Wages and employment of French workers with African origin [§]

Romain Aeberhardt[¶] Denis Fougère[‖]
Julien Pouget[**] Roland Rathelot[††]

May 9, 2008

Abstract

Our study proposes an econometric decomposition of the wage gap and of the difference in employment probabilities between French workers whose both parents had French citizenship at birth and French workers with at least one parent who had the citizenship of an African country at birth. For that purpose, we use data coming from the Formation Qualification Professionnelle (FQP) survey conducted by Insee (Paris) in 2003. Our study is the first to estimate both employment and wage differentials between "native" French workers and children of African migrants. We find that one half of the employment gap is not explained by differences in observable covariates between the two groups and that the unexplained wage difference for those who work is of about 5%. This result is obtained by using a new method yielding more precise results when the sample size of the potentially discriminated group is small.

Keywords: discrimination, wage differentials, second-generation migrants.
JEL: C24, J31, J71.

[§]The authors thank participants at the Alliance Conference on Group Inequality and Discrimination (Columbia University, May 2005), the Séminaire Recherche de l'Insee (Paris, June 2005), the JMA Conference (Nantes, June 2007), EALE Conference (Budapest, August 2007) and EEA Conference (Oslo, September 2007) for their stimulating comments. All the remaining errors are ours. Any opinions expressed here are those of the authors and not of any institution.

[¶]CREST - INSEE (Paris), romain.aeberhardt@insee.fr

[‖]CNRS, CREST-INSEE (Paris), CEPR (London) and IZA (Bonn), denis.fougere@ensae.fr

[**]CREST - INSEE (Paris) and IZA (Bonn), julien.pouget@ensae.fr

[††]CREST - INSEE (Paris), roland.rathelot@insee.fr. Corresponding author. Address: INSEE DESE - Timbre G220 - 15 Bld G. Péri - BP100 - 92244 Malakoff Cedex - France - Tél. : 33 1 41 17 60 17 - Fax. : 33 1 41 17 60 45

1 Introduction

For more than forty 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. The comprehensive survey by Altonji and Blank (1999) presents the main econometric studies dealing with discrimination. There has 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. One of the decomposition methods that is most often used was popularized by Oaxaca (1973) and Blinder (1973). Most U.S. studies conclude that although differences in worker-observable characteristics are important factors of the Black-White wage differential, the current labor market discrimination may account for at least one-third of the overall gap.

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 (2004a) and Neuman and Oaxaca (2004b)). 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 settings and to measure how each group fares in these settings (see Heckman (1998)). These audit studies provide some of the cleanest non-laboratory evidence of differential treatment by race. Bertrand and Mullainathan (2003) performed such a field experiment to measure racial discrimination in the labor market.

In spite of this vast literature on racial discrimination issues, little attention has been devoted to the French case. This lack is partly due to the fact that the French republican and egalitarian political model prevents from defining "ethnic" statistical categories. 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 represent 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 (2005) show that being granted French citizenship has a positive impact on the employment probability of immigrants. This "naturalization premium" seems particularly important for immigrant 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, Pailhé, and Simon (2005), suggest that children of immigrants might also suffer from discrimination in the labor market. Pouget (2005) focuses on the employment in the public sector. Aeberhardt and Pouget (2007) perform a switching regression model of wage determination and occupational employment which leads them to favour 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 unemployment status.

Our paper is the first econometric analysis that examines empirically both employment and wage differences between French workers with different national origins. For that purpose, we use a unique household survey, the Formation Qualification Professionnelle survey (here and after referred as 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, and also accurate information on the residential area, especially the so-called "Zones Urbaines Sensibles" (ZUS) which are distressed areas often concentrating the migrant population. [1] 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 counterfactual groups whose observable covariates are distributed as those of the potentially discriminated group. This method proves to give more precise estimates than the usual ones, but at the cost that we have to give up exact decompositions.

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

2 Methodology

Empirical evidence of wage and participation discrimination toward workers of foreign origin is established through the decomposition method initiated by Oaxaca (1973) and Blinder (1973). Methods taking into account selectivity terms within this framework were introduced by Oaxaca and Ransom (1994), Neuman and Oaxaca (2004a) and Neuman and Oaxaca (2004b). Our contribution is inspired by their work and goes further in that sense.

