J Popul Econ (2010) 23:881–905
DOI 10.1007/s00148-009-0266-3

ORIGINAL PAPER

Wages and employment of French workers with African origin

Romain Aeberhardt · Denis Fougère · Julien Pouget · Roland Rathelot

Received: 2 July 2007 / Accepted: 3 June 2009 /
Published online: 23 July 2009

Abstract Our study focuses on the differences in wages and employment between French workers with French parents and French workers with at least one African parent, using the Formation Qualification Professionnelle survey (Insee, Paris, 2003). We introduce econometric decompositions, which allow us to reach conclusions when the potentially discriminated group is small. Then, we clarify the impact of discrimination at the hiring level in this context. We find that unexplained parts in the employment decompositions are much larger than in the wage decompositions. This suggests that, in France, labor market discrimination is more frequent at the hiring level than in the compensation process.

Keywords Discrimination · Wage differentials · African migrants

JEL Classification C24 · J31 · J71

Responsible editor: James Albrecht

R. Aeberhardt · D. Fougère · J. Pouget · R. Rathelot
Crest-Insee, 15 Bld G. Péri, 92244 Malakoff Cedex, France

D. Fougère
CNRS, Paris, France

D. Fougère
CEPR, London, UK

D. Fougère · J. Pouget
IZA, Bonn, Germany

R. Rathelot (✉)
Dares, 39/43 Quai A. Citroen, 75902 Paris Cedex 15, France
e-mail: roland.rathelot@travail.gouv.fr

1 Introduction

For more than 40 years, economists and econometricians, following Becker (1957), Arrow (1973), and Phelps (1972), have developed theoretical and empirical tools to study discrimination in the labor market. There have been a number of empirical studies in which attempts were made to decompose observed employment rates and earnings differentials into human capital and "discrimination" components. Altonji and Blank (1999) proposes a comprehensive survey of this literature. One of the decomposition methods that is the most often used was popularized by Oaxaca (1973) and Blinder (1973). Using this decomposition method, many US studies conclude that, although differences in worker-observable characteristics are important factors of the Black–White wage differential, current labor market discrimination may account for at least one-third of the overall gap in the USA.

However, these hypothesized "skill" and "treatment" components may lead to ambiguous interpretations. The so-called "treatment" or "discrimination" component may be over-estimated due to unobservable heterogeneity. Another twist in the wage gap decomposition methodology is caused by a potential selectivity bias. This is why more general approaches were proposed (see, for example, the papers by Oaxaca and Ransom (1994), Neuman and Oaxaca (2004), and Neuman and Oaxaca (2005)). Other studies tried to account for the fact that controlling for worker productivity may correspond to inaccurate measures of workers' skills. For instance, Neal and Johnson (1996) use the Armed Forces qualification test as a better measure of skill. This test is taken before entry in the labor market and is, therefore, less likely to be contaminated by worker's choices or labor market discrimination. A different set of studies, known as audit studies, attempts to place comparable minority and non-minority actors into actual social and economic contexts and to measure how each group performs in these contexts (see Heckman 1998). These audit studies, like, for instance, the field experiment led by Bertrand and Mullainathan (2003), provide some of the cleanest non-laboratory evidence of differential treatment by race.

In spite of this vast literature on racial discrimination, little attention has been devoted to the French case. This lack may partly be due to the fact that the French republican and egalitarian ideal is very cautious when dealing with this topic. However, November 2005 riots, occurring simultaneously in various poor suburbs of large cities where immigrants are over-represented, suddenly highlighted the problem of discrimination in the French labor market.

Since 1975, the proportion of immigrants in the population has remained stable in France (7.4% in 1999), but their geographical origin has evolved (Insee, 2005). In 1962, most of them came from Europe (79%), especially from Italy and Spain, and only 15% came from Africa. In 1999, 45% came from Europe and 39% came from Africa, especially from North Africa. Immigrants are more affected by unemployment: their unemployment rate (16.4% in 2002) is twice that of non-immigrants (8.2%). They are more often manual

workers or employees, especially in unskilled jobs, and are over-represented in manufacturing and construction.

In 1999, people born in France with two migrant parents represented 5% of the group of persons aged 66 and less. While 20% of individuals aged 19 to 29 with non-migrant parents are unemployed, the unemployment rate is 30% for those with two migrant parents. However, their labor market situation depends on their parents' country of origin: their unemployment rate is nearly 40% if their parents come from Algeria or Morocco, whereas it is slightly under 20% when they come from Southern Europe (Spain, Italy, Portugal). These numbers naturally raise the question of migrants' children labor market integration, but also of their potential discrimination. The situation of the children of African immigrants in the suburbs of French cities is particularly at stake.

Using longitudinal data coming from the French population censuses, Fougère and Safi (2009) show that being granted French citizenship has a positive impact on the employment probability of immigrants. This "naturalization premium" seems particularly important for migrant groups facing difficulties when entering the labor market, that is, mostly men from sub-Saharan Africa and from Morocco, and women from Turkey and from North Africa. Silberman and Fournier (1999) and Meurs et al. (2006) also suggest that children of immigrants might suffer from discrimination in the labor market. Pouget (2005) focuses on employment in the public sector. Aeberhardt and Pouget (2007) perform a switching regression model of wage determination and occupational employment that leads them to favor an interpretation in terms of occupational segregation, rather than mere wage discrimination. They use business survey data and, therefore, cannot take into account the selectivity bias associated with the non-employment status.

Our paper empirically examines both employment and wage differences between French workers with different national origins. This is the first such econometric analysis on a representative sample of the French population. For that purpose, we use a unique household survey, the *Formation Qualification Professionnelle* survey (hereafter referred as the FQP survey) performed in 2003 by the National Institute for Statistics and Economic Studies (Insee, Paris). This survey contains many socio-demographic and economic variables, as well as accurate information on the residential area, especially the so-called "*Zones Urbaines Sensibles*" (ZUS), which are distressed areas often concentrating the migrant population. The program called "*Zones Urbaines Sensibles*" (ZUS) was launched in 1995; it concerns 751 disadvantaged zones that receive extra public resources and benefit from tax exemptions. In these zones, the unemployment rate is very high (25.4% in 1999, 39.5% for workers aged 15 to 24); the proportion of migrants is also very high (16.5% in 1999, vs. 5.6% in France). In order to identify the potential effects of discrimination, we estimate a selection model allowing for the possible endogeneity of the employment situation. Due to the small sample size of the potentially discriminated group, we introduce a new methodology based on the use of two

counterfactual parameters that are calculated from the estimates of the model parameters associated with the largest (i.e. the non-discriminated) group. This method proves to give more precise estimates than the usual ones, but at the cost of giving up exact decompositions.

The structure of our paper is as follows. Section 2 presents the methodology. Section 3 provides details on the data source. Section 4 outlines the main empirical findings.

2 Methodology

In general, empirical evidence of wage and employment discrimination toward workers of foreign origin is established through the decomposition method initiated by Oaxaca (1973) and Blinder (1973). Much of the literature based on the Oaxaca–Blinder decomposition interprets the part of the wage gap accounted for by differences in coefficients as discrimination, while it would be more appropriate to use the term "unexplained part" of the decomposition. In what follows, we use this latter term as often as possible. Methods taking into account selectivity terms within this framework were introduced by Oaxaca and Ransom (1994), Neuman and Oaxaca (2004), and Neuman and Oaxaca (2005). Our contribution is inspired by their work and goes further in that sense.

