i(1)}{h_j}\right) s_k \right] / \left[\sum_{k|j(k)=j} K\left(\frac{y_k(1) - y_i(1)}{h_j}\right) \right]$$

where $K(\cdot)$ is a kernel chosen to be Epanechnikov and h_j is the gender- j bandwidth fixed to the value given by the rule of thumb. An estimator of $\Psi_j(1)$ denoted $\widehat{\Psi}_j(1)$ is then given by the sample average of $\exp \widehat{E}[s_i | y_i(1), i \in \Omega_j(1)]$ for observations in $\Omega_j(1)$. The value of the access function at rank 1 can then be estimated using the formula: $\widehat{h}(1) = \widehat{\Psi}_f(1) / \widehat{\Psi}_m(1)$.

The manager of the most attractive job derives from a worker i a utility $V_i(1)$ given by (17). He selects the worker i_1 yielding the highest utility. The utility derived from a worker i can be approximated by the sum of $\widehat{E}[s_i | y_i(1), i \in \Omega_{j(i)}(1)]$ and a draw of the match value in an extreme value law. It is then possible to construct the set of gender- j workers available for the job of rank $(N - 1)/N$. If the worker hired by the manager i_1 is a male, we have: $\Omega_f((N - 1)/N) = \Omega_f(1)$ and $\Omega_m((N - 1)/N) = \Omega_m(1) \setminus \{i_1\}$. If the worker hired by the manager is a female, we have: $\Omega_m((N - 1)/N) = \Omega_m(1)$ and $\Omega_f((N - 1)/N) = \Omega_f(1) \setminus \{i_1\}$.

The whole procedure is repeated for the job of rank $(N - 1)/N$ to compute $\widehat{h}((N - 1)/N)$ and determine $\Omega_j((N - 2)/N)$, $j \in \{1, 2\}$. We keep repeating the procedure downwards for the jobs of rank $(N - 2)/N$, $(N - 3)/N$, ..., $1/N$.

²⁵We consider this sample size to keep the simulations tractable with current computer resources.

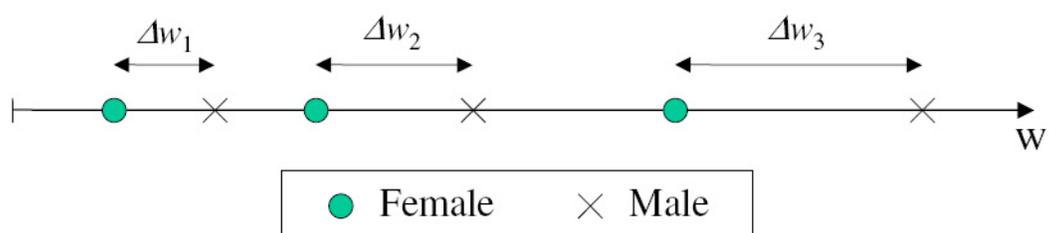
References

- [1] Abowd J., Kramarz F. and S. Roux (2006), “Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data”, *Economic Journal*, 116, pp. 245-285.
- [2] Aigner D. and G. Cain (1977), “Statistical Theories of Discrimination in Labor Markets”, *Industrial and Labor Relations Review*, 30(2), pp. 175-187.
- [3] Albrecht J., Björklund A. and S. Vroman (2003), “Is There a Glass Ceiling in Sweden?”, *Journal of Labour Economics*, 21(1), pp. 145-177.
- [4] Arrow J.K. (1971), *The theory of discrimination*, Working paper n°30A, Princeton University, in O. Aschenfelter & A. Rees Eds, 1974, *Discrimination in Labor Markets*, NJ: Princeton UP.
- [5] Babcock L. and S. Laschever (2003), *Women Don't Ask: Negotiation and the Gender Divide*, Princeton University Press, 240p.
- [6] Baker G., Gibbs M. and B. Holmstrom (1994), “The internal economics of the firm: evidence from personnel data, ”, *The Quarterly Journal of Economics*, 109(4), 881-919.
- [7] Bartel A. (2004), “Human Resource Management and Organizational Performance: Evidence from Retail Banking”, *Industrial and Labor Relations Review*, 57(2), 181-203.
- [8] Becker G. (1971), *The Economics of Discrimination*, Second Edition, The University of Chicago Press, 178p.
- [9] Bertrand M. and K.F. Hallock (2001), “The Gender Gap in Top Corporate Jobs”, *Industrial and Labor Relations Reviews*, 55(1), 3-21.
- [10] Blackaby D.H., Booth A. and J. Frank (2005), “Outside Offers and the Gender Pay Gap: Empirical Evidence from the UK Academic Labour Market”, *The Economic Journal*, 115, F81-F107.
- [11] Cornell B. and I. Welch (1996), “Culture, Information, and Screening Discrimination”, *Journal of Political Economy*, 104(3), pp. 542-571.
- [12] Coate S. and G. Loury (1993), “Will affirmative action policy eliminate negative stereotypes?”, *American Economic Review*, 83(5), pp. 1220-40.
- [13] Combes P., Duranton G., Gobillon L., Puga D. and S. Roux (2012), “The productivity advantages of large cities: distinguishing agglomeration from firm selection”, *Econometrica*, forthcoming.
- [14] Costrell R. and G. Loury (2004), “Distribution of Ability and Earnings in a Hierarchical Job Assignment Model”, *Journal of Political Economy*, 112(6), pp. 1322-1363.