[1] The program called "Zones Urbaines Sensibles" (ZUS) was launched in 1995; it concerns 751 disadvantaged zones which receive public extra 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)

2.1 The model

We denote w_{ij} 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 that is positive if and only if worker i is employed (and thus if the wage is observed).

$$\begin{aligned} E_{ij}^{*} &= Z_i^{'}\gamma_j + \varepsilon_{ij} \\ E_{ij} &= 1_{\{E_{ij}^{*}>0\}} \end{aligned}$$

All observations within the same group are assumed independent and identically distributed. Errors u_{ij} and ε_{ij} are assumed zero-mean. Correlation between u_{ij} and ε_{ij} is *a priori* allowed. We assume for joint normality to estimate the system by maximum likelihood.

$$\left(\begin{array}{c|c} \varepsilon_{ij} \\ u_{ij} \end{array} \quad X_i, Z_i \right) \sim \mathcal{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left[\begin{array}{cc} 1 & \rho_j\sigma_j \\ \rho_j\sigma_j & \sigma_j^2 \end{array} \right] \right) \tag{2}$$

2.2 Decomposition of the employment gap

The decomposition of the difference in employment proportions across groups is a generalized form of the traditional Oaxaca (1973)-Blinder (1973) decomposition:

$$\begin{aligned} \mathbb{E}[E_{iA}] - \mathbb{E}[E_{iB}] &= \mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)] - \mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)] \\ &+ \mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)] - \mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)] \end{aligned} \tag{3}$$

Under simple regularity conditions on the distributions of Z_i, empirical counterparts are the following.

$$\begin{aligned} (1/N_j)\sum_{i\in j} E_i &\overset{p.s.}{\to} \mathbb{E}_{Z_j}[\mathbb{E}(E_{ij}|Z_i)] = \mathbb{E}[E_{ij}] \\ (1/N_B)\sum_{i\in B} \Phi\left(Z_i\hat{\gamma}_A\right) &\overset{p.s.}{\to} \mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)] \end{aligned}$$

Note that this decomposition may be performed only using results from estimation on group A. This is an important feature when, as it is the case in our empirical application, one of the two groups is too small to obtain precise estimates.

2.3 Decomposition of the wage gap

2.3.1 The no-correlation case

In a special case, the correlation between the errors of the two equations is zero. In this case, the wage equation can be estimated with no bias by OLS and the difference between expected

log-wages may be split into two terms.

$$\mathbb{E}[w_{iA}] - \mathbb{E}[w_{iB}] = \left(\mathbb{E}_A[X_i'] - \mathbb{E}_B[X_i']\right)\beta_A + \mathbb{E}_B[X_i](\beta_A - \beta_B) \tag{4}$$

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 are the following.

$$\begin{aligned}
\sum_{i\in j} \frac{E_i w_i}{\sum_{i'\in j} E_{i'}} &\overset{p.s.}{\to} \mathbb{E}_j[X_i']\beta_j = \mathbb{E}[w_{ij}] \\
\sum_{i\in B} \frac{E_i X_i'}{\sum_{i'\in B} E_{i'}} \hat{\beta}_A &\overset{p.s.}{\to} \mathbb{E}_B[X_i']\beta_A
\end{aligned}$$

Note that, as in the decomposition of the employment probabilities, we need only to estimate the model on group A.

2.3.2 The selectivity terms

It is usually not correct to assume that the correlation of the errors of the two equations is zero. Therefore, OLS estimation of the sole wage equation leads to biased results. MLE or Heckman (1979) procedure has to be performed to obtain unbiased estimates. The difference between expected log-wages of employed workers in the two groups can be written as:

$$\begin{aligned}
\mathbb{E}\left[w_{iA}|E_{iA}=1\right] - \mathbb{E}\left[w_{iB}|E_{iB}=1\right] &= \left(\mathbb{E}_A[X_i'|E_{iA}=1] - \mathbb{E}_B[X_i'|E_{iB}=1]\right)\beta_A \\
&+ \mathbb{E}_A[X_i'|E_{iA}=1]\left(\beta_A - \beta_B\right) \\
&+ \rho_A\sigma_A\mathbb{E}_A\left[\lambda_{iA}|E_{iA}=1\right] - \rho_B\sigma_B\mathbb{E}_B\left[\lambda_{iB}|E_{iB}=1\right]
\end{aligned} \tag{5}$$

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

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

where $\varphi(.)$ and Φ are respectively the pdf and the cdf of a normal distribution.

The first two terms of (5) may be interpreted as before. The last one is attributed to the difference in selectivity terms between the two groups. Neuman and Oaxaca (2004a) and Neuman and Oaxaca (2004b) attempt to go one step further in splitting these selectivity terms and to incorporate the chunks in the explained or unexplained component. This approach relies on conventional choices we do not want to make. Moreover, estimation has to be performed on both groups.

2.3.3 Decomposition of the marginal expectations

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

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

This comes, however, at a high cost. Empirical counterparts of marginal expectations are not exactly observed, as wages are imputed for individuals who are not employed. This means that we have to rely on the β's to assign a wage for every individual. Here, because group B is small, estimations on this group will have high standard errors. This will contaminate the decomposition results, for which we thus only achieve a poor precision.

Moreover, a practical problem method is that some of the X variables in the wage equation are not observed when the worker is not employed. For example, firm seniority can obviously not be observed for unemployed workers. When such covariates are not observed, we have to estimate their expected values, given the values of other observed covariates, from a regression model[2] estimated on the group of employed workers (either an OLS or a probit model, depending on the qualitative or quantitative nature of the missing covariate).