2.1 The model

We denote w_{ij} as the log-wage of the individual i in demographic group $j \in \{A, B\}$. Individuals belonging to group B are potentially discriminated. We suppose that the wage is generated by the following model:

$$w_{ij} = X_i^{'}\beta_j + u_{ij} \tag{1}$$

The wage is only observed for employed individuals. A binary variable E_{ij} is set to 1 when i is employed, and 0 otherwise. It is generated by a latent random variable E_{ij}^*, which is positive if and only if worker i is employed (and thus if the wage is observed):

$$E_{ij} = 1_{\left\{E_{ij}^* > 0\right\}} \tag{2}$$

with

$$E_{ij}^* = Z_i^{'}\gamma_j + \varepsilon_{ij} \tag{3}$$

All observations within the same group are assumed to be independent and identically distributed. Errors u_{ij} and ε_{ij} are assumed to have zero mean. Correlation between u_{ij} and ε_{ij} is allowed. The error terms are assumed to be jointly normally distributed, i.e.:

$$\begin{pmatrix} \varepsilon_{ij} \\ u_{ij} \end{pmatrix} \Big| X_i, Z_i \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho_j\sigma_j \\ \rho_j\sigma_j & \sigma_j^2 \end{bmatrix}\right) \tag{4}$$

This last assumption allows us to estimate the model with a maximum likelihood procedure.

2.2 Decomposition of the employment gap

The decomposition of the difference in employment probabilities across groups is a generalized form of the traditional Oaxaca–Blinder decomposition:

$$\begin{aligned}\mathbb{E}[E_{iA}] - \mathbb{E}[E_{iB}] = \mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)] - \mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)] \\ + \mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)] - \mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)] \end{aligned} \quad (5)$$

Under simple regularity conditions on the distributions of Z_i, empirical counterparts of these terms are:

$$(1/N_j)\sum_{i\in j} E_i \overset{a.s.}{\to} \mathbb{E}_{Z_j}[\mathbb{E}(E_{ij}|Z_i)] = \mathbb{E}[E_{ij}] \quad (6)$$

$$(1/N_B)\sum_{i\in B} \Phi\left(Z_i\hat{\gamma}_A\right) \overset{a.s.}{\to} \mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)] \quad (7)$$

Let us remark that this decomposition can be performed by using estimates of the probit model applied to group A only. This property appears to be very useful when one of the two groups is too small for getting precise estimates, as is the case in our empirical application.

2.3 Decomposition of the wage gap

2.3.1 The no-correlation case

When the correlation between the errors of the two equations is zero, the ordinary least squares (OLS) estimates of the wage equation parameters are unbiased. The difference between expected log-wages may then be split into two terms:

$$\mathbb{E}[w_{iA}] - \mathbb{E}[w_{iB}] = \left(\mathbb{E}_A\left[X_i'\right] - \mathbb{E}_B\left[X_i'\right]\right)\beta_A + \mathbb{E}_B\left[X_i'\right](\beta_A - \beta_B) \quad (8)$$

The first term of this sum is due to the average gap in individual characteristics between the two groups. The second one is the unexplained one. It is usually interpreted as the discrimination component.

Under simple regularity conditions on the distributions of X_i, empirical counterparts of these terms are:

$$\sum_{i\in j}\frac{E_i w_i}{\sum_{i'\in j} E_{i'}} \overset{a.s.}{\to} \mathbb{E}_j\left[X_i'\right]\beta_j = \mathbb{E}[w_{ij}] \quad (9)$$

$$\sum_{i\in B}\frac{E_i X_i'}{\sum_{i'\in B} E_{i'}}\hat{\beta}_A \overset{a.s.}{\to} \mathbb{E}_B\left[X_i'\right]\beta_A \quad (10)$$

Let us remark once again that we need only to estimate the model for group A to perform this decomposition.

2.3.2 The selectivity terms

When the correlation between the errors is not zero, an OLS estimation of the sole wage equation gives biased estimates. A maximum likelihood procedure or the Heckman two-step procedure may therefore be performed to overcome this problem. The difference between expected log-wages of employed workers in the two groups can then be written as:

$$\begin{aligned}\mathbb{E}\left[w_{iA}|E_{iA}=1\right]-\mathbb{E}\left[w_{iB}|E_{iB}=1\right] = &\left(\mathbb{E}_A\left[X_i'|E_{iA}=1\right]-\mathbb{E}_B\left[X_i'|E_{iB}=1\right]\right)\beta_A\\ &+\mathbb{E}_A\left[X_i'|E_{iA}=1\right](\beta_A-\beta_B)\\ &+\rho_A\sigma_A\mathbb{E}_A\left[\lambda_{iA}|E_{iA}=1\right]\\ &-\rho_B\sigma_B\mathbb{E}_B\left[\lambda_{iB}|E_{iB}=1\right] \qquad (11)\end{aligned}$$

In this expression, λ_{ij} is the inverse Mills' ratio, defined as:

$$\lambda_{ij}=\frac{\varphi\left(Z_i'\gamma_j\right)}{\Phi\left(Z_i'\gamma_j\right)}, \qquad (12)$$

where φ and Φ are, respectively, the density and the cumulative density function of the standard normal distribution.

The first two terms of Eq. 11 may be interpreted as before. The last one is attributed to the difference in selectivity terms between the two groups. Neuman and Oaxaca (2004) and Neuman and Oaxaca (2005) attempt to go one step further by splitting these selectivity terms and by incorporating the resulting parts either in the explained or in the unexplained component of the decomposition. This approach relies on conventional choices that we do not want to make. Moreover, in this approach, estimation has to be performed in both groups, which may be problematic when the sample size of the potentially discriminated group is small.

2.3.3 Decomposition of the marginal expectations

To avoid this problem, one solution is to work with marginal expectations instead of conditional ones. This allows us to get back to the initial decomposition:

$$\mathbb{E}[w_{iA}]-\mathbb{E}[w_{iB}]=\left(\mathbb{E}_A\left[X_i'\right]-\mathbb{E}_B\left[X_i'\right]\right)\beta_A+\mathbb{E}_B\left[X_i'\right](\beta_A-\beta_B) \qquad (13)$$

However, this approach has a drawback. Since empirical counterparts of marginal expectations are not fully observed, wages of non-employed individuals have to be imputed. Consequently, we must use the estimated β values to assign a wage to each non-employed individual. Since group B is potentially small, parameter estimates for this group have, in general, high standard errors. This small-sample problem contaminates the decomposition terms, which are then quite imprecise.

A further problem is that some of the X variables appearing in the wage equation (for instance, firm seniority) are not observed for unemployed work-

ers. A solution is to estimate the expected values of these covariates for each unemployed worker, given the values of her observed covariates. This may be done first by estimating a regression model for the group of employed workers (either a linear or a probit model, depending on the qualitative or quantitative nature of the missing covariate), and then by simulating the expected value of each unobserved covariate from the estimates of this regression model. In our empirical application, we have checked that the results are not significantly affected by the choice of the imputation method.

2.3.4 Counterfactuals

In this paper, we want to treat the following two problems simultaneously:

1. First, the error terms of our two equations (employment and wages) could be correlated, which means that the wage equation could be affected by a selectivity bias.
2. Second, we would like to run the estimation on group A only, as group B is too small to provide precise estimates.