- [15] Dagsvik J. (1994), “Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes”, *Econometrica*, 62(5), pp. 1179-1205.
- [16] Dejonghe V. and C. Gasnier (1990), “Pratiques salariales et gestion du personnel dans les banques et les assurances”, *Travail et Emploi*, 45(3), pp. 40-46.
- [17] Donald S., Green D. and H. Paarsch (2000), “Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator Distribution Functions in the Presence of Covariates”, *The Review of Economic Studies*, pp. 609-633.
- [18] Fortin N. and T. Lemieux (2002), “Are Women’s Wage Gains Men’s Losses? A Distributional Test”, *American Economic Review*, 90(2), pp. 456-460.
- [19] Gobillon L., Meurs D. and S. Roux (2011), “Estimating gender differences in access to jobs: females trapped at the bottom of the ladder”, CEPR Working Paper 7475.
- [20] Killingsworth M. and C. Reimers (1983), “Race, Ranking, Promotions, and Pay at a Federal Facility: A Logit Analysis”, *Industrial and Labor Relations Review*, 36, pp. 92-107.
- [21] Knott M. (1974), “The distribution of the Cramér-von Mises statistic for small sample sizes”, *Journal of the Royal Statistical Society B*, 36(3), pp. 430–438.
- [22] Lancaster T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press: Cambridge.
- [23] Pekkarinen T. and J. Vartiainen, (2006), “Gender differences in promotion on a job ladder: Evidence from Finnish metalworkers”, *Industrial and Labor Relations Review*, 59(2), pp. 285-301.
- [24] Phelps E. (1972), “The statistical theory of racism and sexism”, *American Economic Review*, 62, pp. 639-651.
- [25] Polachek S. (1981), “Occupatioanl Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure”, *The Review of Economics and Statistics*, 63(1), pp. 60-69.
- [26] Press W., Teukolsky S., Vettering W. and B. Flannery (1992), *Numerical Recipes in C*, Second Edition, Cambridge University Press, 994p.
- [27] Ridder G. and I. Tunalı (1999), “Stratified partial likelihood estimation”, *Journal of Econometrics*, 92(2), 193-232.
- [28] Sattinger M. (1993), “Models of the Distribution of Earnings”, *Journal of Economic Literature*, 31(2), pp. 831-880.
- [29] Silverman B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, 175p.

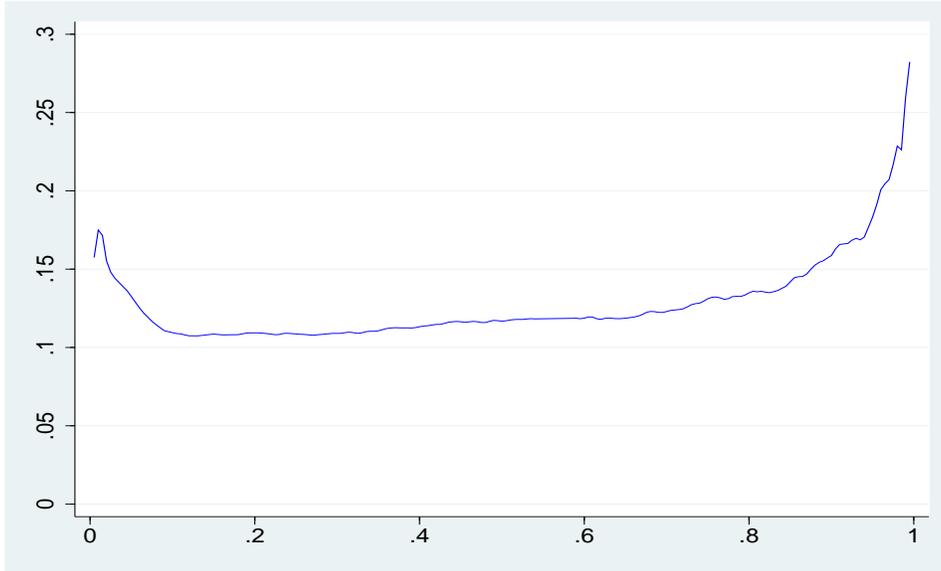
- [30] Teulings C. (1995), “The Wage Distribution in a Model of the Assignment of Skills to Jobs”, *Journal of Political Economy*, 103(2), pp. 280-315.
- [31] Van der Vaart A.W. (1998), *Asymptotic Statistics*, 443p, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Figure 1: Distribution of males and females along a job ladder



Note: Δw_j is the gender wage difference for the j^{th} male-female pair.