2.3.4 Counterfactuals

We have to deal with two issues here:

1. In our application, it is not reasonable to assume that there is no correlation between errors. Thus, we have to deal with selectivity.
2. We only want to carry estimation on group A, as group B is too small to obtain a satisfying precision on the results.

We present here a method using counterfactuals making it possible to meet both conditions. However, this comes at the cost that we have to give up exact decompositions.
We compute the following two differentials:

$$\begin{aligned}
\Delta_1 &= \frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]} \\
\Delta_2 &= \frac{\mathbb{E}_{X_B,Z_B,E_B}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)E_{iB}]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}E_{iB}|X_i,Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]}
\end{aligned}$$

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

$$\begin{aligned}
\Delta_1 &= \frac{\mathbb{E}_{X_A,Z_A}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)\mathbb{E}(E_{iA}|Z_i)]}{\mathbb{E}_{Z_A}[\mathbb{E}(E_{iA}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)\mathbb{E}(E_{iA}|Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]} \\
\Delta_2 &= \frac{\mathbb{E}_{X_B,Z_B,E_B}[\mathbb{E}(w_{iA}|X_i,Z_i,E_{iA}=1)E_{iB}]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]} - \frac{\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iB}|X_i,Z_i,E_{iB}=1)\mathbb{E}(E_{iB}|Z_i)]}{\mathbb{E}_{Z_B}[\mathbb{E}(E_{iB}|Z_i)]}
\end{aligned}$$

Δ_1 may directly be compared to the raw wage gap between the two groups. The second term of Δ_1 is a counterfactual expected wage. It amounts to the average log-wage that a non-discriminated individual would earn if he had the same characteristics X and Z than an average member of group B. Δ_2 corresponds to some unexplained component of the wage gap. The first term of Δ_2 is also a counterfactual wage. It is the average wage of an employed member of group B if he had group A's coefficients in the wage equation. If the two groups had equal returns to their endowments, Δ_2 would be clearly zero, even with different distributions of the

[2] In the empirical application, we have checked that our results are not significantly affected by the choice of the imputation method.

endowments across groups.

Empirical counterparts are the following:

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

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

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

The second empirical counterfactual is supposed to estimate the mean of w_{iA} over the individuals of B such that $E_{iA}=1$. In that sense, the numerator alone $\mathbb{E}_{X_B,Z_B}[\mathbb{E}(w_{iA}E_{iA}|X_i,Z_i)]$ would correspond to a mean over all individuals computed with w_{iA} for those for whom $E_{iA}=1$ and 0 for those for whom $E_{iA}=0$. In order to match an "observed" mean (in which we would not have the 0's), we correct for the proportion of individuals such that $E_{iA}=1$. This explains the term $\mathbb{E}_{Z_B}[\mathbb{E}(E_{iA}|Z_i)]$ at the denominator.

The third empirical counterfactual makes sense if we assume that an individual for whom $E_{iB}=1$ would also have $E_{iA}=1$.

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

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

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

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

The first expression is the average of the observed wages computed on the employed individuals of j only. The second is the weighted average of the expected wages conditional on being employed, computed on all individuals of j with weights equal to their probability of being employed. The third is the average of the expected wages conditional on being employed, computed on the employed individuals of j only.

2.3.5 Decompositions and Entry Barriers

In classical Oaxaca-Blinder type methods, exact decomposition can be performed and raw wage gaps are separated into two (or more) parts. In our case, the selection process creates a second issue, apart from biasing the coefficients. Since the selection process may differ across the two populations, it has to be accounted for in the interpretation of the gaps, otherwise one would compare populations which did not face the same selection process.

