WORKING PAPER

N° 2020-7

# MEASURING DISCRIMINATION IN THE LABOUR MARKET

Emmanuel Duguet

www.tepp.eu

TEPP – Theory and Evaluation of Public Policies - FR CNRS 2042

# Measuring Discrimination in the Labour Market

Emmanuel Duguet[1,2]

[1]Université Paris Est Créteil, ERUDITE EA 437
[2]FR TEPP CNRS 2042

August 2020

**Abstract**

This paper presents a survey of the estimation methods available to measure discrimination in the labour market. We review the most widespread methodology that the profession uses in order to evaluate discrimination. Discrimination may occur at three main stages in the labour market: hiring, wage setting and promotion. Most empirical studies deal with the two first discrimination types. In a first section, we will present the evaluation in hiring discrimination and, in a second section, the evaluation of wage discrimination.

**JEL**: : C18, C51, C93, J23, J31, J71.
**Keywords**: labour market, discrimination, hiring, wage.

# Introduction

Following Heckman (1998), a situation of discrimination against a group is said to arise if an otherwise identical person is treated differently by virtue of that person's group membership, and this group membership by himself has no direct effect on productivity. One important point is that "discrimination is a causal effect defined by a hypothetical *ceteris paribus* conceptual experiment". In other words, discrimination is defined as an injustice. It should not be confused with inequality, even though the two notions are related. The early theoretical analysis of the economics of discrimination dates back to Becker (Becker (1957)). The scope of this chapter is more modest. We will simply review the most widespread methodology that the profession uses in order to evaluate discrimination. Discrimination may occur at three main stages in the labour market: hiring, wage setting and promotion. Most empirical studies deal with the two first discrimination types. In a first section, we will present the evaluation in hiring discrimination and, in a second section, the evaluation of wage discrimination.

# 1 The evaluation of hiring discrimination

Hiring discrimination has considerably attracted attention in Europe since the begining of the 2000's. Baert (2017) surveys 90 studies conducted in Europe alone between 2005 and 2017, and this number is increasing steadily.[1] Several surveys are available, for Europe and other countries, including Bertrand and Duflo (2017), Neumark (2018) and Rich (2014). In this section, we will present the main tools that have been used in this literature.

The measurement of hiring discrimination needs the collection of specific data sets. Standard data bases do not allow for measuring hiring discrimination in a satisfactory manner because of the following problems. First, firm-level data sets include information about the workers that have been hired only. There is no information about the workers whose application has been rejected. Second, there is a self-selection problem. Workers that feel discriminated will tend not to apply to the jobs for which they think they have no chance to be recruited. Third, the opinion of the workers about whether they have been discriminated or not cannot be fully trusted because they generally have no information about their competitors for the job, and they are both judge and jury. Fourth, the information from the recruiters cannot be fully trusted for the same reason, and because discrimination is illegal and they have no incentive to reveal it. Last, two applications are never identical, so that we cannot know for sure whether a candidate was rejected for discrimination or for an objective, non observable, difference in the resume.

In order to answer these five critics, the researchers perform correspondence tests. A correspondence test is an ground experiment. The researcher replies to the ads instead of real candidates. If we wish to test gender discrimination, we should send two resumes, one for each gender, with comparable productive characteristics. This fixes the five previous problems: first, the data is not limited to the candidates that have been successful but also includes all the applicants that failed to reach an interview. Second, there is no self-selection because the applications are sent by the researcher. Third, we observe an objective answer from the firm, not an opinion from the candidate. Fourth, we observe the true behaviour of the recruiter. Fifth, the applications have been designed to be equivalent.

[1] There have been additional studies in Europe, but all of them have not been published in English.

In a correspondence test, we do not send candidates to the interview in order to avoid personality biases (Riach and Rich, 2002). When called, the candidate replies that s-he has already found a job. This is the "callback" variable. Here, it is possible to examine in which order the candidates are called. Some workers may be called in priority, but other workers may still be called when the preferred workers are not available.

Several problems have to be fixed. The candidates must be credible, a sufficient number of observations may be collected and some validity test may be provided. Once the data are collected, an adapted statistical analysis should be performed. Experimental data have specificities that should be accounted for. Among them, the researcher sets the values of the candidates' variables .

We will first recall the simplest tests, then we will show how they can be related to least squares method. Last, we present more advanced methods which can be used to reveal the presence of discriminatory components inside the callback rates. The main point for all theses method is that the answers to all the candidates on a same job add are correlated because the recruiter replies to all of them at the same time. This affects the way to compute the standard errors. We indicate how the statistics should be adapted.