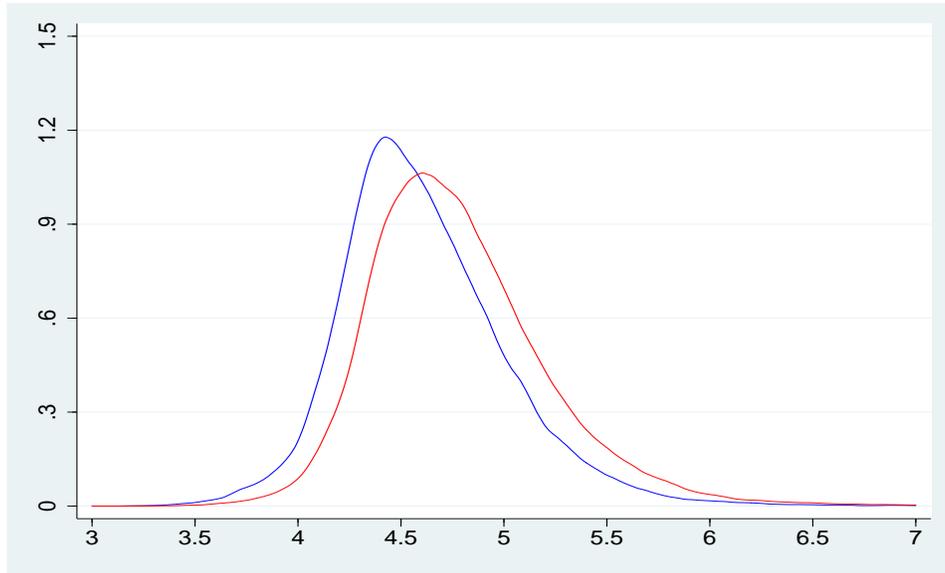
Figure 2: Gender quantile difference (M-F) obtained from the model, $h(u) = .672$



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

Note: the curve represents the gender quantile difference when the wage distribution of job positions is supposed to be the empirical wage distribution in the banking industry (and the proportion of females is fixed to the one in that industry: 28.7%). The gender quantile difference is the difference between the quantile of males and the quantile of females.

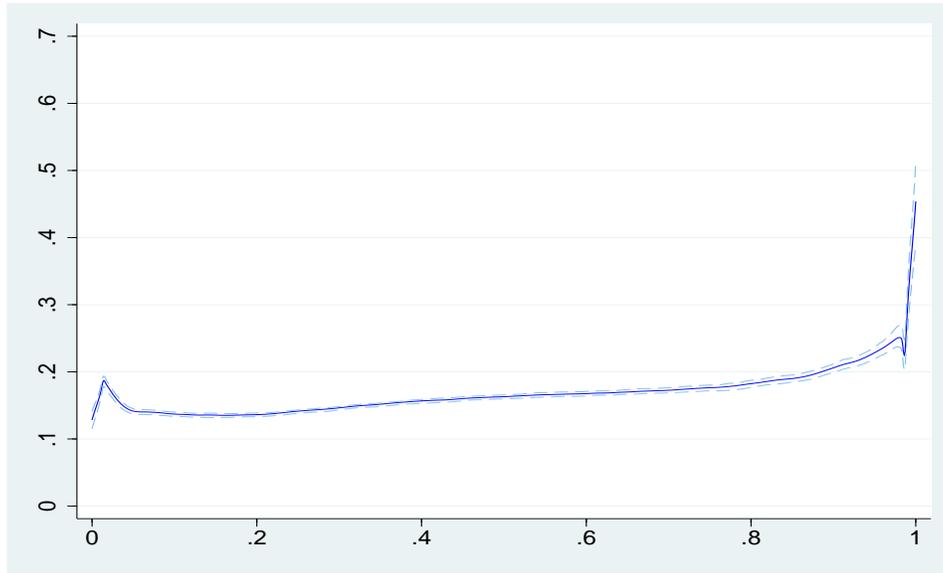
Figure 3: Log-wage densities of the two genders, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: in red, log-wage distribution of males; in blue, log-wage distribution of females.

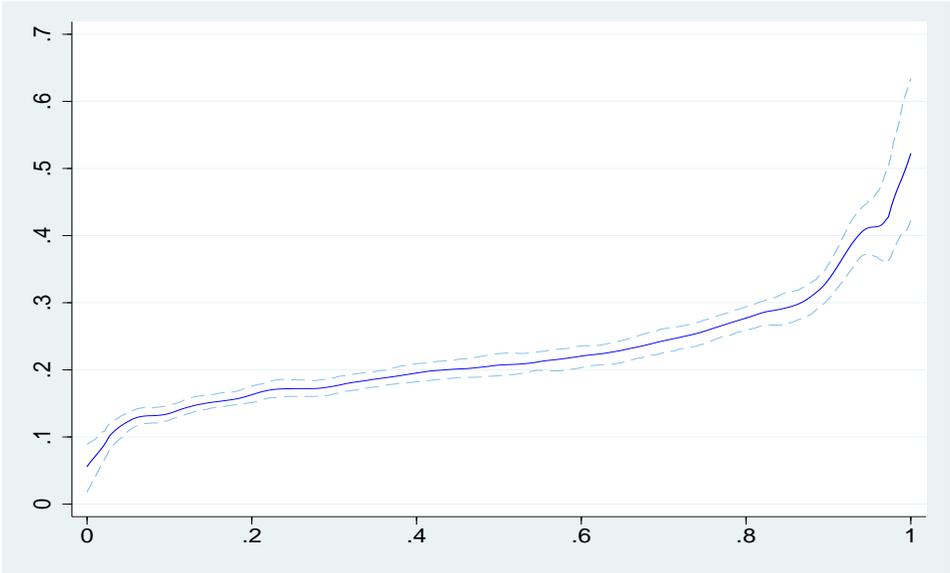
Figure 4: Gender quantile difference (M-F) as a function of rank, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: the gender quantile difference is the difference between the quantile of males and the quantile of females. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