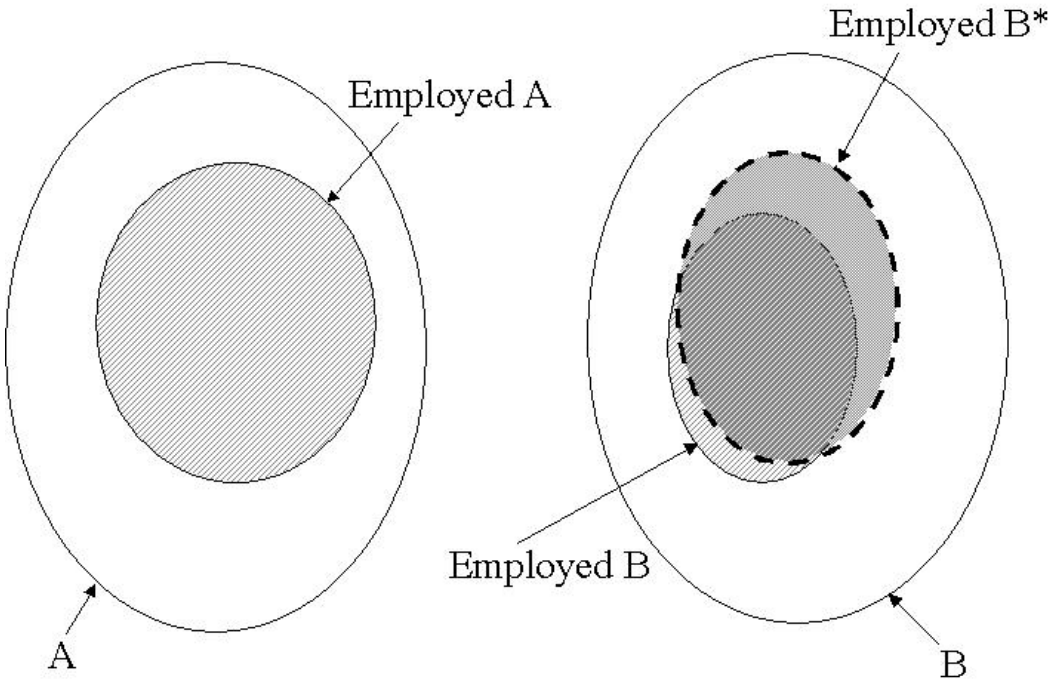

Figure 1: Entry barriers and differences of support

Our method allows us to deal with small samples but we lose the possibility of performing the usual type of decompositions: Δ_1 and Δ_2 could be interpreted as the structural and unexplained part of two different "raw wage gaps" which we can not compute here without estimating the model on the smaller population.

Indeed, the "raw wage gap" corresponding to Δ_1 would be the wage gap that would be observed if the selection process for population B were the same as for population A:

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

And the "raw wage gap" corresponding to Δ_2 would be the wage gap that would be observed if the selection process for population A were the same as for population B:

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

Figure 1 illustrates this issue of the different supports. The observed "raw wage gap" corresponds to the difference in wages between populations A and B, whereas for instance, Δ_1 involves the virtual population B*, which corresponds to the individuals of B who would be working if they faced the same selection as individuals from A.

In particular, the observed mean wage gap cannot be decomposed using $\hat{\Delta}_1$ and $\hat{\Delta}_2$.
If we note:

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

then, $\Delta - \Delta_1$ and $\Delta - \Delta_2$ have no simple interpretation and cannot be interpreted as an unexplained and a structural gap.

However, this does not seem to be a major problem as the terms Δ_1 and Δ_2 are meaningful by themselves. The structural gap may be interpreted as an aggregated indicator of the overall differences of observable endowments between the populations if they both faced the selection process of population A. Since the wages are considered in logarithm, the unexplained gap 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.

3 Data

To our knowledge, the survey "Formation Qualification Professionnelle" (hereafter, 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.

3.1 The Formation Qualification Professionnelle Survey (FQP, 2003)

The 2003 FQP survey follows similar surveys conducted in 1970, 1977, 1985 and 1993 by Insee (Paris), two, three or four years after a population census.

Using a complex sampling design, they 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 usually considered as offering a unique information about the returns to education, the efficiency of the educational system, the impact of social origin on academic and professional success, the impact of vocational training on careers, in terms of mobility or earnings. It also permits to conduct studies on specific populations, e.g. the rise of unemployment among high school drop-outs 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 to link these five topics and observe their interactions. Many questions in the 2003 survey are the same as in the previous surveys, conducted in 1964, 1970, 1977, 1985, and 1993. However, the 2003 survey focuses on professional mobility with a particular emphasis on the professional career in the past five years. Special attention is also put upon organizational and technological changes that employees face during their career.

The reference population consists in all individuals aged 18 to 65, who live in France (metropolitan area) in an ordinary dwelling. Within each dwelling, if there are more than two persons in the scope of the survey, 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 the studies and the enrollment in early retirement plans, but they seemed 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 modeled intermediate decisions, but the quality of the estimates would have probably been very poor given the very small size of this particular sample. 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 information about their professional situation at the time of the interview (2003) and other information that allows to know their situation in the labor market in 2002 (in particular their annual earnings). 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 on those who earned non-wage compensations. Among those who did not work in 2002, we need to identify students and retired people. For the students, we know the period of their studies. For the retired and early retired people, we consider that retirement is an absorbing state (that is those who retired in 2002 are still retired in 2003). Therefore we consider as retired in 2002, those who were retired in 2003 and who had left their last job in 2001 or before. By doing this, there is a risk that we get rid of those who were unemployed during their last year before retirement. In principle, this question could be assessed using the individual calendar of events, but so far, the high rate of non-response for this part of the survey does not allow us to use it efficiently.

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 the French parents is the reference group, and the other one corresponds to the group of potentially discriminated individuals. Since the reference group is relatively large, it allows us to impose conditions on both citizenship and country of birth of the parents, which should improve its homogeneity. Note that, despite the fact that we can identify the "second generation", since the sample size is too small, we present only some descriptive statistics concerning this specific subsample.

