

Aeberhardt, Romain; Fougère, Denis; Pouget, Julien; Rathelot, Roland

**Working Paper**

## Wages and employment of French workers with African origin

IZA Discussion Papers, No. 2898

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Aeberhardt, Romain; Fougère, Denis; Pouget, Julien; Rathelot, Roland (2007) : Wages and employment of French workers with African origin, IZA Discussion Papers, No. 2898, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<http://hdl.handle.net/10419/34591>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 2898

## Wages and Employment of French Workers with African Origin

Romain Aeberhardt  
Denis Fougère  
Julien Pouget  
Roland Rathelot

July 2007

# **Wages and Employment of French Workers with African Origin**

**Romain Aeberhardt**

*CREST-INSEE*

**Denis Fougère**

*CNRS, CREST-INSEE, CEPR and IZA*

**Julien Pouget**

*CREST-INSEE and IZA*

**Roland Rathelot**

*CREST-INSEE*

Discussion Paper No. 2898

July 2007

IZA

P.O. Box 7240

53072 Bonn

Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

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

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

## ABSTRACT

### **Wages and Employment of French Workers with African Origin<sup>\*</sup>**

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 whose at least one parent 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 and one third of the wage gap is not explained by differences in observable covariates between the two groups. This result is obtained by using a new method yielding more precise results when the sample size of the potentially discriminated group is small.

JEL Classification: C24, J31, J71

Keywords: discrimination, wage differentials, second-generation migrants

Corresponding author:

Roland Rathelot  
INSEE DESE  
Timbre G220  
15, Boulevard Gabriel Péri  
92245 Malakoff Cedex  
France  
E-mail: [roland.rathelot@insee.fr](mailto:roland.rathelot@insee.fr)

---

<sup>\*</sup> The authors thank participants at the Alliance Conference on Group Inequality and Discrimination (Columbia University, May 2005), at the Séminaire Recherche de l'INSEE (Paris, June 2005), and at the JMA Conference (Nantes, June 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.

# 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% within 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.<sup>1</sup> 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 a counterfactual group whose observable covariates are distributed as those of the potentially

---

<sup>1</sup>The program called "Zones Urbaines Sensibles" (ZUS) was launched in 1995; it concerns 751 disadvantaged zones that receive public extra resources and that benefit from tax exemptions. In these zones, the unemployment rate is very high (25.4% in 1999, 39.5% for workers aged between 15 and 24); the proportion of migrants is also very high (16.5% in 1999, vs. 5.6% in France)

discriminated group. This method proves to give more precise estimates than the usual ones.

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.

We denote  $\ln(w_{ij})$  the log-wage of the individual  $i$  in (demographic) group  $j$ . Individuals belonging to group  $j$  are assumed to be potentially discriminated. We suppose that the wage is generated by the following model:

$$\ln(w_{ij}) = X'_{ij}\beta_j + u_{ij} \quad (1)$$

where the residual  $u_{ij}$  is assumed to be zero-mean, homoskedastic with variance  $\mathbb{V}(u_{ij}) = \sigma_{u,j}^2$ . Another group of individuals, labeled  $k$ , is the reference group. The log-wages of workers belonging to this group are modeled as:

$$\ln(w_{ik}) = X'_{ik}\beta_k + u_{ik} \quad (2)$$

where the residual  $u_{ik}$  has the same properties as  $u_{ij}$ . All observations from the same group ( $j$  or  $k$ ) are assumed independent and identically distributed.

The difference between expected log-wages can be decomposed as:

$$\mathbb{E}[\ln(w_{ij})] - \mathbb{E}[\ln(w_{ik})] = \mathbb{E}[X'_{ij}](\beta_j - \beta_k) + (\mathbb{E}[X'_{ij}] - \mathbb{E}[X'_{ik}])\beta_k \quad (3)$$

The first term of this sum can be interpreted as the wage gap due to discrimination. The second part of the wage gap is due to the average gap in individual characteristics between the two groups.