The method we introduce allows us to tackle these two problems. However, it does not produce exact decompositions. To be more precise, we focus on the following two parameters:

$$\Delta_1 = \frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]} \tag{14}$$

$$\Delta_2 = \frac{\mathbb{E}_{X_B,Z_B,E_B}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)E_{iB}]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}E_{iB}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} \tag{15}$$

Since $\mathbb{E}(w_{ij}E_{ij}|X_i,Z_i) = \mathbb{E}(w_{ij}|E_{ij}=1,X_i,Z_i)\mathbb{E}(E_{ij}|Z_i)$, these two parameters of interest can also be written as:

$$\Delta_1 = \frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)\mathbb{E}(E_{iA}|Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)\mathbb{E}(E_{iA}|Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]} \tag{16}$$

$$\Delta_2 = \frac{\mathbb{E}_{X_B,Z_B,E_B}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)E_{iB}]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}|X_i,Z_i,E_{iB}=1)\mathbb{E}(E_{iB}|Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} \tag{17}$$

The first term in the right-hand side (r.h.s.) of Δ_1 is the conditional expected wage of employed persons in group A, given that they work. The second term in the same r.h.s. is a counterfactual expected wage. It corresponds to the average log-wage that a "non-discriminated" person would earn if she had the

same characteristics X and Z as the "average" member of group B. The first term in the r.h.s. of Δ_2 is the average wage that an employed member of group B would earn if she had group A's coefficients in the wage equation. Note that this counterfactual term also appears in Neuman and Oaxaca (2005). The second term in the r.h.s. of Δ_2 is the conditional expected wage of employed persons in group B, given that they work. If endowments had equal returns in both groups, the parameter Δ_2 would be clearly zero, even if endowments were differently distributed across groups.

Empirical counterparts of expectations appearing in the r.h.s. of Δ_1 and Δ_2 are:

$$\sum_{i\in j} \frac{w_i E_i}{\sum_{i'\in j} E_{i'}} \xrightarrow{a.s.} \frac{\mathbb{E}_{X_j,Z_j}[\mathbb{E}(w_{ij}E_{ij}|X_i,\, Z_i)]}{\mathbb{E}_{Z_j}[\mathbb{E}(E_{ij}|Z_i)]} \tag{18}$$

$$\sum_{i\in B} \left(\frac{\Phi(Z_i\hat{\gamma}_A)}{\sum_{i'\in B}\Phi(Z_{i'}\hat{\gamma}_A)}\right)\left(X_i\hat{\beta}_A + \hat{\rho}_A\hat{\sigma}_A\frac{\varphi(Z_i\hat{\gamma}_A)}{\Phi(Z_i\hat{\gamma}_A)}\right)$$
$$\xrightarrow{a.s.} \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iA}|X_i,\, Z_i,\, E_{iA}=1)\mathbb{E}(E_{iA}|Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]} \tag{19}$$

$$\sum_{i\in B} \frac{E_i}{\sum_{i'\in B} E_{i'}}\left(X_i\hat{\beta}_A + \hat{\rho}_A\hat{\sigma}_A\frac{\varphi(Z_i\hat{\gamma}_A)}{\Phi(Z_i\hat{\gamma}_A)}\right)$$
$$\xrightarrow{a.s.} \frac{\mathbb{E}_{X_B,Z_B,E_B}[\mathbb{E}(w_{iA}|X_i,\, Z_i,\, E_{iA}=1)E_{iB}]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} \tag{20}$$

These three empirical conterfactuals only make use of the parameter estimates of the model for the largest group, here denoted A. For instance, the second empirical counterfactual consistently estimates the conditional expectation of the wage w_{iA} for a person i in group B, given that she would be employed if she were a member of group A (which corresponds to $E_{iA}=1$). The third empirical counterfactual makes sense if we assume that an employed individual i of the potentially "discriminated" group B (i.e. $E_{iB}=1$) would also be employed if she were a member of the "non-discriminated" group A (i.e. $E_{iA}=1$).

The choice of these empirical counterfactuals comes from the fact that, under usual regularity conditions on X_j and Z_j, the following three expressions have the same limit.

$$\sum_{i\in j} \frac{w_i E_i}{\sum_{i'\in j} E_{i'}} \tag{21}$$

$$\sum_{i\in j} \left(\frac{\Phi(Z_i\hat{\gamma}_j)}{\sum_{i'\in j}\Phi(Z_{i'}\hat{\gamma}_j)}\right)\left(X_i\hat{\beta}_j + \hat{\rho}_j\hat{\sigma}_j\frac{\varphi(Z_i\hat{\gamma}_j)}{\Phi(Z_i\hat{\gamma}_j)}\right) \tag{22}$$

$$\sum_{i\in j} \frac{E_i}{\sum_{i'\in j} E_{i'}}\left(X_i\hat{\beta}_j + \hat{\rho}_j\hat{\sigma}_j\frac{\varphi(Z_i\hat{\gamma}_j)}{\Phi(Z_i\hat{\gamma}_j)}\right) \tag{23}$$

The first expression is simply the average wage of employed individuals in group j. The second expression is the weighted mean of the conditional expected wages of all members of group j given that they would be employed, the weight being equal to their probability of being employed. The third expression is the unweighted average of the conditional expected wages of the employed individuals in group j, given that they are employed.

2.3.5 Decompositions and entry barriers

The selection process creates a specific problem with respect to the decomposition of the raw wage gap that is observed between the two subpopulations. Since the selection process may differ across the two populations, it has to be accounted for in the interpretation of the gap; otherwise, one would compare populations that do not face the same selection process. This remark invites us to reinterpret the two parameters that we have introduced above. More precisely, we show below that Δ_1 and Δ_2 correspond, respectively, to the structural part and to the unexplained part of two different "raw wage gaps" that cannot be computed without estimating the model on the smallest group (i.e., the potentially discriminated group). On the contrary, Δ_1 and Δ_2 may be estimated without using parameter estimates of the model corresponding to the smallest group.

The "raw wage gap" corresponding to Δ_1 is the wage gap that would be observed if the selection process for population B were the same as the selection process affecting population A. This first "raw wage gap" is then equal to:

$$\frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]} \tag{24}$$

Figure 1 illustrates this issue. The observed "raw wage gap" corresponds to the difference between mean wages of employed persons in groups A and B, whereas Δ_1 involves the virtual population B^*, which corresponds to the individuals of group B who would be working if they faced the same selection as individuals from group A.

Similarly, the "raw wage gap" corresponding to Δ_2 is the wage gap that would be observed if the selection process for population A were the same as the selection process for population B. This second "raw wage gap" is then equal to:

$$\frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}E_{iB}|X_i,Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iB}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}E_{iB}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} \tag{25}$$

The observed mean wage gap (i.e. the observed "raw wage gap") is equal to:

$$\Delta = \frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}E_{iB}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} \tag{26}$$