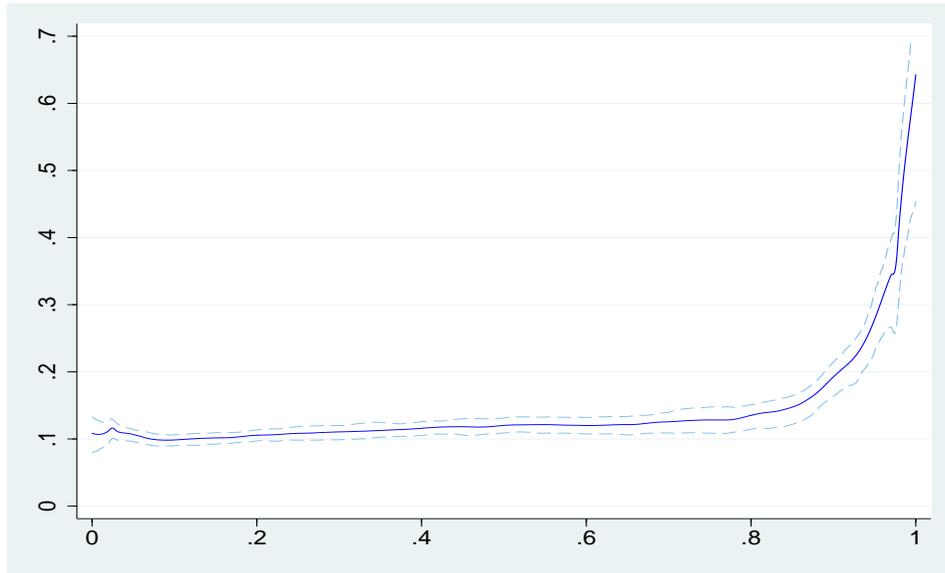
Figure 5: Gender quantile difference (M-F) as a function of rank, insurance industry



Source: DADS, 2003, full-time executives of the insurance industry aged 40-45.

Note: the gender quantile difference is the difference between the quantile of males and the quantile of females. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

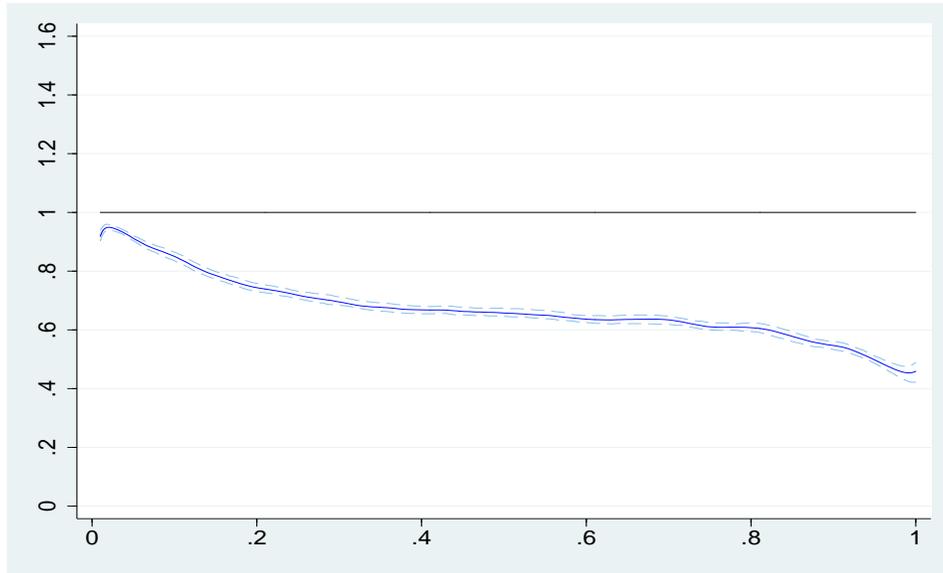
Figure 6: Gender quantile difference (M-F) as a function of rank, banking industry



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

Note: the gender quantile difference is the difference between the quantile of males and the quantile of females. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

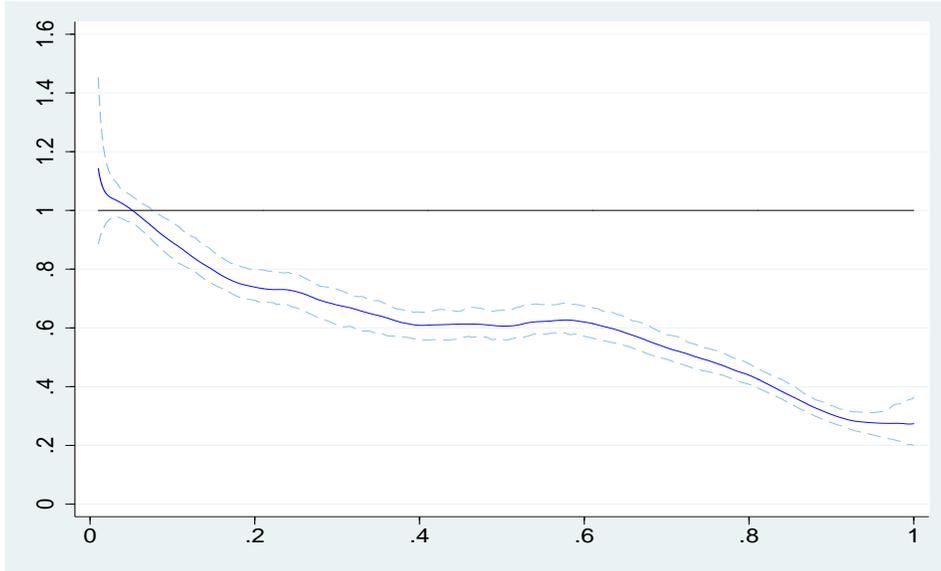
Figure 7: Access function (F/M) as a function of job rank, pooled industries