OLS estimations of parameters  $\beta_j$  and  $\beta_k$  are potentially biased, since only the participants' wages are observed and taken into account in the regression. To get unbiased estimates, one can specify a two-equation model, with a selection equation and a wage equation:

$$\text{for group } j: \begin{cases} Y_{ij}^* = Z'_{ij}\gamma_j + \varepsilon_{ij} \\ \ln(w_{ij}) = X'_{ij}\beta_j + u_{ij} \end{cases} \quad (4)$$

$$\text{for group } k: \begin{cases} Y_{ik}^* = Z'_{ik}\gamma_k + \varepsilon_{ik} \\ \ln(w_{ik}) = X'_{ik}\beta_k + u_{ik} \end{cases} \quad (5)$$

The first equation of the system 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). In other words, worker  $i$  is employed if and only if  $Y_{ij} = 1$  (*i.e.*  $Y_{ij}^* > 0$ ). She is not employed if and only if  $Y_{ij} = 0$  (*i.e.*  $Y_{ij}^* < 0$ ). A similar model is applied to individuals in group  $k$ . We add an assumption on the joint distribution of residuals  $u$  and  $\varepsilon$ . Vectors  $(\varepsilon_{ij}, u_{ij})$  and  $(\varepsilon_{ik}, u_{ik})$  are assumed to be generated by a bivariate normal distribution,

$$\begin{pmatrix} \varepsilon_{ij} \\ u_{ij} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho_j \sigma_{u,j} \\ \rho_j \sigma_{u,j} & \sigma_{u,j}^2 \end{bmatrix} \right) \quad (6)$$

and

$$\begin{pmatrix} \varepsilon_{ik} \\ u_{ik} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho_k \sigma_{u,k} \\ \rho_k \sigma_{u,k} & \sigma_{u,k}^2 \end{bmatrix} \right) \quad (7)$$

Under this set of assumptions, the difference between expected log-wages of employed workers in the two groups can be written as:

$$\begin{aligned} \mathbb{E}[\ln(w_{ij}) | Y_{ij} = 1] - \mathbb{E}[\ln(w_{ik}) | Y_{ik} = 1] &= \mathbb{E}[X'_{ij}](\beta_j - \beta_k) + (\mathbb{E}[X'_{ij}] - \mathbb{E}[X'_{ik}]) \beta_k \quad (8) \\ &+ (\rho_j \sigma_{u,j} \mathbb{E}[\lambda_j] - \rho_k \sigma_{u,k} \mathbb{E}[\lambda_k]) \end{aligned}$$

In this expression, terms  $\lambda_j$  and  $\lambda_k$  are the inverse Mills' ratios, defined as:

$$\lambda_j = \frac{\varphi(Z'_{ij}\gamma_j)}{\Phi(Z'_{ij}\gamma_j)} \quad \text{and} \quad \lambda_k = \frac{\varphi(Z'_{ik}\gamma_k)}{\Phi(Z'_{ik}\gamma_k)} \quad (9)$$

Just as before, the second term in decomposition (8) can be understood as the part of the wage gap explained by average differences in observable characteristics, while the first term can be attributed to “discrimination”. The last term in this expression is attributed to the difference in selectivity terms between the two groups. Parameters of models (4)+(6) and (5)+(7) can be estimated either by a maximum likelihood procedure, or by a two-step consistent method.

This Tobit approach includes selectivity terms which make the results more difficult to interpret. Neuman and Oaxaca (2004a) and Neuman and Oaxaca (2004b) try to deal with these selectivity terms and to interpret them. They incorporate subcomponents of the Mills' ratios into the explained part and into the discrimination element so that some or all of the selectivity terms vanish. This approach relies on specific choices we do not want to make.

To avoid these drawbacks, we decompose the difference between the unconditional expected log-wages, namely  $\mathbb{E}[\ln(w_{ij})] - \mathbb{E}[\ln(w_{ik})]$ . Estimating the  $\beta$ 's using a Tobit model provides



consistent estimates of the unconditional expectation of this log-wage difference:

$$\mathbb{E}[\ln(w_{ij})] - \mathbb{E}[\ln(w_{ik})] = \mathbb{E}[X'_{ij}](\beta_j - \beta_k) + (\mathbb{E}[X'_{ij}] - \mathbb{E}[X'_{ik}]) \beta_k \quad (10)$$

A practical problem with this 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 estimated their expected values (given the values of other observed covariates) from a regression model (either an OLS or a probit model, depending on the qualitative or quantitative nature of the missing covariate) estimated within the group of employed workers. Then we have checked that our results are not significantly affected by this imputation method.