3.2.3 Unemployment

As explained before, the 2003 FQP survey informs accurately on the situation at the date of the interview, and a calendar describes the past five years of professional life (but it has too many missing values to be used directly). Here we describe briefly a method for distinguishing

unemployed individuals from persons who were inactive. This distinction is used to compare the different sub-populations when calculating descriptive statistics, but it is not used for estimating the model. The difficulty is to find, among those who did not work in 2002, those who were effectively 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 of them left it less than five years ago and others more than five years ago. For the latter, we have only very few information and we consider as unemployed those who were unemployed at the time of the interview. For those whose last job occurred in the last five years, we have more information, including their current situation and the reason why they left their last job. We consider as unemployed those who were unemployed when they left their last occupation and were 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 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 with African origin are relatively much more numerous in all precarious situations: 9.3 % of them were unemployed for twelve months whereas only 3.5 % were in that situation in the reference population. 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

The variable of interest is the logarithm of the individual wage. More precisely, we use the wage in full-time full-year equivalent. Distributions and means of this variable for the different sub-populations are shown in Table 3. This table also shows labor market status and workers' occupations for both groups.

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. Household composition is different in the two subgroups (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 opposite, women without children are less represented in the latter group.

For persons with 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). There are much more people without any diploma among individuals of African origin. The rest of the education distribution looks the same, except for vocational degrees which are relatively less common for those with African origin. Between the two groups, there is a huge difference both in terms of concentration around Paris and in the number of people residing in a "Zone Urbaine Sensible" (ZUS): individuals with African origin are much more concentrated in the Paris region and in ZUS areas.
Table 8 in Appendix presents sample sizes in more detail.

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
Diploma		
Graduate	11,7	11,1
Some College	11,1	8,5
Completed High School	16,0	14,5
Vocational Degree	26,4	20,8
Junior High School	9,5	11,1
No Diploma	25,3	34,0
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
Residence		
Not Poor × Rest of France	81,9	46,6
Not Poor × Paris and Suburbs	13,3	30,2
Poor × Rest of France	3,5	15,1
Poor × Paris and Suburbs	1,2	8,1
Situation on the Labor Market		
12 months FT	59,5	48,0
12 months PT	9,9	6,0
12 months FT/PT	1,3	1,2
12 months unemployed	3,5	9,3
some work (various situations)	13,7	18,7
no work (various situations)	11,9	16,8

Note: All statistics are computed using individual weights. All sub-columns sum to 100 %.
Reading: Among French individuals whose both parents are French at birth, 54.2 % are women.
Source: Formation Qualification Professionnelle survey (FQP), 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
Diploma				
Graduate	6,7	12,8	5,9	13,4
Some College	5,6	12,3	5,5	9,8
Completed High School	11,9	16,9	11,0	16,1
Vocational Degree	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
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
Residence				
Not ZUS × Rest of France	83,4	81,6	50,5	44,9
Not ZUS × Paris and Suburbs	9,4	14,1	21,6	34,0
ZUS × Rest of France	5,9	3,0	20,1	12,9
ZUS × Paris and Suburbs	1,3	1,2	7,7	8,2

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

Table 3: Differences in observable covariates between individuals living in and out of a ZUS

National origin of the parents	France		Africa	
	not ZUS	ZUS	not ZUS	ZUS
Employed individuals only	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
Professional Category				
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
Employee	30,8	36,3	31,0	40,8
Skilled Worker	16,4	17,7	19,1	21,4
Unskilled Worker	6,8	8,4	9,0	13,6
Earnings				
Mean	18 636	15 735	16 986	11 869
First Quarter	11 639	10 976	10 539	6 980
Median	16 189	14 700	14 450	11 500
Third Quarter	22 867	19 967	20 399	15 245
Wage (Full-Time Full-Year equivalent)				
Mean	21 526	18 248	19 442	14 924
First Quarter	13 150	12 522	12 000	9 661
Median	17 544	15 688	15 245	12 958
Third Quarter	24 080	21 000	22 000	17 658
Full Sample	21189	1066	687	207
Situation on the labor market				
12 months FT	59,9	52,3	50,5	39,6
12 months PT	10,0	8,0	6,0	6,3
12 months FT/PT	1,3	1,5	1,2	1,4
12 months unemployed	3,3	7,3	8,2	13,0
some work (various situations)	13,7	14,8	18,0	20,8
no work (various situations)	11,7	16,0	16,2	18,8

Note: All statistics are computed using individual weights. All 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: Formation Qualification Professionnelle survey (FQP), Insee, Paris, 2003.

4 Results

We alternatively estimate the Tobit model by a two-step Heckman-type procedure and by a maximum likelihood procedure. Our model is identified thanks to the introduction into the selection equation of variables which are supposed to have an impact on the employment probability but not directly on the wage. Socio-demographic variables (living in couple, having children, whether the spouse is working or not) seem to be valid instruments, since their impacts on employment are significant.

Estimation is separately done for both groups: 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 the results of our estimations, before using them to assess the potential existence of some discrimination in the labor market. We also run OLS estimations in order to measure the impact of the selectivity bias. Sample sizes drastically vary from one group to the other, bringing about the risk to jeopardize the significance of our estimates. This leads us to gather men and women in one single sample.