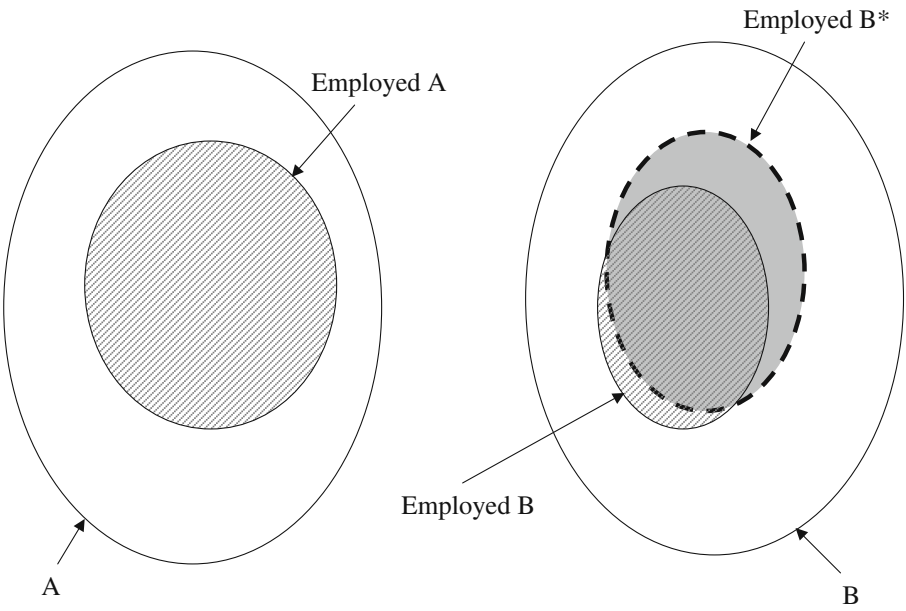

Fig. 1 Entry barriers and differences of support

Let us remark that this observed wage gap cannot be decomposed using Δ_1 and Δ_2. Consequently, $\Delta - \Delta_1$ and $\Delta - \Delta_2$ have no simple interpretation. However, this is not a major problem, since Δ_1 and Δ_2 are meaningful by themselves. The structural gap Δ_1 may be interpreted as an aggregated indicator of the effects of the overall differences in observable endowments between the two populations in the hypothetical situation in which both would face the selection process of population A. Since the wages are considered in logarithms, the unexplained gap Δ_2 can be seen, at the first order, as the mean of the relative difference between the observed wage and the expected wage for the potentially discriminated population B. All estimations and computations of the present work have been done both with SAS and R (see R Development Core Team 2007; Toomet and Henningsen 2008).

3 Data

3.1 The Formation Qualification Professionnelle survey (FQP, 2003)

To our knowledge, the survey FQP performed by Insee (Paris) in 2003 is the first major French survey that collects information on national origin of persons, as well as information on wages and employment for a representative sample of the French population. Similar surveys were conducted in 1970,

1977, 1985, and 1993 by Insee (Paris), 2, 3, or 4 years after a population census. However, these previous surveys do not contain any information on the national origin of individuals.

Using a complex sampling design, the FQP surveys cover all men and women in metropolitan France with a quite substantial number of individual face-to-face interviews (39,285 in 2003). In France, these surveys are used to study the returns to education, the efficiency of the educational system, the impact of social origin on academic and professional success, or the impact of vocational training on careers. They also permit to conduct studies on specific populations, e.g., the rise of unemployment among high school dropouts in the nineties. The questionnaire is made of five parts: professional mobility, initial education, vocational training, social origin, and earnings. FQP is the only survey that allows one to link these five topics and to observe their interactions. Many questions in the 2003 survey are the same as in the previous surveys. However, the 2003 survey focuses on professional mobility with a particular emphasis on the professional career in the past 5 years. Special attention is also put upon organizational and technological changes that employees face during their career.

The reference population consists in all individuals between 18 and 65 years old, who live in France (metropolitan area) in an ordinary dwelling. Within each dwelling, if more than two persons can be potentially interviewed, only two are randomly drawn and surveyed. The initial sample comprises 40,000 dwellings. Due to vacancies and refusal of participation, the final sample contains about 40,000 individuals. The survey is conducted in face-to-face interviews using CAPI (computer assisted personal interviewing). After the description of the household, which takes about 3 minutes, the survey questionnaire takes about 30 minutes per person. The data collection took place between April and July 2003.

3.2 Sample and groups considered for the analysis

3.2.1 Scope of the study

Our final sample includes all wage earners as well as the non-working population, except students and retired individuals. This choice can of course be challenged because of the potential endogeneity of the decision regarding the length of studies and the enrollment in early retirement plans, but they seem to be appropriate for our study.

The model distinguishes between those who receive wages and those who do not. Therefore we exclude from our analysis those who receive only non-wage compensations (they account for a very small part of the population). Here again, we could have modelled intermediate decisions, but the quality of the estimates would have been probably very poor given the very small size of this particular subsample. We also leave aside those who do not answer the wage question and those who say they do not know it.

In this survey, sampled individuals are asked about their professional situation at the time of the interview (2003) and other information that allows us to know their situation in the labor market in 2002 (in particular their annual earnings in 2002). Since earnings are key variables in our study and are available only for the year 2002, we have to reconstruct explicitly the individual situation in the labor market during that year. One question allows us to know directly whether the person worked in 2002 and earned a wage. It also informs us on those who earned non-wage compensations. Moreover, among those who did not work in 2002, we can distinguish students and retired people.

3.2.2 Sub-populations of interest

Most of our results concern two subsamples of French individuals. First, those with at least one parent who had the citizenship of an African country at birth (Maghreb included), second, those whose both parents were French at birth and born in France. We exclude those for whom the citizenship at birth of at least one of the parents is unknown, except if only one citizenship is known and corresponds to an African country.

The group with French parents is the reference group, while the other group corresponds to the group of potentially discriminated individuals. Note that, despite the fact that we can identify the subsample of "second generation" migrants, this subsample is too small for being treated separately in the estimation procedure. Thus we present only some descriptive statistics concerning this specific subsample.

3.2.3 Unemployment

The 2003 FQP survey informs accurately on the individual labor market situation at the date of the interview, while a calendar describes the past five years of the individual labor market history. However, this calendar has in general too many missing values to be used directly. Here we describe briefly a method to distinguish unemployed individuals from persons who were inactive. This distinction is used to compare the different subpopulations when calculating descriptive statistics, but it is not used to estimate the model. The difficulty is to find, among those who did not work in 2002, those who were actually unemployed.

First we distinguish between the individuals who worked in 2002 and those who did not. Among those who did not, we check if they ever worked before. Among those who never worked, we keep only the unemployed who were not students in 2002. Among those who had a job in the past, some left this job less than five years ago and others more than five years ago. We have only very few information about the latter group. Consequently, we consider as unemployed those who were unemployed at the time of the interview. For those whose last job was occupied continuously during the last five years, we

have more information, including their current situation and the reason why they left this last job. Among them, we consider as unemployed those who became unemployed after leaving their last job and who are still unemployed at the time of the interview. A few people declare themselves as unemployed just after leaving their last job but are out of the labor force (retired, back to school or university, or inactive) at the time of the interview. And among those who declare themselves as unemployed, some left their job for health or family reasons, *i.e.* another reason than a layoff, a quit or the termination of a temporary labor contract. In that case we do not know whether these individuals participated in the labor market in 2002, and thus we exclude them from the unemployed group. We might therefore slightly underestimate the number of unemployed people by putting some of them into the inactive group.

As shown in Table 1, individuals of African origin are relatively much more numerous in all precarious situations: in 2002, 9.3% of them were unemployed for twelve months whereas only 3.5% were in that situation in the reference group. They are also much more likely to be inactive or to have worked less than twelve months during the year.

3.2.4 Outcomes and covariates

We distinguish between different categories of individuals, according to their gender, their marital status, the presence of children and the presence of a working spouse in the household. Household composition is different across groups (see Tables 1 and 2).

In particular, single women with children, but also men and women having children and a non-working spouse, are relatively more numerous among persons of African origin. On the contrary, women without children are less represented in the latter group.

For individuals of African origin, the distribution of ages is shifted to the left: they are more numerous in the youngest age groups (see Tables 1 and 2). Table 3 reports exact sample sizes.

There are much more people without any diploma among individuals of African origin. Proportions of individuals with higher educational levels are more similar, except for vocational degrees which are relatively less common in the group of persons having an African origin. Table 4 shows that there is a huge difference between the two groups, both in terms of concentration around Paris and in the proportions of people residing in a ZUS: individuals of African origin are much more concentrated in the Paris region and in ZUS areas.