A second drawback of the Oaxaca-Blinder method occurs when the sample size of one group is too small to provide precise estimates. In our case, the sample size of the potentially discriminated group  $j$  is actually quite small. Therefore we derive an estimator for the discrimination component of the wage gap that is not dependent on  $\beta_j$ . For this purpose we consider a counterfactual population  $l$ , which has exactly the same distribution of observable covariates as population  $j$  (the potentially discriminated group), and whose employment and wage are generated by the following model:

$$\begin{cases} Y_{il}^* = Z'_{il}\gamma_k + \varepsilon_{il} \\ \ln(w_{il}) = X'_{il}\beta_k + u_{il} \end{cases} \quad (11)$$

This means that the market returns to the characteristics of population  $l$  are the same as those of group  $k$  (the reference group). From this definition,  $\mathbb{E}[\ln(w_{il}) | Y_{il} = 1]$  is the counterfactual that we want to identify, namely the expected wage of a worker with covariates of type  $j$  (the minority group) and whose returns are those of an individual of the reference group  $k$ . Consequently, decomposing the conditional wage gap using this counterfactual results in the following wage equation:

$$\begin{aligned} \mathbb{E}[\ln(w_{ik}) | Y_{ik} = 1] - \mathbb{E}[\ln(w_{ij}) | Y_{ij} = 1] &= \underbrace{\mathbb{E}[\ln(w_{ik}) | Y_{ik} = 1] - \mathbb{E}[\ln(w_{il}) | Y_{il} = 1]}_{\text{explained component}} \\ &+ \underbrace{\mathbb{E}[\ln(w_{il}) | Y_{il} = 1] - \mathbb{E}[\ln(w_{ij}) | Y_{ij} = 1]}_{\text{unexplained component}} \end{aligned}$$

For a given individual, we have:

$$\mathbb{E}[\ln(w_{il}) | X_{il}, Z_{il}, Y_{il} = 1] = X_{il}\beta_k + \rho_k \sigma_{u,k} \frac{\varphi(Z_{il}\gamma_k)}{\Phi(Z_{il}\gamma_k)}$$

Here we propose to estimate a counterfactual mean wage, representing the observed mean wage among the individuals who would actually be working if their wages were generated by the same model as that of population  $k$ . Let us assume that we observe completely the  $X_{ij}$ ,  $X_{ik}$ ,  $Z_{ij}$ , and

$Z_{ik}$  covariates, and that we estimate consistently  $\beta_k, \gamma_k, \sigma_{u,k}, \rho_k$  by  $\hat{\beta}_k, \hat{\gamma}_k, \hat{\sigma}_{u,k}, \hat{\rho}_k$ . Then an estimator  $\hat{w}_l$  for this counterfactual mean  $\hat{w}_l$  is:

$$\hat{w}_l = \sum_{i \in j} \left( \frac{\Phi(Z_{ij}\hat{\gamma}_k)}{\sum_i \Phi(Z_{ij}\hat{\gamma}_k)} \right) \left[ X_{ij}\hat{\beta}_k + \hat{\rho}_k \hat{\sigma}_{u,k} \frac{\varphi(Z_{ij}\hat{\gamma}_k)}{\Phi(Z_{ij}\hat{\gamma}_k)} \right] \quad (12)$$

This estimator measures the overall average wage of employed individuals. It is computed among the whole population and each element of the sum is weighted by the probability of being employed.

In order to justify this expression, we need to prove that, for the reference population, this weighted average on the whole population corresponds to the average of the observed wages (calculated only for the individuals whose wage is observed). In other words, we need to show that:

$$\frac{\sum_{i=1}^N \ln(w_i) \mathbb{I}_{\{Y_i=1\}}}{\sum_{i=1}^N Y_i} \stackrel{N \rightarrow \infty}{\sim} \frac{1}{\sum_{i=1}^N \Phi(Z'_i \gamma)} \sum_{i=1}^N \Phi(Z'_i \gamma) \left[ X'_i \beta + \rho \sigma_u \frac{\varphi(Z'_i \gamma)}{\Phi(Z'_i \hat{\gamma})} \right]$$

where the first term corresponds to the mean of the observed wages calculated over the subsample of working individuals, and the second term is the expression that we use to estimate this expectation on any other subpopulation (including working and non working individuals). In this expression,  $g \stackrel{N \rightarrow \infty}{\sim} f$  means  $f - g = o(g)$  when  $N \rightarrow \infty$ .