## 1.1 The correspondence test

Several fake candidates are sent in answer to a job advertisement. The data is collected daily until the end of the experiment, which lasts several months in general. Many precaution must be taken in order to avoid detection. Among them:

- time consistency: the diplomas and the information about tenure must be consistent with the age of the candidates;
- the resumes must be similar but not identical, and examined by a professional associated with the correspondence test;
- the postal addresses must exhibit similar transportation times and neighbourhood reputation;
- other causes of discrimination must be avoided. If we test the existence of gender discrimination, we should send candidates from the same origin in order to avoid that origin discrimination weakens the identification of gender discrimination;
- resume templates may be rotated among the candidates;
- no photo should be used, since it can influence the answer (Rich, 2018);
- letters must be posted from different post offices;
- emails should be sent from different or undetectable servers;
- each application is followed by one person only;
- the testers must be tested;
- the confidentiality must be complete until all the data are collected.

## 1.2 Comparisons of proportions

Let us consider the two-candidate test. We have performed a correspondence test and have collected data about the callback of two candidates denoted $j \in \mathbf{J} = \{A, B\}$. For each job ad, we have two answers from the recruiter. The candidates can receive two answers only:[2] yes (coded 1) or no (coded 0). We summarize the situation in Table 1:

Table 1: Correspondence test

| $j = B$ / $j = A$ | yes (1) | no (0) |
|---|---|---|
| yes (1) | $p(1,1)$, equal treatment | $p(1,0)$, $B$ discriminated |
| no (0) | $p(0,1)$, $A$ discriminated | $p(0,0)$, equal treatment |

where $p(d_A, d_B)$ is the proportion of answer for which candidate A was answered $d_A \in \{0,1\}$ and candidate $B$ was answered $d_B \in \{0,1\}$. Equal treatment happens when both candidates have the same answers. The corresponding proportion equals $p(0,0)+p(1,1)$. The other proportions indicate an unequal treatment, which is interpreted as a discrimination because the applications have been made equivalent. In practice, the researchers use the *net* discrimination coefficient against candidate $B$:

$$\begin{aligned} D(A,B) &= \Pr(\text{A called back}) - \Pr(\text{B called back}) \\ &= p(1,1) + p(1,0) - (p(1,1) + p(0,1)) \\ &= p(1,0) - p(0,1) \end{aligned}$$

which is the excess probability of discriminatory cases against candidate $B$.[3] Testing the equality of the callbacks of the two candidates is equivalent to test $p(1,0) = p(0,1)$. If a correspondence test is done on a given occupation, this means that we define discrimination at the job market level rather than at the firm level. $D(A,B) > 0$ means that there is discrimination against candidate $B$ at the job market level and $D(A,B) < 0$ that there is discrimination against candidate $A$.

Now let us consider the discrimination test globally. What critical value should we use? The null hypothesis is the absence of discrimination $D(A,B) = 0$ against $D(A,B) \neq 0$. Therefore we may defined the type I and II errors as in Table 2.

Table 2: Error types

| Truth / Conclusion | $D(A,B) = 0$ | $D(A,B) \neq 0$ |
|---|---|---|
| $D(A,B) = 0$ | Correct | Type II error ($\beta$) |
| $D(A,B) \neq 0$ | Type I error ($\alpha$) | Correct |

The standard test will take the absence of discrimination as the null hypothesis ($H_0 : D(A,B) = 0$). Then, we should distinguish two types of errors. The Type I error ($\alpha$) occurs when we reject the null hypothesis while it is true; in our context, it means that we conclude that there is discrimination while there is none. The type II error happens when we accept the null hypothesis while it is wrong; in other words, we conclude

[2] By a standard convention, the absence of answer is interpreted as a no.

[3] Also notice that $D(A,B) = -D(B,A)$.

that there is no discrimination while there is some ($\beta$). The power of the test is $\zeta = 1-\beta$, it is the probability to conclude that there is discrimination when this is true. Standard test theory fixes $\alpha$ and let $\beta$ (and so $\zeta$) free. Moreover, the lower $\alpha$, the lower $\zeta$. On small samples, a small type I error can correspond to a very small power. By setting a low value for $\alpha$, like 1%, we can end up with a very small probability to find discrimination when there is some. A first approach keeps the same sample size and argues that $\alpha$ should be taken relatively "high", like 5% or 10%. A second approach proposes to determine an optimal sample size for a given testing problem. For a given $\alpha$, we will get a higher $\zeta$ if the sample size is larger.