The variable of interest is the logarithm of the individual wage. More precisely, we use the wage in a full-time, full-year equivalent. Distributions and means of this variable for the different sub-populations are shown in Table 4. This table also shows labor market status and workers' occupations for both groups.

Table 1 Descriptive statistics

National origin of the parents	France	Africa
Number of observations	22,255	894
Gender		
Female	54.2	53.9
Male	45.8	46.1
Age		
Less than 20	0.5	1.6
20 to 29	15.4	24.6
30 to 39	29.4	34.9
40 to 49	27.0	24.9
50 to 59	24.0	11.9
60 and more	3.7	2.1
Educational level		
Graduate	11.7	11.1
College	11.1	8.5
High school	16.0	14.5
Vocational school	26.4	20.8
Junior high school	9.5	11.1
No diploma	25.3	34.0
Size of the household		
Single man without children	7.1	5.7
Single man with children	2.1	3.0
Single woman without children	7.1	6.7
Single woman with children	5.5	8.4
Man with working spouse, with children	17.0	12.1
Man with working spouse, without children	7.3	4.8
Man with non working spouse, with children	8.9	18.7
Man with non working spouse, without children	3.5	1.8
Woman with working spouse, with children	23.2	22.5
Woman with working spouse, without children	9.0	4.9
Woman with non working spouse, with children	4.2	9.2
Woman with non working spouse, without children	5.2	2.2
Poverty and region of residence		
Not poor, living out of the Paris region	81.9	46.6
Not poor, living in Paris and suburbs	13.3	30.2
Poor, living out of the Paris region	3.5	15.1
Poor, living in Paris and suburbs	1.2	8.1
Labor market situation in 2002		
Employed in a full-time job during 12 months	59.5	48.0
Employed in a part-time job during 12 months	9.9	6.0
Employed 12 months partly part-time, partly full-time	1.3	1.2
Unemployed during 12 months	3.5	9.3
Some work during the year (various situations)	13.7	18.7
Out of the labor force	11.9	16.8

All statistics are computed using individual weights. Proportions in columns sum to 100%.
Reading: Among French individuals whose both parents are French at birth, 54.2% are women.
Source: FQP survey, Insee, Paris, 2003

Table 2 Differences in observable covariates between employed and non-employed individuals

National origin of the parents	France		Africa	
	Non-employed	Employed	Non-employed	Employed
Number of observations	3,988	18,267	273	621
Gender				
Female	77.0	49.2	72.2	45.9
Male	23.0	50.8	27.8	54.1
Age				
Less than 20	1.6	0.3	4.4	0.3
20 to 29	14.4	15.6	30.4	22.1
30 to 39	21.9	31.0	33.7	35.4
40 to 49	19.4	28.6	16.8	28.5
50 to 59	31.6	22.4	11.7	11.9
60 and more	11.1	2.1	2.9	1.8
Educational level				
Graduate	6.7	12.8	5.9	13.4
College	5.6	12.3	5.5	9.8
High school	11.9	16.9	11.0	16.1
Vocational school	23.1	27.2	17.9	22.1
Junior high school	10.3	9.3	12.5	10.5
No diploma	42.5	21.6	47.3	28.2
Size of the household				
Single man without children	5.9	7.3	5.1	6.0
Single man with children	2.4	2.0	3.7	2.7
Single woman without children	6.8	7.1	4.8	7.6
Single woman with children	7.9	4.9	12.5	6.6
Man with working spouse, with children	3.3	20.0	4.0	15.6
Man with working spouse, without children	2.9	8.3	1.1	6.4
Man with non working spouse, with children	5.5	9.6	12.8	21.3
Man with non working spouse, without children	3.0	3.6	1.1	2.1
Woman with working spouse, with children	30.2	21.7	31.1	18.7
Woman with working spouse, without children	9.3	8.9	5.1	4.8
Woman with non working spouse, with children	8.8	3.2	15.4	6.4
Woman with non working spouse, without children	14.1	3.3	3.3	1.8
Region of residence				
Living out of a ZUS and out of the Paris region	83.4	81.6	50.5	44.9
Living out of a ZUS, but in Paris and suburbs	9.4	14.1	21.6	34.0
Living in a ZUS, but out of the Paris region	5.9	3.0	20.1	12.9
Living in a ZUS, in Paris and suburbs	1.3	1.2	7.7	8.2

All statistics are computed using individual weights. Proportions in sub-columns sum to 100%.
Reading: Among non-employed French individuals whose both parents are French at birth, 77.0% are women.
Source: FQP survey, Insee, Paris, 2003

Table 3 Number of observations (full sample)

National origin of parents	France	Africa
Number of observations	22,255	894
Gender		
Female	12,054	482
Male	10,201	412
Age		
Less than 20	113	14
20 to 29	3,422	220
30 to 39	6,542	312
40 to 49	6,001	223
50 to 59	5,346	106
60 and more	831	19
Educational level		
Graduate	2,600	99
College	2,472	76
High school	3,556	130
Vocational school	5,886	186
Junior high school	2,109	99
No diploma	5,632	304
Size of the household		
Single man without children	1,576	51
Single man with children	457	27
Single woman without children	1,578	60
Single woman with children	1,214	75
Man with working spouse, with children	3,781	108
Man with working spouse, without children	1,633	43
Man with non working spouse, with children	1,976	167
Man with non working spouse, without children	778	16
Woman with working spouse, with children	5,174	201
Woman with working spouse, without children	1,999	44
Woman with non working spouse with children	932	82
Woman with non working spouse, without children	1,157	20
Poverty and region of residence		
Not poor, living out of the Paris region	18,233	417
Not poor, living in Paris and suburbs	2,956	270
Poor, living out of the Paris region	788	135
Poor, living in Paris and suburbs	278	72
Labor market situation in 2002		
Employed in a full-time job during 12 months	13,246	429
Employed in a part-time job during 12 months	2,208	54
Employed 12 months partly part-time, partly full-time	298	11
Unemployed during 12 months	787	83
Some work during the year (various situations)	3,059	167
Out of the labor force	2,657	150

The figures correspond to the exact numbers of observations in each stratum. Interpretation: Among French individuals whose both parents are French at birth, there are 12,054 women and 10,201 men in the sample. Source: FQP survey, Insee, Paris, 2003

Table 4 Differences in observable covariates between individuals living in and out of a ZUS

National origin of the parents	France		Africa	
	out of a ZUS	ZUS	out of a ZUS	ZUS
Number of employed individuals	17,489	778	490	131
Working time (%)				
Part time	17.1	16.6	14.4	17.5
Full time	82.9	83.4	85.6	82.5
Occupation (%)				
Craftsman	2.6	1.3	3.6	1.0
Executive	15.8	9.4	16.6	1.0
Intermediate	27.6	26.9	20.7	22.3
White-collar	30.8	36.3	31.0	40.8
Skilled blue-collar	16.4	17.7	19.1	21.4
Unskilled blue-collar	6.8	8.4	9.0	13.6
Earnings (in euros)				
Mean	18,636	15,735	16,986	11,869
First quartile	11,639	10,976	10,539	6,980
Median	16,189	14,700	14,450	11,500
Third Quartile	22,867	19,967	20,399	15,245
Full-time full-year equivalent wage				
Mean	21,526	18,248	19,442	14,924
First quartile	13,150	12,522	12,000	9,661
Median	17,544	15,688	15,245	12,958
Third quartile	24,080	21,000	22,000	17,658
Total number of observations	21,189	1,066	687	207
Labor market situation in 2002 (%)				
Employed in a full-time job during 12 months	59.9	52.3	50.5	39.6
Employed in a part-time job during 12 months	10.0	8.0	6.0	6.3
Employed 12 months partly part-time, partly full-time	1.3	1.5	1.2	1.4
Unemployed during 12 months	3.3	7.3	8.2	13.0
Some work during the year (various situations)	13.7	14.8	18.0	20.8
Out of the labor force	11.7	16.0	16.2	18.8