Indeed, because  $\mathbb{E}(Y_i) = \Phi(Z'_i \gamma)$ , the Lindberg-Feller central limit theorem with unequal variances implies that:

$$\frac{1}{N} \sum_{i=1}^N Y_i \stackrel{N \rightarrow \infty}{\sim} \frac{1}{N} \sum_{i=1}^N \Phi(Z'_i \gamma)$$

Similarly,

$$\begin{aligned} \mathbb{E}(\ln(w_i) \mathbb{I}_{\{Y_i=1\}}) &= Pr(Y_i = 1) \mathbb{E}(\ln(w_i) \mathbb{I}_{\{Y_i=1\}} | Y_i = 1) + \underbrace{Pr(Y_i = 0) \mathbb{E}(\ln(w_i) \mathbb{I}_{\{Y_i=1\}} | Y_i = 0)}_{=0} \\ &= Pr(Y_i = 1) \mathbb{E}(\ln(w_i) | Y_i = 1) \\ &= \Phi(Z'_i \gamma) \left[ X'_i \beta + \rho \sigma_u \frac{\varphi(Z'_i \gamma)}{\Phi(Z'_i \hat{\gamma})} \right] \end{aligned}$$

which gives:

$$\frac{1}{N} \sum_{i=1}^N \ln(w_i) \mathbb{I}_{\{Y_i=1\}} \stackrel{N \rightarrow \infty}{\sim} \frac{1}{N} \sum_{i=1}^N \Phi(Z'_i \gamma) \left[ X'_i \beta + \rho \sigma_u \frac{\varphi(Z'_i \gamma)}{\Phi(Z'_i \hat{\gamma})} \right]$$

and finally, due to the above result, we get:

$$\frac{\sum_{i=1}^N \ln(w_i) \mathbb{I}_{\{Y_i=1\}}}{\sum_{i=1}^N Y_i} \stackrel{N \rightarrow \infty}{\sim} \frac{1}{\sum_{i=1}^N \Phi(Z'_i \gamma)} \sum_{i=1}^N \Phi(Z'_i \gamma) \left[ X'_i \beta + \rho \sigma_u \frac{\varphi(Z'_i \gamma)}{\Phi(Z'_i \hat{\gamma})} \right]$$

### 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 across 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 than 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 between 18 and 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 wage earners and 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 has worked in 2002 and has earned some 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 with respect to this calendar 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 have 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 have distinguished 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 origins, 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 origins. 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 origins are much more concentrated in the Paris region and in ZUS areas.

Table 7 in the Appendix presents sample sizes in more detail.

Table 1: Descriptive statistics

		%
National origin of parents	France	Africa
Number of individuals	22 255	796
Gender		
<i>Female</i>	54.2	54.1
<i>Male</i>	45.8	45.9
Age		
<i>Less than 20</i>	0.5	1.4
<i>20 to 29</i>	15.4	24.6
<i>30 to 39</i>	29.4	35.6
<i>40 to 49</i>	27.0	25.5
<i>50 to 59</i>	24.0	10.8
<i>60 and more</i>	3.7	2.1
Diploma		
<i>Graduate</i>	11.7	10.6
<i>Some college</i>	11.1	9.0
<i>Completed high-school</i>	16.0	13.4
<i>Vocational high-school</i>	26.4	21.7
<i>Junior high-school</i>	9.5	11.4
<i>No diploma</i>	25.3	33.8
Household		
<i>Single man without children</i>	7.1	6.3
<i>Single man with children</i>	2.1	3.4
<i>Single woman without children</i>	7.1	6.8
<i>Single woman with children</i>	5.5	9.2
<i>Man with a working spouse, with children</i>	17.0	11.4
<i>Man with a working spouse, without children</i>	7.3	5.2
<i>Man with a non-working spouse, with children</i>	8.9	18.2
<i>Man with a non-working spouse, without children</i>	3.5	1.4
<i>Woman with a working spouse, with children</i>	23.2	22.5
<i>Woman with a working spouse, without children</i>	9.0	4.6
<i>Woman with a non-working spouse, with children</i>	4.2	8.9
<i>Woman with non working spouse without children</i>	5.2	2.1
Residence		
<i>Not in a ZUS, not in the Paris region</i>	81.9	47.1
<i>Not in a ZUS, but in the Paris region</i>	13.3	29.1
<i>In a ZUS, but not in the Paris region</i>	3.5	15.1
<i>In a ZUS and in the Paris region</i>	1.2	8.7
Situation in the Labor Market		
<i>Employed 12 months full-time</i>	59.5	47.7
<i>Employed 12 months part-time</i>	9.9	6.0
<i>Employed 12 months full- and part-time</i>	1.3	1.1
<i>12 months unemployed</i>	3.5	9.3
<i>Some employment (various situations)</i>	13.7	19.6
<i>No employment (various situations)</i>	11.9	16.2