Consider the first approach. Ideally, we would like to measure the cost of each error type and to minimize a loss function. What is the cost of a type I error? If we conclude that there is discrimination while there is none, the cost should be close do 0. Indeed, if there is no discrimination, nobody can prove that there is some before a court and no firm should be condemned for this. There should be no important prejudice. A type II error can also be costly, since it implies that we conclude to no discrimination while there is some. This conclusion could weaken the arguments of the workers discriminated against since discrimination is hard to prove. Moreover, this cost should be multiplied by the size of the population. Considering gender discrimination, even a small discrimination coefficient would imply a very high cost because of the large size of the population. Inconclusive results could even be used to weaken anti-discrimination policies. This gives an incentive not to take a low $\alpha$.

But the most satisfactory answer seems to determine an optimal sample size before to run the experiment, so as to avoid adjusting the $\alpha$ on a qualitative, and somewhat arbitrary, basis. We show below how to compute a minimal sample size, after presenting the paired Student tests.

Once that the data have been collected we may seek to measure discrimination. However some caution is to be taken. First we must keep in mind that the answers to all the candidates on a same job are correlated, since the same recruiter replies to all the candidates. This precludes to use independent samples tests or standard ordinary least squares (henceforth, OLS) methods. We will use *paired* tests. A common practice consists in using stacked OLS regressions. We show that it provides both a good measurement of discrimination and a bad estimation of the variances. The reason is that standard OLS assume the independence of all observations when computing the standard errors and, for this reason, cannot be used for a valid inference. Instead, one should use either paired OLS regressions or clustered standard errors. We show that the method of Arellano (1987), developed originally for panel data, can be used to compute the standard error used in paired Student statistics. The next section describes how it is possible to extract information from more complicated correspondence tests, and to reveal the discriminatory component inside the callback rates. This part may also be useful for the tests where all the candidates could not be sent on all the job adds. we show how to make an optimal use of all the information available.

**Proportions tests.** Since we send several candidates on the same job ad, we can compare them easily with a paired test. The advantage of paired tests is that, on the one hand, they use the differences of treatment between two candidates so that their interpretation is straightforward and, on the other hand, the paired tests do not require to account explicitly for the correlation among the answers to the candidates, since the statistic includes the right correction.

We consider two candidates selected among the total number of candidates. Consider a job add $i \in \mathbf{I}$, where $\mathbf{I}$ denotes the index set of all adds and let I denote their number. The set of candidates is denoted $\mathbf{J}$ and their number J. For candidate $j \in \mathbf{J}$ the answer is coded as a dummy variable denoted $d_{ji} \in \{0,1\}$, where, by convention, $d_{ji} = 0$ represents a "no" and $d_{ji} = 1$ a "yes". The theoretical success probabilities associated with the candidates are denoted $p_j$. By definition $d_{ji}$ follows a Bernoulli process with probability $p_j$. Consider two candidates, $\mathbf{J} = \{A, B\}$, we wish to perform the following test:

$$\begin{aligned} \mathrm{H}_0 &: p_A - p_B = 0 \\ \mathrm{H}_a &: p_A - p_B \neq 0 \end{aligned}$$

**Paired Student test.** The null hypothesis is the absence of discrimination since both candidates are treated equally on the same ads. In case of rejection, we can have either discrimination against candidate A ($p_A < p_B$) or discrimination against candidate B ($p_A > p_B$). The test can be done with a Student test. However, the reader should keep in mind that we need a *paired* Student test. Indeed, the answers to all the candidates on a given ad are correlated because they are given by the same recruiter. Therefore, we cannot assume that the two candidates come from two independent samples, as is usually done with the two-sample Student test. By convention, we will denote the means with a bar, so that the empirical callback rate of candidate $j$ is denoted:

$$\bar{d}_j = \frac{1}{\mathrm{I}} \sum_{I \in \mathbf{I}} d_{ji}, \ j \in \mathbf{J}$$