All statistics are computed using individual weights. Proportions in sub-columns sum to 100%.
Reading: Among French workers living in a ZUS and whose both parents are French at birth, 83.4% work full-time.
Source: FQP survey, Insee, Paris, 2003

4 Results

We alternatively estimate our two-equation model with a two-step Heckman-type (H2S hereafter) procedure and a maximum likelihood estimation (MLE hereafter) procedure. These two procedures have different advantages and drawbacks. As pointed out by Wooldridge (Wooldridge 2002, p. 566), the MLE is more efficient than the two-step procedure under joint normality of the error terms, and it produces standard errors and likelihood ratio statistics that can be used directly. The drawbacks are that it is less robust than the two-step procedure and that it is sometimes difficult to get the problem to converge.

Our model is identified thanks to the introduction into the selection equation of variables that are supposed to have an impact on the employment probability but not directly on the wage. These variables are dummy variables

indicating whether the person lives in a couple, whether he/she has children, and whether his/her spouse is working. Mroz (1987) and Hyslop (1999), among others, provide strong evidence that such variables are indeed exogenous, at least with respect to the participation decision. Moreover, Table 5 shows that they have a statistically significant effect on the individual employment probability. Thus, they may be considered as valid instruments.

Estimation is separately conducted for both groups, namely, French individuals whose parents were both French at birth and French individuals with at least one parent who had the citizenship of an African country at birth. We first comment on the parameter estimates of our two-equation model, before using them to decompose the employment and wage gaps. We also run OLS estimations in order to assess the importance of the selectivity bias. Because the size of the potentially discriminated group is small, we put together men and women.

Results of the employment equation are available in Table 5. In each group, a higher education increases the probability to be employed. Individuals of African origin and without any education are slightly less employed than comparable individuals with French parents. Potential labor market experience has a positive but concave impact on the employment probability. Potential experience is probably a better measure of working experience for the native French population than for those of African origin, especially because these persons may not have been living in France for their entire working life, and because working life gained abroad may not be relevant for wages earned in France. We are aware of the measurement error bias that could result from the use of potential experience. However, we should also admit that the use of exact working experience (which is not observed in our data base anyway) is not less problematic because of its potential endogeneity.

Socio-demographic variables are also statistically significant determinants of this probability. Several variables are interacted: gender, marital status, presence of children in the household, and employment status of the spouse. Our results are similar to those obtained in previous studies. Single women with children are less employed than single men and single women without children (the reference situation). This result is verified in each group but even more pronounced for women whose father or mother was African at birth. Men with a working spouse and with children are more often employed, whereas women in the same situation behave the opposite way: this pattern is similar in both groups. Men with a working spouse and without children tend to work more than the reference category, but the gap is higher among those of African origin. Women with a working spouse behave the same way in both groups: they are less often employed when they raise children and they are as often employed as the reference category when they do not. The gap between the two groups increases for women whose spouse is not employed. Whereas women of African origin are less often employed when they have children, their employment probability is significantly different from the reference population when they do not have children.

Table 5 Estimates of the employment equation parameters

Covariates	French parents		African parents	
	H2S	MLE	H2S	MLE
Intercept	0.66[a]	0.66[a]	0.30	0.30
	(0.05)	(0.05)	(0.22)	(0.22)
Household composition				
Single men and single women without children	Ref.	Ref.	Ref.	Ref.
Single women with children	−0.44[a]	−0.43[a]	−0.56[b]	−0.56[b]
	(0.05)	(0.05)	(0.20)	(0.20)
Men with a working spouse, with children	0.54[a]	0.55[a]	0.39	0.39
	(0.05)	(0.05)	(0.22)	(0.22)
Men with a working spouse, without children	0.50[a]	0.50[a]	0.96[b]	0.96[b]
	(0.06)	(0.06)	(0.34)	(0.34)
Men with a non-working spouse, with children	0.24[a]	0.24[a]	0.24	0.24
	(0.05)	(0.05)	(0.17)	(0.17)
Men with a non-working spouse, without children	0.15[c]	0.16[c]	0.34	0.34
	(0.06)	(0.06)	(0.39)	(0.39)
Women with a working spouse, with children	−0.53[a]	−0.53[a]	−0.53[a]	−0.53[a]
	(0.03)	(0.03)	(0.16)	(0.16)
Women with a working spouse, without children	−0.05	−0.05	−0.10	−0.10
	(0.04)	(0.04)	(0.24)	(0.24)
Women with a non-working spouse, with children	−0.49[a]	−0.49[a]	−0.60[b]	−0.60[b]
	(0.05)	(0.05)	(0.19)	(0.20)
Women with a non-working spouse, without children	−0.57[a]	−0.57[a]	−0.31	−0.31
	(0.05)	(0.05)	(0.32)	(0.32)
Residence				
Out of a ZUS, out of the Paris region	Ref.	Ref.	Ref.	Ref.
Out of a ZUS, out of the Paris region	0.12[a]	0.12[a]	0.26[c]	0.26[c]
	(0.04)	(0.04)	(0.12)	(0.12)
In a ZUS, but out of the Paris region	−0.24[a]	−0.24[a]	0.09	0.09
	(0.05)	(0.05)	(0.14)	(0.14)
In a ZUS, and in the Paris region	−0.04	−0.04	0.34	0.34
	(0.09)	(0.09)	(0.19)	(0.19)
Experience	0.12[a]	0.12[a]	0.11[a]	0.11[a]
	(0.00)	(0.00)	(0.01)	(0.01)
Experience squared	−0.28[a]	−0.28[a]	−0.24[a]	−0.24[a]
	(0.01)	(0.01)	(0.04)	(0.04)
Educational level				
University graduate	Ref.	Ref.	Ref.	Ref.
College	0.06	0.06	0.10	0.10
	(0.05)	(0.05)	(0.24)	(0.24)
High-school	−0.18[a]	−0.18[a]	−0.12	−0.12
	(0.05)	(0.05)	(0.21)	(0.21)
Vocational school	−0.45[a]	−0.45[a]	−0.27	−0.27
	(0.04)	(0.04)	(0.20)	(0.20)
Junior high-school	−0.52[a]	−0.52[a]	−0.52[c]	−0.52[c]
	(0.05)	(0.05)	(0.22)	(0.22)
No diploma	−0.74[a]	−0.74[a]	−0.66[a]	−0.66[a]
	(0.04)	(0.04)	(0.19)	(0.19)
Number of observations	22,255	22,255	894	894

Standard errors are between parentheses.
Source: FQP survey, Insee, Paris, 2003

[a]Statistically significant at the 1% level

[b]Statistically significant at the 5% level

[c]Statistically significant at the 10% level

The area where a person lives also has an impact on her employment probability. To characterize the area of residence, we consider interactions between two variables: living in the region of Paris (called *Île-de-France*) and living in a ZUS-disadvantaged area. The reference situation is the case in which the person lives neither in *Île-de-France* nor in a ZUS. For individuals with French-born parents, living in *Île-de-France* improves the employment probability, whereas living in a ZUS located outside *Île-de-France* drastically diminishes it. For this group, there is no statistically significant difference in the employment probabilities of persons living in a ZUS in the region of Paris and the reference category. Individuals of African origin living in Paris but not in a disadvantaged area (i.e., in a ZUS) have a higher employment probability, but this probability is still higher for these persons when they live in disadvantaged areas located in the Paris region (although the estimated parameter is only significant at the 10% level). Those living outside the region of Paris have a lower employment probability, which may help to explain why they concentrate in the vicinity of Paris.