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 parents	France		Africa	
	non-employed	employed	non-employed	employed
Number of individuals	3 988	18 267	239	557
Gender				
<i>Female</i>	77.0	49.2	72.8	46.1
<i>Male</i>	23.0	50.8	27.2	53.9
Age				
<i>Less than 20</i>	1.6	0.3	3.8	0.4
<i>20 to 29</i>	14.4	15.6	30.1	22.3
<i>30 to 39</i>	21.9	31.0	36.0	35.4
<i>40 to 49</i>	19.4	28.6	16.7	29.3
<i>50 to 59</i>	31.6	22.4	10.5	11.0
<i>60 and more</i>	11.1	2.1	2.9	1.8
Diploma				
<i>Graduate</i>	6.7	12.8	4.6	13.1
<i>Some college</i>	5.6	12.3	6.3	10.2
<i>Completed high-school</i>	11.9	16.9	10.5	14.7
<i>Vocational high-school</i>	23.1	27.2	18.4	23.2
<i>Junior high-school</i>	10.3	9.3	13.0	10.8
<i>No diploma</i>	42.5	21.6	47.3	28.0
Household				
<i>Single man without children</i>	5.9	7.3	5.9	6.5
<i>Single man with children</i>	2.4	2.0	4.2	3.1
<i>Single woman without children</i>	6.8	7.1	4.2	7.9
<i>Single woman with children</i>	7.9	4.9	13.8	7.2
<i>Man with a working spouse, with children</i>	3.3	20.0	3.3	14.9
<i>Man with a working spouse, without children</i>	2.9	8.3	1.3	6.8
<i>Man with a non-working spouse, with children</i>	5.5	9.6	11.7	21.0
<i>Man with a non-working spouse, without children</i>	3.0	3.6	0.8	1.6
<i>Woman with a working spouse, with children</i>	30.2	21.7	31.0	18.9
<i>Woman with a working spouse, without children</i>	9.3	8.9	4.6	4.7
<i>Woman with a non-working spouse, with children</i>	8.8	3.2	15.9	5.9
<i>Woman with a non-working spouse, without children</i>	14.1	3.3	3.3	1.6
Residence				
<i>Not in a ZUS, not in the Paris region</i>	83.4	81.6	50.2	45.8
<i>Not in a ZUS, but in the Paris region</i>	9.4	14.1	20.9	32.7
<i>In a ZUS, but not in the Paris region</i>	5.9	3.0	20.5	12.7
<i>In a ZUS, and in the Paris region</i>	1.3	1.2	8.4	8.8

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 parents	France		Africa	
	non-ZUS	ZUS	non-ZUS	ZUS
Number of individuals	17489	778	437	120
Working time (%)				
<i>Part-time</i>	17.1	16.6	14.6	16.1
<i>Full-time</i>	82.9	83.4	85.4	83.9
Professional category (%)				
<i>Craftsman</i>	2.6	1.3	3.0	0.0
<i>Executive</i>	15.8	9.4	15.9	1.1
<i>Intermediate</i>	27.6	26.9	20.7	19.4
<i>White-collar</i>	30.8	36.3	32.8	44.1
<i>Skilled blue-collar</i>	16.4	17.7	18.9	21.5
<i>Unskilled blue-collar</i>	6.8	8.4	8.6	14.0
Earnings (euros)				
<i>Mean</i>	18 636	15 735	16 379	11 847
<i>First quarter</i>	11 639	10 976	10 000	7 000
<i>Median</i>	16 189	14 700	14 156	11 500
<i>Third quarter</i>	22 867	19 967	19 970	15 245
Wage (full-time full-year equivalent, euros)				
<i>Mean</i>	21 526	18 248	18 763	14 776
<i>First quarter</i>	13 150	12 522	12 000	9 512
<i>Median</i>	17 544	15 688	14 940	12 958
<i>Third quarter</i>	24 080	21 000	21 747	17 658
Employment status (%)				
<i>Employed 12 months full-time</i>	59.9	52.3	50.2	39.7
<i>Employed 12 months part-time</i>	10.0	8.0	6.1	5.8
<i>Employed 12 months full- and part-time</i>	1.3	1.5	1.0	1.6
<i>12 months unemployed</i>	3.3	7.3	8.2	12.7
<i>Some employment (various situations)</i>	13.7	14.8	19.3	20.6
<i>No employment (various situations)</i>	11.7	16.0	15.2	19.6