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: see Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

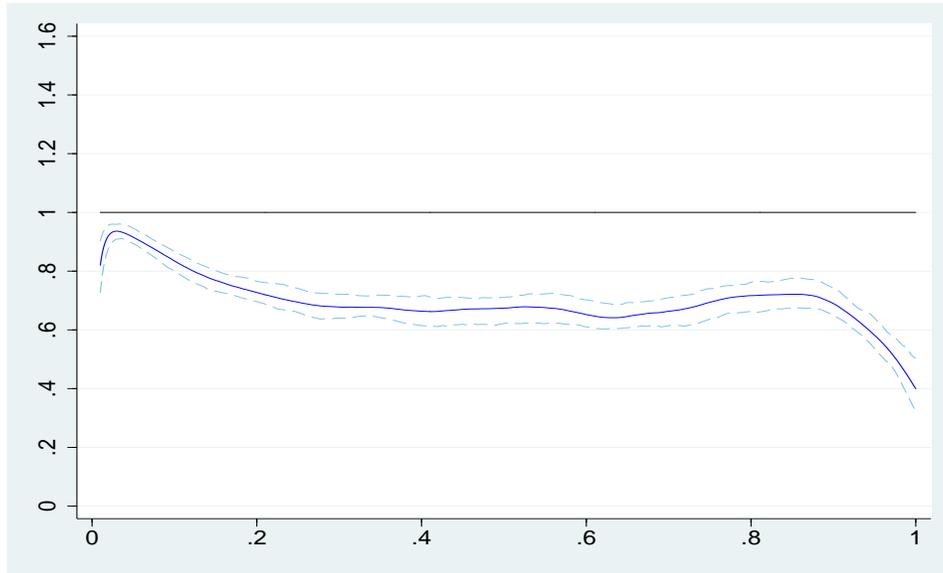
Figure 8: Access function (F/M) as a function of job rank, insurance industry



Source: DADS, 2003, full-time executives of the insurance industry aged 40-45.

Note: see Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

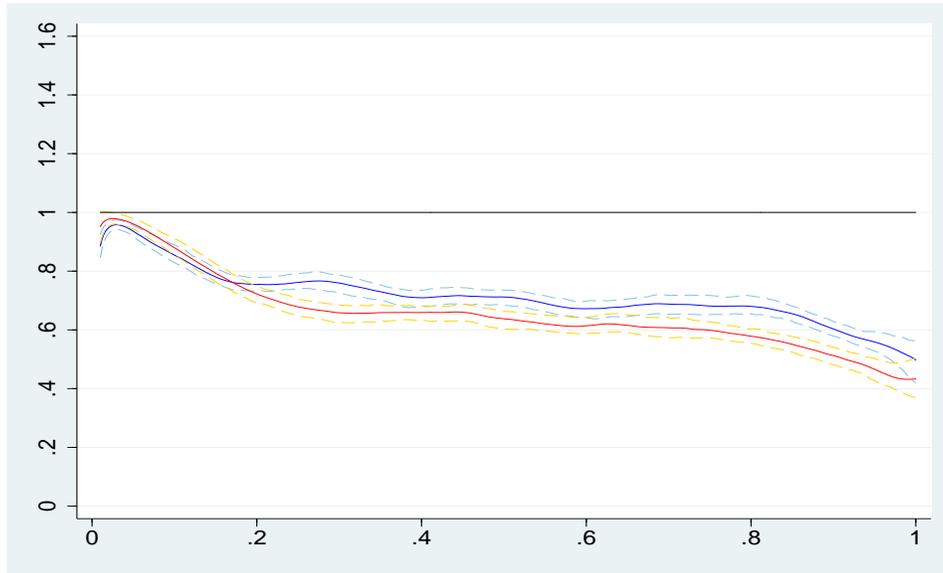
Figure 9: Access function (F/M) as a function of job rank, banking industry



Source: DADS, 2003, full-time executives of the banking industry aged 40-45.

Note: see Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

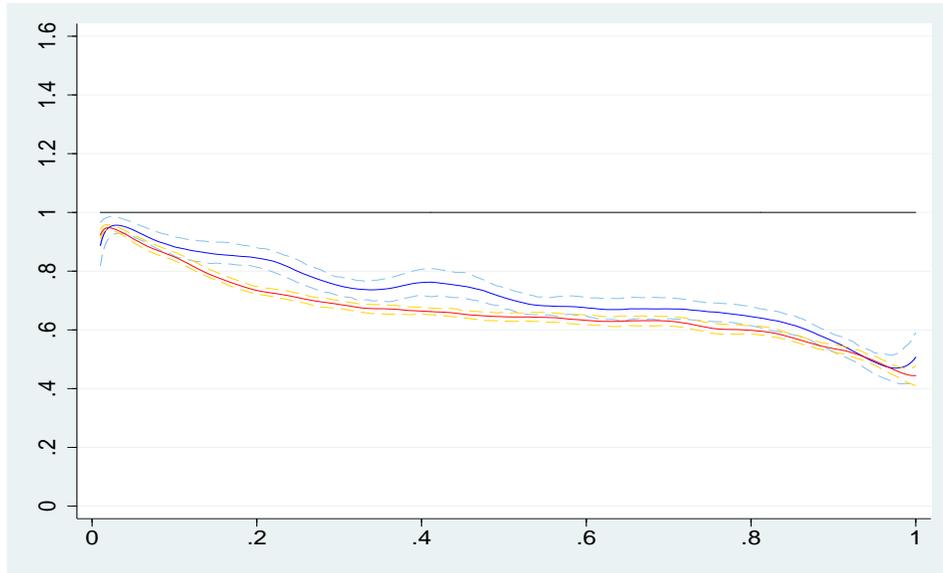
Figure 10: Access function (F/M) as a function of job rank, pooled industries, by age group



Source: DADS, 2003, full-time executives of the Private Sector aged 40 or 45.