Parameter estimates of the wage equation for both groups are reported in Table 6. Effects of potential experience and education are as usual: hump-shaped for potential experience, increasing with the level of education. We introduce firm seniority in the equation, even though such a variable may be potentially endogenous (see Beffy et al. 2006, for empirical evidence on this issue). We find that there exists a wage premium for workers who have been employed for more than 5 years in a firm. As usual, we also note that women earn less than men. Interestingly, there are no major differences in the coefficients associated with gender, seniority, experience, and education across the two groups. The main differences concern the intercept, and the coefficients associated with the occupation of a full-time job, with a college education, and with residence in a ZUS (i.e., in a disadvantaged) area out of the Paris region (namely, the wage of workers of African origin is significantly lower when they live in a ZUS area that is located outside the Paris region). When using the two-step estimation procedure, we find that the estimated coefficients associated with the inverse Mills' ratios in the wage equations are, respectively, equal to -0.055 for the French native group, and to 0.022 for the group of individuals of African origin. Both estimates are statistically significant at the 10% level, which indicates the presence of a selectivity bias. These results are confirmed by the values of the ML estimates of the correlation coefficients between the error terms of the two equations of our model, which are both statistically significant at the 10% level.

Now we get to the main results, those concerning the decomposition of wage and employment gaps between the two groups. Decompositions are summarized in Tables 7 and 8. Table 7 reports usual wage and employment decompositions, while Table 8 presents the earnings gaps estimated with our alternative approach. The gaps calculated for the population called "second generation only" are generally wider; this is particularly true for the structural term Δ_1. The "second generation" includes persons who have at least one

Table 6 Estimates of the wage equation parameters

Covariates	French parents			African parents		
	OLS	H2S	MLE	OLS	H2S	MLE
Intercept	9.96[a]	9.98[a]	9.96[a]	10.01[a]	10.00[a]	10.01[a]
	(0.02)	(0.03)	(0.02)	(0.11)	(0.15)	(0.12)
Working time						
Part-time	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Full-time	−0.06[a]	−0.07[a]	−0.07[a]	−0.08	−0.08	−0.08
	(0.01)	(0.01)	(0.01)	(0.06)	(0.06)	(0.06)
Gender						
Men	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Women	−0.25[a]	−0.24[a]	−0.25[a]	−0.17[a]	−0.18[b]	−0.18[c]
	(0.01)	(0.01)	(0.01)	(0.05)	(0.07)	(0.05)
Residence						
Out of a ZUS, out of the Paris region	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Out of a ZUS, in the Paris region	0.22[a]	0.22[a]	0.22[a]	0.21[a]	0.22[a]	0.21[a]
	(0.01)	(0.01)	(0.01)	(0.05)	(0.05)	(0.05)
In a ZUS, out of the Paris region	−0.05[b]	−0.04	−0.05[b]	−0.20[c]	−0.20[c]	−0.20[c]
	(0.02)	(0.02)	(0.02)	(0.07)	(0.07)	(0.07)
In a ZUS, in the Paris region	0.11[c]	0.11[c]	0.11[c]	0.10	0.10	0.10
	(0.03)	(0.03)	(0.03)	(0.09)	(0.09)	(0.08)
Experience	0.02[a]	0.02[a]	0.02[a]	0.02[b]	0.02	0.02[b]
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Experience squared	−0.04[a]	−0.03[a]	−0.04[a]	−0.03	−0.03	−0.03
	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
Seniority						
Less than 1 year	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
1 to 5 years	0.06[a]	0.06[a]	0.06[a]	0.04	0.04	0.04
	(0.01)	(0.01)	(0.01)	(0.07)	(0.07)	(0.07)
5 to 10 years	0.11[a]	0.11[a]	0.11[a]	0.07	0.07	0.07
	(0.02)	(0.02)	(0.02)	(0.08)	(0.08)	(0.08)
More than 10 years	0.25[a]	0.25[a]	0.25[a]	0.21[c]	0.21[c]	0.21[c]
	(0.02)	(0.02)	(0.02)	(0.08)	(0.08)	(0.08)
Educational level						
University graduate	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
College	−0.19[a]	−0.19[a]	−0.19[a]	−0.34[a]	−0.34[a]	−0.34[a]
	(0.02)	(0.02)	(0.02)	(0.10)	(0.09)	(0.09)
High-school	−0.38[a]	−0.37[a]	−0.38[a]	−0.52[a]	−0.52[a]	−0.52[a]
	(0.01)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)
Vocational school	−0.60[a]	−0.59[a]	−0.60[a]	−0.55[a]	−0.55[a]	−0.55[a]
	(0.01)	(0.01)	(0.01)	(0.08)	(0.08)	(0.08)
Junior high-school	−0.51[a]	−0.50[a]	−0.50[a]	−0.55[a]	−0.56[a]	−0.55[a]
	(0.02)	(0.02)	(0.02)	(0.09)	(0.10)	(0.09)
No diploma	−0.75[a]	−0.73[a]	−0.75[a]	−0.61[a]	−0.62[a]	−0.61[a]
	(0.01)	(0.02)	(0.02)	(0.08)	(0.09)	(0.08)
Number of observations	18,267	18,267	22,255	621	621	894

Standard errors are between parentheses.
Source: FQP survey, Insee, Paris, 2003

[a]Statistically significant at the 1% level

[b]Statistically significant at the 5% level

[c]Statistically significant at the 10% level

Table 7 Decomposition of the earnings gap and of the employment gap between French workers with French parents and French workers with at least one African parent

Estimation method	Raw gap	Explained	Unexplained	95% confidence interval for the explained part
Employment				
Full sample	0.126	0.065	0.062	[0.055 , 0.074]
Second generation only	0.155	0.093	0.062	[0.082 , 0.104]
Wages—full sample				
OLS	0.120	0.055	0.065	[0.043 , 0.067]
Marginal (MLE)	0.140	75%	25%	[44% , 233%]
Wages—second generation only				
OLS	0.197	0.133	0.064	[0.120 , 0.146]
Marginal (MLE)	0.233	84%	16%	[−196% , 493%]

For the marginal MLE decomposition, the raw gap is estimated and not observed. Confidence intervals at the 95% level are obtained by a pseudo-bootstrap technique.
Source: FQP survey, Insee, Paris, 2003

African parent but who were born in France. This subpopulation is very small; consequently, the corresponding estimates have very poor precision.

Table 7 reports the decomposition of the wage gap when the wage equation is estimated by OLS. In this case, almost half of the gap is not explained by the differences between mean values of covariates. However, if there exists a selection process correlated with the wage formation process, the OLS estimator is biased. The marginal Blinder–Oaxaca decomposition, which requires estimating the wage and employment equations within each group separately, is also shown in Table 7. Results obtained through H2S and MLE procedures are similar (here, we report only the MLE results). They contrast with the OLS estimates since the explained part grows up now to 75%. Thus, the unexplained part is limited to approximately one quarter of the total wage gap.