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) 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 having gained 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 general, coefficients are not statistically different between the two groups. 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. 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 Paris region (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 Paris region and the reference category. The situation is different for individuals with African parents. If they live in the Paris region, but outside of a ZUS, their wage is significantly higher. It is lower if they live in a ZUS that is located outside of the Paris region. 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 Paris region (although the estimated parameter is only significant at the 90% level). Those living outside of the Paris region 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 Befy, 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. Part-time workers have lower wages than full-time workers, which is consistent with the definition we take for the wage. 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 (which may reflect that part-time workers with African parents work less hours than part-time workers with French parents) 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. The first line (labelled OLS) of this table 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.

Lines 2 and 3 of Table 6 refer to the marginal Blinder-Oaxaca decomposition that requires to estimate the wage and employment equations within each group separately. Results obtained through H2S and MLE procedures are similar. They contrast from the OLS estimates in that the explained part grows up to almost 75%. Thus the unexplained part is limited to approximately one quarter of the total wage gap. Line 4 of Table 6 refers to our decomposition method which is only based on the estimation of the wage and employment equations for the reference group (see the line “counterfactual MLE”). This decomposition yields more precise estimates,

Table 4: Estimates of the employment equation parameters

Covariates	French parents		African parents	
	H2S	MLE	H2S	MLE
Intercept	0.657*** (0.046)	0.656*** (0.046)	0.444* (0.239)	0.444* (0.239)
Household composition				
<i>Single men and single women w/o children</i>	Ref.	Ref.	Ref.	Ref.
<i>Single women with children</i>	-0.436*** (0.050)	-0.435*** (0.050)	-0.550*** (0.203)	-0.549*** (0.203)
<i>Men with a working spouse, with children</i>	0.545*** (0.050)	0.545*** (0.050)	0.483** (0.244)	0.486** (0.245)
<i>Men with a working spouse, without children</i>	0.502*** (0.056)	0.503*** (0.056)	0.915*** (0.345)	0.914*** (0.345)
<i>Men with a non-working spouse, with children</i>	0.241*** (0.049)	0.240*** (0.049)	0.298* (0.180)	0.297* (0.180)
<i>Men with a non-working spouse, without children</i>	0.153** (0.064)	0.158** (0.064)	0.564 (0.506)	0.564 (0.506)
<i>Women with a working spouse, with children</i>	-0.526*** (0.035)	-0.526*** (0.035)	-0.499*** (0.163)	-0.498*** (0.163)
<i>Women with a working spouse, without children</i>	-0.054 (0.044)	-0.053 (0.044)	-0.109 (0.259)	-0.108 (0.259)
<i>Women with a non-working spouse, with children</i>	-0.495*** (0.053)	-0.494*** (0.053)	-0.664*** (0.206)	-0.668*** (0.209)
<i>Women with a non-working spouse, without children</i>	-0.570*** (0.049)	-0.570*** (0.049)	-0.435 (0.349)	-0.433 (0.349)
Residence location				
<i>Not in a ZUS, not in the Paris region</i>	Ref.	Ref.	Ref.	Ref.
<i>Not in a ZUS, but in the Paris region</i>	0.121*** (0.035)	0.120*** (0.035)	0.282** (0.127)	0.282** (0.127)
<i>In a ZUS, but not in the Paris region</i>	-0.238*** (0.054)	-0.238*** (0.054)	0.102 (0.153)	0.102 (0.153)
<i>In a ZUS, in the Paris region</i>	-0.044 (0.095)	-0.044 (0.095)	0.346* (0.194)	0.347* (0.194)
Experience	0.122*** (0.003)	0.122*** (0.003)	0.101*** (0.014)	0.101*** (0.014)
Experience squared	-0.276*** (0.008)	-0.276*** (0.008)	-0.222*** (0.039)	-0.222*** (0.039)
Diploma				
<i>University graduate</i>	Ref.	Ref.	Ref.	Ref.
<i>College</i>	0.056 (0.053)	0.057 (0.053)	-0.061 (0.262)	-0.061 (0.262)
<i>Completed high-school</i>	-0.181*** (0.046)	-0.180*** (0.046)	-0.227 (0.237)	-0.227 (0.237)
<i>Vocational high-school</i>	-0.448*** (0.043)	-0.446*** (0.044)	-0.395* (0.222)	-0.394* (0.222)
<i>Junior high-school</i>	-0.519*** (0.051)	-0.518*** (0.051)	-0.635*** (0.240)	-0.634*** (0.240)
<i>No diploma</i>	-0.738*** (0.043)	-0.737*** (0.043)	-0.811*** (0.209)	-0.809*** (0.210)
Nobs	22321	22321	796	796