When we have paired data, we can compute the difference of answers between the two candidates for each ad and take the mean on this difference. Consider two candidates for the job ad $i$, we observe the couple of answers $(d_{Ai}, d_{Bi}),\ i \in \mathbf{I}$. The difference between the answers to candidates A and B, denoted $\delta_i = d_{Ai} - d_{Bi}$ can take three values :

$$\delta_i = \begin{cases} 1 & \text{candidate A prefered} \\ 0 & \text{equal treatment} \\ -1 & \text{candidate B prefered} \end{cases}$$

The paired test is simply the Student test of a zero mean on this difference since:

$$\mathrm{E}(\delta) = \mathrm{E}(d_A - d_B) = p_A - p_B.$$

Since $\bar{\delta} = \bar{d}_A - \bar{d}_B$, testing that this quantity is close to 0 is equivalent to test for the absence of discrimination. The paired Student statistic, denoted $T_p$ equals:

$$T_p = \frac{|\bar{\delta}|}{\sqrt{\hat{\mathrm{V}}(\bar{\delta})}}$$

where $\hat{\mathrm{V}}$ denotes the empirical variance. The following unbiased estimator is commonly used:

$$\hat{\mathrm{V}}(\bar{\delta}) = \frac{\hat{\mathrm{V}}(\delta)}{\mathrm{I}} = \frac{1}{\mathrm{I}(\mathrm{I}-1)} \sum_{i \in \mathbf{I}} (\delta_i - \bar{\delta})^2$$

One can easily show that :

$$\hat{\mathrm{V}}(\delta) = \hat{\mathrm{V}}(d_A) + \hat{\mathrm{V}}(d_B) - 2\hat{\mathrm{Cov}}(d_A, d_B)$$

where $\hat{\text{Cov}}$ is the empirical covariance.[4] Notice that the two-sample Student statistic is obtained only when $\hat{\text{Cov}}(d_A, d_B) = 0$. The $T_p$ statistic is used to test the absence of net discrimination. On large samples, we use the normal approximation.[5] For a test at the $\alpha$ level, we will reject the null hypothesis when $T_p \geq z_{1-\alpha/2}$, where $z$ is the quantile of the standard normal distribution.[6] Notice that we measure *net* discrimination since some firms may prefer candidate $A$ and others candidate $B$. What we measure is whether, at the job market level, comparable candidates are treated equally.

**Minimum sample size.** The impossibility to control the power of the test may be undesirable from a social viewpoint. In this section we show how to compute the sample size needed, for a test with a level $\alpha$, a power $\zeta$ and a minimum detectable value for $\delta = p_A - p_B$. We consider the following test:

$$\begin{aligned} \mathrm{H}_0 &: \delta = 0 \\ \mathrm{H}_a &: \delta = \delta^* \end{aligned}$$

where $\delta^* > 0$ is a minimum detectable value. It represents the difference threshold from which we consider that there is a relevant amount of discrimination. We wish to reject $\mathrm{H}_0$ with probability $\alpha$ when $\mathrm{H}_0$ is true and to reject $\mathrm{H}_0$ with probability $\zeta$ when $\mathrm{H}_a$ is true. In order to perform the test, we have a sample of treatment differences $\delta_i$, and compute their mean $\bar{\delta}$. We assume that $\delta_i$ has variance $\sigma^2$. Using a normal approximation, $\bar{\delta}$, should be distributed with a variance equal to $\sigma^2/\mathrm{I}$, with a 0 mean under the null hypothesis and a $\delta^*$ mean under the alternative hypothesis. We should have:

$$\Pr\left(\bar{\delta} > \frac{\sigma u_{1-\alpha}}{\sqrt{\mathrm{I}}} \middle| \mathrm{H}_0\right) = \alpha$$

where $u_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution (1.645 for $\alpha = 0.05$). Under the null hypothesis, $\bar{\delta}$ should converge to zero and this equality should be satisfied. Now, consider the power inequality. We should have:

$$\Pr\left(\bar{\delta} > \frac{\sigma u_{1-\alpha}}{\sqrt{\mathrm{I}}} \middle| \mathrm{H}_a\right) \geq \zeta.$$