Table 8 refers to our counterfactual approach, which is only based on the estimation of the wage and employment equations for the reference group. This method yields more precise estimates, since the reference group is large enough, but we lose the possibility of performing usual decompositions. The results show that the structural gap Δ_1 is relatively large, especially for the "second generation." But still, there remains an unexplained difference Δ_2 of about 5% for the employed individuals.

Table 8 Counterfactual approach: earnings differentials between French workers with French parents and French workers with at least one African parent

Sample	Raw gap	Structural gap ($\hat{\Delta}_1$)	95% conf. int. for $\hat{\Delta}_1$	Unexplained gap ($\hat{\Delta}_2$)	95% conf. int. for $\hat{\Delta}_2$
Full sample	0.120	0.079	[0.066 , 0.091]	0.055	[0.043 , 0.067]
Second generation only	0.197	0.170	[0.157 , 0.183]	0.053	[0.041 , 0.065]

As explained in Subsection 2.3.5, this table does not report usual Oaxaca–Blinder decompositions; in particular, the sum of $\hat{\Delta}_1$ and $\hat{\Delta}_2$ is not equal to the raw gap. Confidence intervals at the 95% level are obtained by a pseudo-bootstrap technique.
Source: FQP survey, Insee, Paris, 2003

Concerning the employment probability, all the decompositions suggest that the unexplained part is higher than for the wage gap, around 47%. In the literature dealing with discrimination, there are few papers presenting explicit confidence intervals. A possible explanation is that traditional Oaxaca–Blinder decompositions usually provide imprecise results when the sample size is too small: 95% confidence intervals often include the [0, 1] interval. This is the main advantage of our "counterfactual" method: it provides more precise estimates, even when one of the groups has a small size. In our case, a pseudo-bootstrap technique applied to the marginal MLE decomposition shows that the explained part of the wage gap ranges from 44% to 233%, which means that we are pretty confident to be able to explain at least 44% of the wage gap, and maybe all of it.

Our results prove that there exists a strong difference in the employment probabilities of the two groups, which is largely unexplained by usual covariates. Once workers are hired, there is still a wage gap between the two groups. However, this gap is lower than the gap between the employment probabilities.

5 Conclusion

Our paper contains the first estimates of the wage gap and of the employment probability gap between French workers whose both parents were French at birth and French workers with at least one parent who had the nationality of an African country at birth. Data come from the survey "*Formation et Qualification Professionnelle*" conducted by Insee (*Institut National de la Statistique et des Études Économiques*, Paris) in 2003.

In general, econometric methods of wage decompositions yield imprecise estimates resulting from the small sample size of the minority groups. In order to circumvent this problem, we propose a new method relying on the use of two counterfactual parameters that are calculated from the estimates of the model parameters associated with the largest (i.e. the non-discriminated) group only. The first parameter corresponds to the structural part of the "raw wage gap" that would be observed if the selection process applied to the potentially discriminated group were the same as the selection process affecting the non-discriminated group. The second parameter corresponds to the unexplained part of the "raw wage gap" that would be observed if the selection process affecting the non-discriminated group were the same as the selection process of the potentially discriminated group. Using this new method, we obtain estimates that prove to be more precise than usual estimates, even if the sample size of the potentially discriminated group is rather small. This comes, however, at the cost of giving up exact decompositions.

Our estimates suggest that one half of the employment gap is not explained by differences in usual covariates, such as age, gender, education, potential experience, residential area, etc., while the unexplained wage difference for those who work is of about 5%. These results are in line with those obtained

from audit studies on the hiring process, suggesting that the French labor market is characterized by a substantial discrimination against second-generation African workers who apply for vacant jobs (see, for instance, Cediey et al. 2008). In other terms, if there is discrimination against first- and second-generation migrants in the French labor market, this discrimination is more present at entry into employment than in the job.

Acknowledgements All correspondence to Roland Rathelot. The authors thank Jim Albrecht, Manon Domingues dos Santos, Guy Laroque, David Neumark, Sébastien Roux, Patrick Simon, François-Charles Wolff, and two anonymous referees for helpful discussions and remarks. Participants at Alliance Conference on Group Inequality and Discrimination (Columbia University, May 2005), INSEE Research Seminar (Paris, June 2005), Soccult Workshop (Paris, February 2006), INED Seminar (Paris, May 2006), MIRE Seminar (Paris, October 2006), JMA Conference (Fribourg, June 2007), EALE Conference (Budapest, August 2007), EEA Conference (Oslo, September 2007), TEPP Conference (Evry-sur-Seine, December 2007), and at the 2nd IZA Migration Topic Week (Bonn, May 2008) provided useful comments. All the remaining errors are ours.

References

Aeberhardt R, Pouget J (2007) National origin wage differentials in France: evidence from matched employer-employee data. IZA discussion paper no 2779

Altonji J, Blank R (1999) Race and gender in the labor market. In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3C. Elsevier, Amsterdam, pp 3143–3259

Arrow KJ (1973) The theory of discrimination. In: Ashenfelter O, Rees A (eds) Discrimination in labor markets. Princeton University Press, Princeton, pp 3–33

Becker G (1957) The economics of discrimination. University of Chicago Press, Chicago

Beffy M, Buchinsky M, Fougère D, Kamionka T, Kramarz F (2006) The returns to seniority in France (and why are they lower than in the United States). IZA discussion paper no 1935

Bertrand M, Mullainathan S (2003) Are Emily and Greg More employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am Econ Rev 94(4):159–190

Blinder A (1973) Wage discrimination: reduced form and structural estimates. J Hum Resour 8(4):436–455

Cediey E, Foroni F, Garner H (2008) Discriminations à l'embauche fondées sur l'origine à l'encontre de jeunes français(es) peu qualifié(e)s. Premières Informations - Premières Synthèses 06.3, DARES, Paris

Fougère D, Safi M (2009) Naturalization and employment of immigrants in France (1968–1999). Int J Manpow 30(1–2):83–96

Heckman JJ (1998) Detecting discrimination. J Econ Perspect 12(2):101–116

Hyslop DR (1999) State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. Econometrica 67(6):1255–1294

Meurs D, Pailhé A, Simon P (2006) The persistence of intergenerational inequalities linked to immigration: labour market outcomes for immigrants and their descendants in France. Population 61(5):645–682

Mroz TA (1987) The sensitivity of an empirical model of married women's hours to economic and statistical assumptions. Econometrica 55(4):765–799

Neal DA, Johnson WR (1996) The role of premarket factors in black-white wage differences. J Polit Econ 104(5):869–895

Neuman S, Oaxaca R (2004) Wage decompositions with selectivity corrected wage equations: a methodological note. J Econ Inequal 2(1):3–10

Neuman S, Oaxaca R (2005) Wage differentials in the 1990s in Israel: endowments, discrimination, and selectivity. Int J Manpow 26(3):217–236

Oaxaca R (1973) Male-female wage differentials in urban labor markets. Int Econ Rev 14(3):693–709

Oaxaca R, Ransom M (1994) On discrimination and the decomposition of wage differentials. J Econom 61(1):5–21
Phelps ES (1972) The statistical theory of racism and sexism. Am Econ Rev 62(4):659–661
Pouget J (2005) La Fonction publique : vers plus de diversité? In: France, portrait social, édition 2005–2006. Insee, Paris, pp 143–162
R Development Core Team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Silberman R, Fournier I (1999) Les enfants d'immigrés sur le marché du travail. Les mécanismes d'une discrimination sélective. Form Empl 65:31–55
Toomet O, Henningsen A (2008) Sample selection models in R: package sampleSelection. J Stat Softw 27(7):1–23
Wooldridge JM (2002) Econometric analysis of cross section and panel data. MIT, Cambridge