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	9.958*** (0.021)	9.980*** (0.027)	9.962*** (0.022)	9.976*** (0.114)	9.998*** (0.146)	9.982*** (0.122)
Working time						
<i>Part-time</i>	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<i>Full-time</i>	-0.065*** (0.011)	-0.067*** (0.011)	-0.065*** (0.011)	-0.038 (0.069)	-0.040 (0.068)	-0.039 (0.068)
Gender						
<i>Men</i>	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<i>Women</i>	-0.253*** (0.008)	-0.242*** (0.012)	-0.251*** (0.009)	-0.177*** (0.050)	-0.164** (0.078)	-0.174*** (0.058)
Residence location						
<i>Not in a ZUS, not in the Paris region</i>	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<i>Not in a ZUS, but in the Paris region</i>	0.220*** (0.011)	0.219*** (0.011)	0.220*** (0.011)	0.191*** (0.055)	0.186*** (0.058)	0.190*** (0.055)
<i>In a ZUS, but not in the Paris region</i>	-0.046** (0.022)	-0.042* (0.023)	-0.045** (0.023)	-0.199*** (0.076)	-0.199*** (0.075)	-0.199*** (0.075)
<i>In a ZUS and in the Paris region</i>	0.105*** (0.035)	0.106*** (0.035)	0.105*** (0.035)	0.076 (0.088)	0.071 (0.090)	0.074 (0.087)
Experience	0.024*** (0.001)	0.022*** (0.002)	0.024*** (0.002)	0.012* (0.007)	0.011 (0.011)	0.012 (0.008)
Experience squared	-0.038*** (0.003)	-0.033*** (0.005)	-0.037*** (0.003)	-0.022 (0.018)	-0.018 (0.024)	-0.021 (0.020)
Seniority						
<i>Less than 1 year</i>	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<i>1 to 5 years</i>	0.062*** (0.015)	0.062*** (0.015)	0.062*** (0.015)	0.061 (0.069)	0.061 (0.068)	0.061 (0.068)
<i>5 to 10 years</i>	0.112*** (0.016)	0.112*** (0.016)	0.112*** (0.016)	0.102 (0.085)	0.101 (0.084)	0.102 (0.084)
<i>More than 10 years</i>	0.246*** (0.016)	0.247*** (0.016)	0.246*** (0.016)	0.208** (0.085)	0.208** (0.084)	0.208** (0.084)
Diploma						
<i>University graduate</i>	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
<i>College</i>	-0.190*** (0.015)	-0.191*** (0.015)	-0.190*** (0.015)	-0.345*** (0.100)	-0.345*** (0.099)	-0.345*** (0.099)
<i>Completed high-school</i>	-0.377*** (0.014)	-0.374*** (0.014)	-0.376*** (0.014)	-0.562*** (0.090)	-0.559*** (0.090)	-0.561*** (0.089)
<i>Vocational high-school</i>	-0.601*** (0.014)	-0.594*** (0.015)	-0.600*** (0.014)	-0.536*** (0.083)	-0.529*** (0.086)	-0.534*** (0.083)
<i>Junior high-school</i>	-0.506*** (0.017)	-0.498*** (0.018)	-0.505*** (0.017)	-0.541*** (0.098)	-0.529*** (0.109)	-0.538*** (0.100)
<i>No diploma</i>	-0.748*** (0.015)	-0.734*** (0.018)	-0.745*** (0.015)	-0.598*** (0.081)	-0.583*** (0.105)	-0.595*** (0.087)
Nobs	22 321	22 321	22 321	796	796	796

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.

since the reference group is large enough. The results show that the explained part is almost equal to two thirds, somewhere between the two other estimates. The last line refers to the same decomposition but it focuses on the “second generation” only. The “second generation” includes persons who have at least one African parent but who have gained the French citizenship at birth. This subpopulation is too small to be analysed with the other methods, but our method may still be applied since it does not involve any estimation within the minority group.