Note: in blue: workers aged 40; in red: workers aged 45. See Section 3 for details on the estimation method. Bounds of the confidence intervals estimated by bootstrap (100 replications) are represented by dashed lines.

Figure 11: Access function (F/M) as a function of job rank, pooled industries, by country of birth

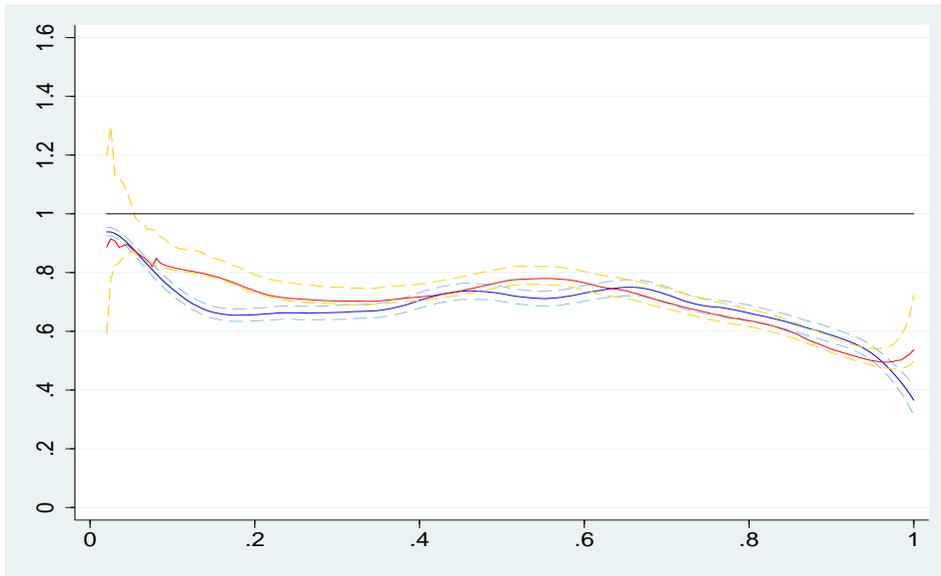


Source: DADS, 2003, full-time executives of the Private Sector.

Note: in blue: born in France; in red: born in a foreign country. See Section 3 for details on the estimation method.

Bounds of the confidence intervals estimated by bootstrap (100 replications) are represented by dashed lines.

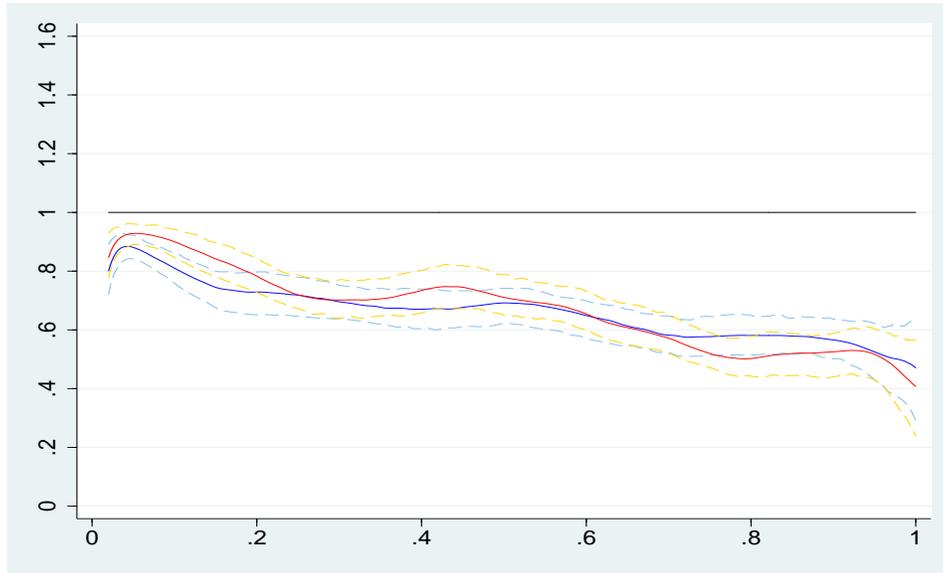
Figure 12: Average access function (F/M) as a function of job rank, large firms, national market and segmented submarkets



Source: DADS, 2003, full-time executives aged 40-45 in firms employing more than 150 such executives.