Results of the employment equation are available in Table 4. In each group, a higher education increases the probability to be employed. Individuals of African origin without any education are slightly less employed than comparable individuals with French parents. Potential experience has a positive but concave impact on the employment probability. Socio-demographic variables are also significant determinants of this probability. Several variables are interacted: gender, marital status, having children, whether the spouse is employed or not. 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 with an 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.

The area where a person lives has also an impact on her employment probability. To characterize the residence, we consider interactions between two variables: living in the region of Paris (called *Ile-de-France*) and living in a ZUS disadvantaged area. The reference situation is the case in which the person lives neither in Ile-de-France nor in a ZUS. For individuals with French born parents, living in Ile-de-France improves the employment probability whereas living in a ZUS located outside Ile-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. The situation is different for individuals with African parents. If they live in the region of Paris, but outside a ZUS, their wage is significantly higher. It is lower if they live in a ZUS that is located outside the region of Paris. The results for the employment equation are somewhat different. Individuals with an African origin living in Paris but not in a disadvantaged area have a higher employment probability, but it is still higher for persons living in disadvantaged areas located in the region of Paris (although the estimated parame-

ter is only significant at the 90% 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 5. 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, Buchinsky, Fougère, Kamionka, and Kramarz (2006), for empirical evidence on this issue). We clearly observe a wage premium for workers who have been employed more than five 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 between the two groups. The main differences concern the intercept, the full-time coefficient and the coefficient associated with a college degree (versus a post-graduate educational level).

Now we get to the main results, those concerning the decomposition of wage and employment gaps between the two groups. They are summarized in Table 6 and 7. Table 6 shows classical wage and employment decompositions, while table 7 shows the earnings differentials resulting from our counterfactual approach. The gaps with the population labelled "second generation only" are generally wider, but the structural part is also larger. The "second generation" includes persons who have at least one African parent but who were born in France. This subpopulation is very small and the results which require an estimation on this subsample have a very poor precision.

The OLS decomposition in table 6 gives a decomposition of the wage gap when the wage equation is estimated by ordinary least squares. 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 to estimate the wage and employment equations within each group separately is also shown in table 6. Results obtained through H2S and MLE procedures are similar (here we report only the MLE results). They contrast from the OLS estimates in that the explained part grows up to 75%. Thus the unexplained part is limited to approximately one quarter of the total wage gap.

Table 7 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 is relatively large, especially for the "second generation", but still, there remains an unexplained difference of about 5% for the individuals who work.

Concerning the employment probability, all the decompositions suggest that the unexplained

Table 4: Estimates of the employment equation parameters

Covariates	French parents		African parents	
	H2S	MLE	H2S	MLE
Intercept	0.66^{***} (0.05)	0.66^{***} (0.05)	0.30 (0.22)	0.30 (0.22)
Household composition				
Single men and single women w/o children	Ref.	Ref.	Ref.	Ref.
Single Women with children	-0.44^{***} (0.05)	-0.43^{***} (0.05)	-0.56^{**} (0.20)	-0.56^{**} (0.20)
Men with a working spouse, with children	0.54^{***} (0.05)	0.55^{***} (0.05)	0.39 (0.22)	0.39 (0.22)
Men with a working spouse, without children	0.50^{***} (0.06)	0.50^{***} (0.06)	0.96^{**} (0.34)	0.96^{**} (0.34)
Men with a non-working spouse, with children	0.24^{***} (0.05)	0.24^{***} (0.05)	0.24 (0.17)	0.24 (0.17)
Men with a non-working spouse, without children	0.15^{*} (0.06)	0.16^{*} (0.06)	0.34 (0.39)	0.34 (0.39)
Women with a working spouse, with children	-0.53^{***} (0.03)	-0.53^{***} (0.03)	-0.53^{***} (0.16)	-0.53^{***} (0.16)
Women with a working spouse, without children	-0.05 (0.04)	-0.05 (0.04)	-0.10 (0.24)	-0.10 (0.24)
Women with a non-working spouse, with children	-0.49^{***} (0.05)	-0.49^{***} (0.05)	-0.60^{**} (0.19)	-0.60^{**} (0.20)
Women with a non-working spouse, without children	-0.57^{***} (0.05)	-0.57^{***} (0.05)	-0.31 (0.32)	-0.31 (0.32)
Residence location				
Not in a ZUS, not in the region of Paris	Ref.	Ref.	Ref.	Ref.
Not in a ZUS, but in the region of Paris	0.12^{***} (0.04)	0.12^{***} (0.04)	0.26^{*} (0.12)	0.26^{*} (0.12)
In a ZUS, but not in the region of Paris	-0.24^{***} (0.05)	-0.24^{***} (0.05)	0.09 (0.14)	0.09 (0.14)
In a ZUS, in the region of Paris	-0.04 (0.09)	-0.04 (0.09)	0.34 (0.19)	0.34 (0.19)
Experience	0.12^{***} (0.00)	0.12^{***} (0.00)	0.11^{***} (0.01)	0.11^{***} (0.01)
Experience squared	-0.28^{***} (0.01)	-0.28^{***} (0.01)	-0.24^{***} (0.04)	-0.24^{***} (0.04)
Diploma				
University graduate	Ref.	Ref.	Ref.	Ref.
College	0.06 (0.05)	0.06 (0.05)	0.10 (0.24)	0.10 (0.24)
Completed high-school	-0.18^{***} (0.05)	-0.18^{***} (0.05)	-0.12 (0.21)	-0.12 (0.21)
Vocational high-school	-0.45^{***} (0.04)	-0.45^{***} (0.04)	-0.27 (0.20)	-0.27 (0.20)
Junior high-school	-0.52^{***} (0.05)	-0.52^{***} (0.05)	-0.52^{*} (0.22)	-0.52^{*} (0.22)
No diploma	-0.74^{***} (0.04)	-0.74^{***} (0.04)	-0.66^{***} (0.19)	-0.66^{***} (0.19)
Nobs	22255	22255	894	894