Bootstrapping within the estimated distribution of the estimated coefficients, we are able to compute 95% confidence intervals: such an interval goes from 52% to 69% for the full sample, and from 76% to 89% when we focus on the “second generation” only. Our results all converge to the fact that unobserved factors represent between about one quarter and one half of the wage gap.

Concerning the employment probability, all the decompositions suggest that the unexplained part is even higher, around 47%.<sup>2</sup>

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

---

<sup>2</sup>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 42% to 240%, which means that we are pretty confident to be able to explain at least 42% of the wage gap and maybe all of it.

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	Wage gap decomposition			Employment gap decomposition		
	raw gap	explained	unexplained	raw gap	explained	unexplained
<i>Full sample</i>						
OLS	0.145	0.074	0.071	-	-	-
H2S	0.149	0.110	0.039	0.121	0.064	0.057
MLE	0.153	0.113	0.040	0.121	0.064	0.057
Counterfactual MLE	0.145	0.088	0.057	0.121	0.065	0.056
<i>Second generation only</i>						
Counterfactual MLE	0.208	0.171	0.037	0.150	0.087	0.063

Notes: The first line refers to the traditional OLS Oaxaca-Blinder decomposition. The second and the third lines refer to the marginal decomposition of the Tobit model, as described by equation (10). The fourth line refers to our “counterfactual” decomposition, where the counterfactual is estimated following equation (12). The last line refers to the same decomposition but focuses on the “second generation” only.

Source: Formation Qualification Professionnelle survey (FQP), INSEE, Paris, 2003.

## 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 whose at least one parent gained at birth the nationality of an African country. 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 that prove to be more precise than usual estimates, even if the sample size of the potentially discriminated group is rather small. Our estimates suggest that one half of the employment gap and one third of the wage gap is not explained by differences in usual covariates, such as age, gender, education, potential experience, residential area, etc. 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, Amadiou (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. (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.
- 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. (1972): “The Statistical Theory of Racism and Sexism,” *American Economic Review*, 62, 639–651.
- 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 7: Number of observations (full sample)

National origin of parents	France	Africa
Total number of individuals	18 267	557
Gender		
<i>Female</i>	12 054	431
<i>Male</i>	10 201	365
Age		
<i>Less than 20</i>	113	11
<i>20 to 29</i>	3 422	196
<i>30 to 39</i>	6 542	283
<i>40 to 49</i>	6 001	203
<i>50 to 59</i>	5 346	86
<i>60 and more</i>	831	17
Diploma		
<i>Graduate</i>	2 600	84
<i>Some college</i>	2 472	72
<i>Completed high-school</i>	3 556	107
<i>Vocational high-school</i>	5 886	173
<i>Junior high-school</i>	2 109	91
<i>No diploma</i>	5 632	269
Household		
<i>Single man without children</i>	1 576	50
<i>Single man with children</i>	457	27
<i>Single woman without children</i>	1 578	54
<i>Single woman with children</i>	1 214	73
<i>Man with a working spouse, with children</i>	3 781	91
<i>Man with a working spouse, without children</i>	1 633	41
<i>Man with a non-working spouse, with children</i>	1 976	145
<i>Man with a non-working spouse, without children</i>	778	11
<i>Woman with a working spouse, with children</i>	5 174	179
<i>Woman with a working spouse, without children</i>	1 999	37
<i>Woman with a non-working spouse, with children</i>	932	71
<i>Woman with a non-working spouse, without children</i>	1 157	17
Residence		
<i>Not in a ZUS, not in the Paris region</i>	18 233	375
<i>Not in a ZUS, but in the Paris region</i>	2 956	232
<i>In a ZUS, not in the Paris region</i>	788	120
<i>In a ZUS, but in the Paris region</i>	278	69
Labor market status		
<i>Employed 12 months full-time</i>	13 246	380
<i>Employed 12 months part-time</i>	2 208	48
<i>Employed 12 months full- and part-time</i>	298	9
<i>12 months unemployed</i>	787	74
<i>Some employment (various situations)</i>	3 059	156
<i>No employment (various situations)</i>	2 657	129

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.