We have the following equivalence:

$$\bar{\delta} > \frac{\sigma u_{1-\alpha}}{\sqrt{\mathrm{I}}} \Leftrightarrow \sqrt{\mathrm{I}} \times \frac{\bar{\delta} - \delta^*}{\sigma} > u_{1-\alpha} - \sqrt{\mathrm{I}} \times \frac{\delta^*}{\sigma}$$

since the left-hand term follow a standard normal distribution under $\mathrm{H}_a$. We get:

$$\begin{aligned} \Pr\left(\sqrt{\mathrm{I}} \times \frac{\bar{\delta} - \delta^*}{\sigma} > u_{1-\alpha} - \sqrt{\mathrm{I}} \times \frac{\delta^*}{\sigma}\right) &= 1 - \Phi\left(u_{1-\alpha} - \sqrt{\mathrm{I}} \times \frac{\delta^*}{\sigma}\right) \\ &= \Phi\left(\sqrt{\mathrm{I}} \times \frac{\delta^*}{\sigma} - u_{1-\alpha}\right) \end{aligned}$$

[4] An unbiased estimator is:

$$\hat{\text{Cov}}(d_A, d_B) = \frac{1}{I-1} \sum_{i_1 \in \mathbf{I}} \sum_{i_2 \in \mathbf{I}} (d_{Ai1} - \bar{d}_A)(d_{Bi2} - \bar{d}_B)$$

.

[5] The Student distribution with $N$ degrees of freedom converges to the standard normal distribution when $N \to +\infty$.

[6] For $\alpha$ =1%, 5% or 10%, we get the respective critical values $z$ =2.58, 1.96 and 1.645.

so that:

$$\Phi\left(\sqrt{\mathrm{I}} \times \frac{\delta^*}{\sigma} - u_{1-\alpha}\right) \geq \zeta \Leftrightarrow \mathrm{I} \geq \left(\frac{\sigma}{\delta^*}(\Phi^{-1}(\zeta) + u_{1-\alpha})\right)^2.$$

Let us take an example. Let $p_A = 0.50$ and $p_A - p_B = \delta^* = 0.10$. We also let $\alpha = \beta = 10\%$ so that the power is set at 90% ($z_{0.9} = 1.28$). We just need the variance of the statistic. The callback dummies verify $\mathrm{V}(d_j) = p_j(1-p_j)$ and we should compute :

$$\mathrm{V}(d_A - d_B) = \mathrm{V}(d_A) + \mathrm{V}(d_A) - 2\mathrm{Cov}(d_A, d_B)$$

and we do not know their covariance. But we can use the following bound on the correlation coefficient:

$$-1 \leq \frac{\mathrm{Cov}(d_A, d_B)}{\sqrt{\mathrm{V}(d_A)\mathrm{V}(d_B)}} \leq 1$$

which implies that the variances :

$$\sigma^2 = \mathrm{V}(d_A - d_B) \leq \mathrm{V}(d_A) + \mathrm{V}(d_A) + 2\sqrt{\mathrm{V}(d_A)\mathrm{V}(d_B)} = \tilde{\sigma}^2$$

and we can use it to get an a conservative value for $I$. In our example, we get:

$$\mathrm{V}(d_A) = 0.25,\ \mathrm{V}(d_B) = 0.24, \tilde{\sigma}^2 = 0.98,$$

so that:

$$\mathrm{I} \geq \frac{0.98}{0.01} \times (3.29)^2 \simeq 642.$$

which is a relatively high value. A median assumption may ignore the covariance (since it can be negative or positive) and in this case we would be $\sigma^2 = 0.49$ and $\mathrm{I} \geq 321$.

**Paired OLS regression.** It is possible to perform the previous test by running a very simple regression. One should just regress the treatment difference variable $\delta_i$ on the constant term. The model is: $\delta_i = b_0 \times 1 + u_i$ where $b_0$ is the intercept and $u_i$ the disturbance. Let $e_\mathrm{I}$ be the column unit vector of size I and $\delta$ the column vector of all $\delta_i$. The OLS formula gives:

$$\begin{aligned}\hat{b}_0 &= (e_\mathrm{I}' e_\mathrm{I})^{-1} e_\mathrm{I}' \delta \\ &= (\sum_{i \in \mathbf{I}} 1^2)^{-1} \sum_{i \in \mathbf{I}} 1 \times \delta_i \\ &= \frac{1}{\mathrm{I}} \sum_{i \in \mathbf{I}} \delta_i \\ &= \bar{\delta}\end{aligned}$$