Note: in blue: access function for the national market; in red: average access function computed across segmented submarkets (each submarket being a large firm). See Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

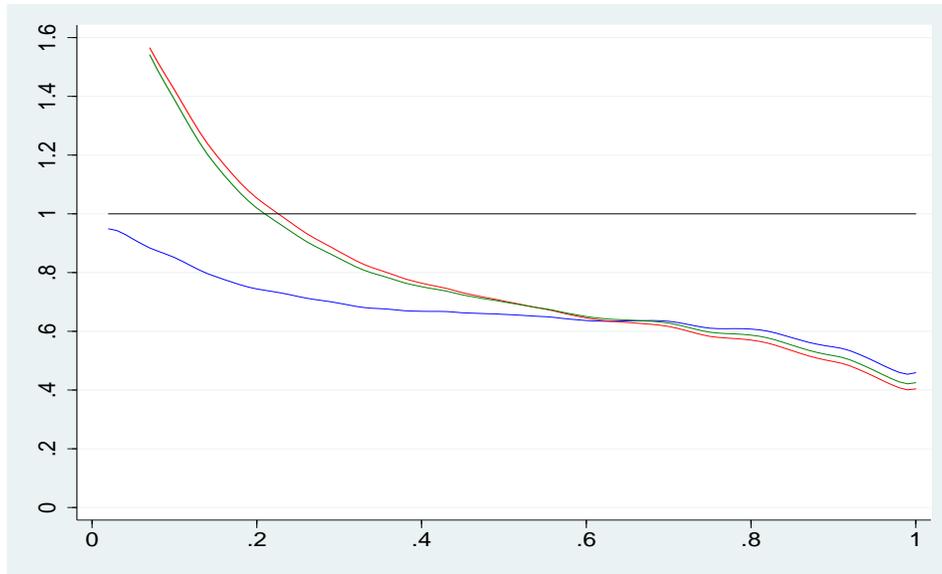
Figure 13: Average access function (F/M) as a function of job rank, by experience group



Source: DADS, 2003, full-time executives of the Private Sector aged 40-45, born in October of an even year.

Note: in blue: labor market experience above median; in red: labor market experience below median. See Section 3 for details on the estimation method. Bounds of the confidence interval estimated by bootstrap (100 replications) are represented by dashed lines.

Figure 14: Access function obtained when workers occupying an intermediate position are included in the sample



Source: DADS, 2003, full-time workers occupying an executive or an intermediate position in the Private Sector, aged 40-45, and born in October of an even year.

Note: in blue: access function for executive workers only. In red: access function obtained when the sample includes executives and workers occupying an intermediate position. In green: access function obtained when the sample includes executives and workers occupying an intermediate position with at least ten years of experience in intermediate positions.

Figure 15: Access function (F/M) from simulation of statistical discrimination

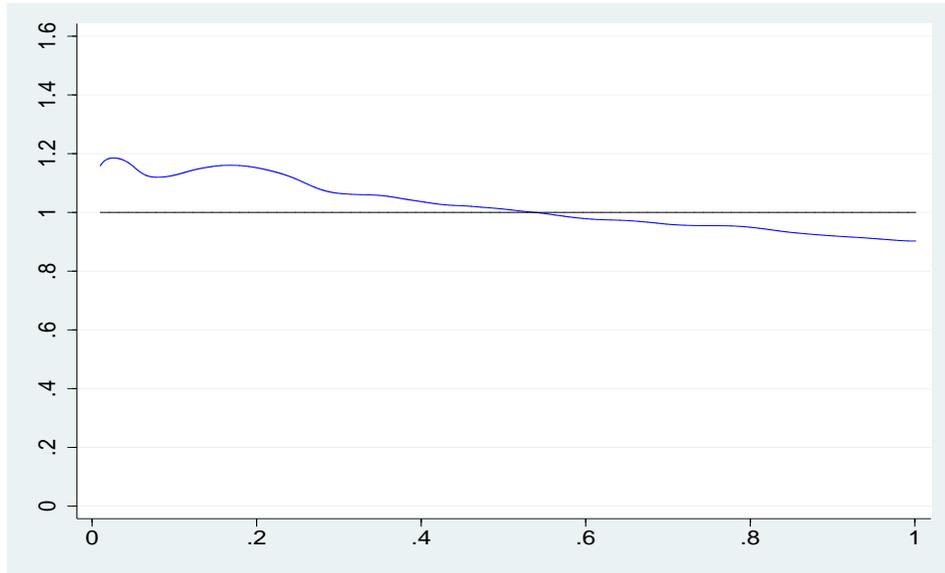


Table 1: Descriptive statistics by subgroup of firms

Sector	Nb. firms	Nb. jobs	% females	Daily wages, all					
				Mean	Std	p50	p5	p90	p95
All firms	86989	354968	22.4	139	602	109	64	200	252
Large firms	175	90345	22.2	132	119	113	69	196	245
Banking	545	18628	28.7	142	449	104	68	197	269
Banking, large firms	18	9024	31.0	155	298	112	68	223	320
Insurance	507	9360	36.9	125	74	107	67	186	238
Insurance, large firms	12	4611	39.0	119	64	106	68	175	216

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: pX corresponds to the X^{th} percentile.

Table 1: Descriptive statistics by subgroup of firms (cont.)

Sector	Daily wages, Females						Daily wages, Males					
	Mean	Std	p50	p5	p90	p95	Mean	Std	p50	p5	p90	p95
All firms	119	434	96	58	169	209	145	642	113	66	208	263
Large firms	116	83	103	62	170	208	136	126	116	72	202	257
Banking	120	211	95	64	171	220	150	514	108	71	208	292
Banking, large firms	126	134	103	63	190	245	168	347	117	71	240	359
Insurance	105	49	94	63	147	172	136	84	115	71	206	262
Insurance, large firms	102	41	93	64	141	158	131	72	114	72	193	245

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: pX corresponds to the X^{th} percentile.