Notes: $*$ *means significant at the 90% level,* $**$ *means significant at the 95% level and* $***$ *means significant at the 99% level. Standard errors are between parentheses.*
Source: Formation Qualification Professionnelle survey (FQP), Insee, Paris, 2003.

Table 5: Estimates of the wage equation parameters

Covariates	French parents			African parents		
	OLS	H2S	MLE	OLS	H2S	MLE
Intercept	9.96*** (0.02)	9.98*** (0.03)	9.96*** (0.02)	10.01*** (0.11)	10.00*** (0.15)	10.01*** (0.12)
Working time						
Part-time	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Full-time	−0.06*** (0.01)	−0.07*** (0.01)	−0.07*** (0.01)	−0.08 (0.06)	−0.08 (0.06)	−0.08 (0.06)
Gender						
Men	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Women	−0.25*** (0.01)	−0.24*** (0.01)	−0.25*** (0.01)	−0.17*** (0.05)	−0.18* (0.07)	−0.18** (0.05)
Residence location						
Not in a ZUS, not in the region of Paris	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Not in a ZUS, but in the region of Paris	0.22*** (0.01)	0.22*** (0.01)	0.22*** (0.01)	0.21*** (0.05)	0.22*** (0.05)	0.21*** (0.05)
In a ZUS, but not in the region of Paris	−0.05* (0.02)	−0.04 (0.02)	−0.05* (0.02)	−0.20** (0.07)	−0.20** (0.07)	−0.20** (0.07)
In a ZUS, in the region of Paris	0.11** (0.03)	0.11** (0.03)	0.11** (0.03)	0.10 (0.09)	0.10 (0.09)	0.10 (0.08)
Experience	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02* (0.01)	0.02 (0.01)	0.02* (0.01)
Experience squared	−0.04*** (0.00)	−0.03*** (0.00)	−0.04*** (0.00)	−0.03 (0.02)	−0.03 (0.02)	−0.03 (0.02)
Seniority						
Less than 1 year	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
1 to 5 years	0.06*** (0.01)	0.06*** (0.01)	0.06*** (0.01)	0.04 (0.07)	0.04 (0.07)	0.04 (0.07)
5 to 10 years	0.11*** (0.02)	0.11*** (0.02)	0.11*** (0.02)	0.07 (0.08)	0.07 (0.08)	0.07 (0.08)
More than 10 years	0.25*** (0.02)	0.25*** (0.02)	0.25*** (0.02)	0.21** (0.08)	0.21** (0.08)	0.21** (0.08)
Diploma						
University graduate	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
College	−0.19*** (0.02)	−0.19*** (0.02)	−0.19*** (0.02)	−0.34*** (0.10)	−0.34*** (0.09)	−0.34*** (0.09)
Completed high-school	−0.38*** (0.01)	−0.37*** (0.01)	−0.38*** (0.01)	−0.52*** (0.08)	−0.52*** (0.08)	−0.52*** (0.08)
Vocational high-school	−0.60*** (0.01)	−0.59*** (0.01)	−0.60*** (0.01)	−0.55*** (0.08)	−0.55*** (0.08)	−0.55*** (0.08)
Junior high-school	−0.51*** (0.02)	−0.50*** (0.02)	−0.50*** (0.02)	−0.55*** (0.09)	−0.56*** (0.10)	−0.55*** (0.09)
No diploma	−0.75*** (0.01)	−0.73*** (0.02)	−0.75*** (0.02)	−0.61*** (0.08)	−0.62*** (0.09)	−0.61*** (0.08)
Nobs	18267	18267	22255	621	621	894

Notes: $*$ means significant at the 90% level, $**$ means significant at the 95% level and $***$ means significant at the 99% level. Standard errors are between parentheses.
Source: Formation Qualification Professionnelle survey (FQP), Insee, Paris, 2003.