The residual of this regression equals $\hat{u}_i = \delta_i - \hat{b}_0 = \delta_i - \bar{\delta}$. Furthermore, denote $\hat{\sigma}^2$ the empirical variance of the residual, we have:

$$\hat{\mathrm{V}}(\hat{b}_0) = \hat{\sigma}^2 (e_\mathrm{I}' e_\mathrm{I})^{-1} = \frac{\hat{\sigma}^2}{\mathrm{I}}$$

and standard OLS software will provide:

$$\hat{\sigma}^2 = \frac{1}{\mathrm{I}-1} \sum_{i \in \mathbf{I}} \hat{u}_i^2 = \frac{1}{\mathrm{I}-1} \sum_{i \in \mathbf{I}} (\delta_i - \bar{\delta})^2$$

so that we get $\hat{\mathrm{V}}(\hat{b}_0) = \hat{\mathrm{V}}(\bar{\delta})$. The Student statistic of the paired OLS regression is identical to the paired Student statistic. Unfortunately, this regression is not always used. Instead, stacked regressions are often preferred. We show in the next section that they provide the wrong variance and how it can be corrected.

**Stacked OLS regressions.** Stacked regressions are popular in the applied literature. In this setting, one performs a regression of the callback dummy variable on a constant term and a dummy variable for the reference group. Consider our example, we define a dummy variable $x_{ji}$ equal to 1 for belonging to group $A$ (equal to 0 otherwise):

$$\forall i,\ x_{ji} = \begin{cases} 1 & \text{if } j = A \\ 0 & \text{if } j = B \end{cases}$$

where we notice that the definition does not depends on $i$. This is because, in a field experiment, the explanatory variables are *constructed*: we send the two candidates on the same ads. Another interesting point is the number of observations: we stack the I answers of candidates A and B, so that the number of observations in the regression in now 2I. In order to get the probability difference, we write a linear probability model as:

$$d_{ji} = c_0 + c_1 x_{ji} + u_{ji}$$

with $\mathrm{E}(u_{ji}) = 0$ without loss of generality provided the model includes a constant term. We easily get:

$$\begin{aligned} \mathrm{E}(d_{Ai}) &= c_0 + c_1 \\ \mathrm{E}(d_{Bi}) &= c_0 \end{aligned}$$

so that, using $\mathrm{E}(d_{ji}) = p_j$, we get :

$$\begin{aligned} c_0 &= p_B \\ c_1 &= p_A - p_B \end{aligned}$$

therefore the theoretical value of the OLS coefficient $c_1$ gives the discrimination coefficient. The relationship is also valid for the empirical counterparts. We rewrite the model as:

$$d_{ji} = X_{ji} c + u_{ji}$$

with $X_{ji} = (1, x_{ji})$ and $c' = (c_0, c_1)$. Applying OLS is equivalent to solve the system $X'X\hat{c} = X'd$. The cross products are specific: $e'_{2\mathrm{I}} e_{2\mathrm{I}} = 2\mathrm{I}$ since there are 2I observations after stacking. For $e'_{2\mathrm{I}} x = x' e_{2\mathrm{I}}$, we just need to consider that $x_{Ai} = 1$ and $x_{Bi} = 0, \forall i$. This gives:

$$\begin{aligned} e'_{2\mathrm{I}} x &= \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} x_{ji} \\ &= \sum_{i \in \mathbf{I}} 1 \\ &= I \end{aligned}$$

and for $x'x$ we only need to remark that $x_{ji} = x_{ji}^2$ since it is a binary variable. Therefore:

$$\begin{aligned} x'x &= \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} x_{ji}^2 \\ &= \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} x_{ji} \\ &= \mathrm{I}. \end{aligned}$$

For the right-hand cross products, we get:

$$
\begin{aligned}
e_1'd &= \sum_{i\in\mathbf{I}}\sum_{j\in\mathbf{J}} d_{ji} \\
&= \sum_{i\in\mathbf{I}} d_{Ai} + \sum_{i\in\mathbf{I}} d_{Bi} \\
&= \mathrm{I}\times(\bar{d}_A+\bar{d}_B)
\end{aligned}
$$

the total number of callbacks, and

$$
\begin{aligned}
x'y &= \sum_{i\in\mathbf{I}}\sum_{j\in\mathbf{J}} x_{ji}d_{ji} \\
&= \sum_{i\in\mathbf{I}} d_{Ai} \\
&= I\times\bar{d}_A
\end{aligned}
$$

the number of callbacks received by candidate 1. Let $\mathrm{I}_j = \mathrm{I}\times\bar{d}_j$ be the number of callbacks for the candidate $j$, we should solve:

$$
\begin{pmatrix} 2\mathrm{I} & \mathrm{I} \\ \mathrm{I} & \mathrm{I} \end{pmatrix}\begin{pmatrix} \hat{c}_0 \\ \hat{c}_1 \end{pmatrix} = \begin{pmatrix} \mathrm{I}_A+\mathrm{I}_B \\ \mathrm{I}_A \end{pmatrix}.
$$