Table 2: Values of the access function at different ranks

	p5	p10	p25	p50	p75	p90	p95	p99
All firms	.91 [.90,.92]	.85 [.84,.86]	.72 [.70,.73]	.66 [.65,.67]	.61 [.60,.62]	.55 [.53,.56]	.50 [.48,.51]	.45 [.42,.48]
Large firms	.89 [.87,.90]	.74 [.72,.76]	.66 [.64,.69]	.73 [.70,.75]	.69 [.66,.71]	.59 [.56,.61]	.52 [.50,.55]	.41 [.37,.45]
Banking	.92 [.89,.95]	.83 [.80,.87]	.69 [.66,.74]	.67 [.62,.71]	.69 [.64,.74]	.69 [.65,.74]	.58 [.54,.64]	.44 [.38,.53]
Banking, large firms	.95 [.91,.99]	.86 [.82,.91]	.70 [.65,.76]	.73 [.67,.79]	.69 [.63,.75]	.62 [.56,.68]	.51 [.45,.57]	.33 [.25,.42]
Insurance	1.00 [.96,1.05]	.89 [.84,.96]	.72 [.67,.78]	.61 [.56,.66]	.49 [.45,.53]	.31 [.28,.34]	.28 [.24,.31]	.27 [.21,.35]
Insurance, large firms	1.02 [.96,1.08]	.92 [.85,.99]	.73 [.66,.81]	.59 [.53,.67]	.52 [.46,.58]	.29 [.25,.33]	.22 [.18,.27]	.23 [.15,.32]

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: pX corresponds to the X^{th} percentile. See Section 3 for the estimation method. Bounds of the confidence intervals reported in brackets are estimated by bootstrap (100 replications).

Table 3: Linear specification of the access function

Sector	National market			Segmented markets		
	Const	Slope	Stat	Const	Slope	Stat
All firms	.80 [.79,.81]	.28 [.26,.29]	.453			
Large firms	.74 [.72,.76]	.09 [.07,.14]	.828	.69 [.64,.70]	.05 [-.02,.07]	.434
Banking	.71 [.67,.75]	.07 [-.01,.15]	.066			
Banking, large firms	.83 [.77,.89]	.26 [.16,.39]	.077	.77 [.67,.82]	.25 [.11,.33]	.300
Insurance	.90 [.85,.95]	.60 [.54,.69]	.078			
Insurance, large firms	.93 [.82,1.01]	.66 [.53,.79]	.056	.74 [.60,.79]	.41 [.23,.51]	.339

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: we report the estimated coefficients of a linear specification of the access function, $h(u) = a - bu$ where “Const” refers to a and “Slope” refers to b . Bounds of the confidence intervals are estimated by bootstrap (100 replications) and are given in brackets. We also report the statistic of a specification test for which the threshold at the 5% level is .461. The method used to estimate the coefficients and to compute the test statistic is detailed in Appendix B.

Table A.1: Descriptive statistics on age and country of birth by gender

Sector	All				Females				Males			
	Age		Foreign		Age		Foreign		Age		Foreign	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
All firms	42.4	1.7	.10	.31	42.4	1.7	.11	.31	42.4	1.7	.10	.30
Large firms	42.5	1.7	.10	.30	42.4	1.7	.11	.31	42.5	1.7	.10	.30
Banking	42.4	1.7	.09	.29	42.3	1.7	.10	.30	42.5	1.7	.09	.28
Banking, large firms	42.4	1.7	.09	.29	42.3	1.7	.10	.30	42.4	1.7	.09	.29
Insurance	42.5	1.7	.08	.27	42.4	1.7	.08	.27	42.5	1.7	.08	.27
Insurance, large firms	42.5	1.7	.07	.26	42.4	1.7	.07	.25	42.6	1.7	.07	.26

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45.

Note: "Foreign" refers to being born in a foreign country.

Table A.2: Descriptive statistics on the panel of workers born on octobre of an even year

All						Females						Males					
Age			Foreign			Age			Foreign			Age			Foreign		
Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std
43.0	43.0	1.6	.12	0	.32	42.9	43	1.6	.13	0	.34	43.0	43.0	1.6	.11	0	.32
Wage			Experience			Wage			Experience			Wage			Experience		
Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std
134	114	87	15.4	16.5	4.5	116	101	61	15.4	16.4	4.3	139	118	93	15.4	16.5	4.6

Source: DADS, 2003, full-time executives of the Private Sector aged 40-45 born in october of an even year.

Note: “Med” refers to the median. “Foreign” refers to being born in a foreign country.

Table A.3: Highest qualifications by gender computed from the 2003 Labour Force Survey

Highest qualifications	Males	Females	All
Graduate	38.4%	39.9%	38.8%
Undergraduate	24.9%	29.0%	25.9%
High school completed	17.1%	17.3%	17.2%
Below high school	19.6%	13.9%	18.2%
Number of observations	1,281	404	1,685

Source: 2003 French Labour Force Survey, full-time executives of the Private Sector aged 40-45.

Note: Statistics are weighted for the sample to be representative of the population.

Table A.4: Age at the end of initial schooling by gender computed from the 2003 Labour Force Survey

Statistic	Males	Females	All
Mean	21.47	21.20	21.41
Standard error	.09	.15	.08
Number of observations	1,281	404	1,685

Source: 2003 French Labour Force Survey, full-time executives of the Private Sector aged 40-45.

Note: Statistics are weighted for the sample to be representative of the population.