Table 6: Decomposition of the earnings gap and the employment gap between French workers with French parents and French workers with at least one African parent

Estimation method	raw gap	explained	unexplained	C.I. (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%]

Note: For the marginal MLE decomposition, the raw gap is estimated and not observed. Bootstrapping gives a set of gaps, explained and unexplained parts. Here we report the mean estimated gap and the mean percentages of the gap which are explained or not.
Source: Formation Qualification Professionnelle survey (FQP), Insee, Paris, 2003.

Table 7: Counterfactual approach: earnings differentials between French workers with French parents and French workers with at least one African parent

Sample	raw gap	Structural Gap $\hat{\Delta}_1$	C.I. $\hat{\Delta}_1$	Unexplained gap $\hat{\Delta}_2$	C.I. $\hat{\Delta}_2$
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]

Note: As explained in 2.3.5, this table does not show classical decompositions, and in particular, $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are not supposed to add up to the raw gap.
Source: Formation Qualification Professionnelle survey (FQP), Insee, Paris, 2003.

part is higher than for the wage gap, around 47%.[3]

All this tends to prove that there exists a strong difference in the employment probabilities of the two groups, which may be partly unexplained by usual covariates. Once workers are hired, there is still a wage gap between the two groups. This gap is slightly lower than the gap between the employment probabilities.

[3]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 unprecise 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, bootstrapping for the marginal MLE decomposition gives an explained part of the wage gap ranging 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.

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 a counterfactual group whose observable covariates are distributed as those of the potentially discriminated persons but whose slope coefficients are those of the reference group. Using this counterfactual group, we obtain estimates which 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 that we have to give up the usual 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. and that 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, Amadieu (2004)).

References

Aeberhardt, R., and J. Pouget (2007): "National Origin Wage Differentials in France: Evidence from Matched Employer-Employee Data," IZA Discussion Paper 2779.

Altonji, J., and R. Blank (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3C, pp. 3143–3259. Elsevier, Amsterdam.

Amadieu, J.-F. (2004): "Enquête Testing sur CV," Université Paris-I.

Arrow, K. J. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter, and A. Rees, pp. 3–33. Princeton University Press.

Becker, G. (1957): *The Economics of Discrimination*. University of Chicago Press.

Beffy, M., M. Buchinsky, D. Fougère, T. Kamionka, and F. Kramarz (2006): "The Returns to Seniority in France (and Why they are Lower than in the United States)," IZA Discussion Paper No 1935.

Bertrand, M., and S. Mullainathan (2003): "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 94(4), 159–190.

Blinder, A. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8, 436–455.

Fougère, D., and M. Safi (2005): "L'acquisition de la nationalité française : quels effets sur l'accès à l'emploi des immigrés ?," in *France Portrait Social, édition 2005-2006*, pp. 163–184. Insee, Paris.

Heckman, J. (1998): "Detecting Discrimination," *Journal of Economic Perspectives*, 12, 101–116.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

Meurs, D., A. Pailhé, and P. Simon (2005): "Mobilité intergénérationnelle et persistance des inégalités. L'accès à l'emploi des immigrés et de leurs descendants en France," INED Working Paper.

Neal, D., and W. Johnson (1996): "The Role of Premarket Factors in Black-White wages differences," *Journal of political Economy*, 104, 869–895.

Neuman, S., and R. Oaxaca (2004a): "Wage Decompositions with Selectivity Corrected Wage Equations: A Methodological Note," *Journal of Economic Inequality*, 2, 3–10.

——— (2004b): "Wage Differentials in the 1990s in Israel: Endowments, Discrimination, and selectivity," IZA Working Paper no. 1362.

Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709.

Oaxaca, R., and M. Ransom (1994): "On Discrimination and the decomposition of wage Differentials," *Journal of Econometrics*, 65, 5–21.

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659–661.

Pouget, J. (2005): "La Fonction publique : vers plus de diversité ?," in *France, Portrait Social, édition 2005-2006*, pp. 143–162. Insee, Paris.

Silberman, R., and I. Fournier (1999): "Les enfants d'immigrés sur le marché du travail, Les mécanismes d'une discrimination sélective," *Formation Emploi*, 65, 31–55.

Table 8: 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
Diploma		
Graduate	2 600	99
Some College	2 472	76
Completed High School	3 556	130
Vocational Degree	5 886	186
Junior High School	2 109	99
No Diploma	5 632	304
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
Residence		
Not Poor × Rest of France	18 233	417
Not Poor × Paris and Suburbs	2 956	270
Poor × Rest of France	788	135
Poor × Paris and Suburbs	278	72
Situation on the Labor Market		
12 months FT	13 246	429
12 months PT	2 208	54
12 months FT/PT	298	11
12 months unemployed	787	83
some work (various situations)	3 059	167
no work (various situations)	2 657	150

Note: The figures correspond to the exact number of observations in the sample.
Interpretation: Among French individuals whose both parents are French at birth, there are 12 054 women and 10 201 men in the sample.
Source: Formation Qualification Professionnelle survey (FQP), Insee, Paris, 2003.