Using:

$$
\begin{pmatrix} 2I & I \\ I & I \end{pmatrix}^{-1} = \frac{1}{I}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix},
$$

we get :

$$
\begin{pmatrix} \hat{c}_0 \\ \hat{c}_1 \end{pmatrix} = \frac{1}{\mathrm{I}}\begin{pmatrix} \mathrm{I}_B \\ \mathrm{I}_A-\mathrm{I}_B \end{pmatrix} = \begin{pmatrix} \bar{d}_B \\ \bar{d}_A-\bar{d}_B \end{pmatrix}
$$

the OLS regression coefficient $\hat{c}_1$ of the group dummy variable $x_{ij}$ equals $\bar{\delta}$ and measures discrimination.

The only problem with OLS is that it gives wrong standard errors. We will first show that it gives the two-sample Student variance and how to fix the problem with a clusterisation of the covariance matrix. For the variance, we need to compute $\hat{\sigma}_v^2(X'X)^{-1}$. Here we simply notice that $\forall i$:

$$
\begin{aligned}
\hat{u}_{Ai} &= d_{Ai} - (\hat{c}_0+\hat{c}_1) = d_{Ai} - \bar{d}_A \\
\hat{u}_{Bi} &= d_{Bi} - (\hat{c}_0) = d_{Bi} - \bar{d}_B
\end{aligned}
$$

For the variance we should consider that there are 2I observations and 2 parameters, which makes $2\mathrm{I}-2 = 2(\mathrm{I}-1)$ degrees of freedom:

$$
\begin{aligned}
\hat{\sigma}_u^2 &= \frac{1}{2(\mathrm{I}-1)}\sum_{i\in\mathbf{I}}\sum_{j\in\mathbf{J}} \hat{u}_{ji}^2 \\
&= \frac{1}{2(\mathrm{I}-1)}\sum_{i\in\mathbf{I}} (d_{Ai}-\bar{d}_A)^2 + (d_{Bi}-\bar{d}_B)^2 \\
&= \frac{1}{2}(\hat{\mathrm{V}}(d_A)+\hat{\mathrm{V}}(d_B))
\end{aligned}
$$

and

$$\hat{\mathrm{V}}(\hat{c}) = \frac{1}{2}(\hat{\mathrm{V}}(d_A) + \hat{\mathrm{V}}(d_B)) \times \frac{1}{\mathrm{I}} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

so that :

$$\hat{\mathrm{V}}(\hat{c}_1) = \frac{1}{\mathrm{I}}(\hat{\mathrm{V}}(d_A) + \hat{\mathrm{V}}(d_B))$$

which is the variance used in the two-sample Student statistic. It is not valid in this context because the answers to a same ad are not independent from each other. We show how to correct it in the next section.

**Clustered variances.** We need to account for the correlation between the callback dummies from the same ad. When the disturbance covariance matrix is $\mathrm{V}(u) = \sigma_u^2 \mathrm{Id}_\mathrm{I}$, the OLS variance formula is $\mathrm{V}(\hat{c}) = \sigma_u^2 (X'X)^{-1}$ but when there are correlations between the observations, we must write $\mathrm{V}(u) = \Omega$ and the OLS variance formula becomes $\mathrm{V}(\hat{c}) = (X'X)^{-1} X'\Omega X (X'X)^{-1}$. In our case $\Omega$ has a specific shape, it is block diagonal, since the callback decisions are assumed to be correlated on the *same* ad only. Let $\Omega_i$ be the $2 \times 2$ ad-level covariance matrix, $d_i' = (d_{Ai}, d_{Bi})$ the callback vector of ad $i$ and

$$X_i = \begin{pmatrix} 1 & x_{Ai} \\ 1 & x_{Bi} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \; U_n = \begin{pmatrix} u_{Ai} \\ u_{Bi} \end{pmatrix}$$

the explanatory variables matrix and the disturbance vector for ad $i$. Notice that the explanatory variable $x$ is chosen by the econometrician and takes alternatively the values 1 and 0 on each ad. The OLS variance equals:

$$\mathrm{V}(\hat{c}) = \left( \sum_{i \in \mathbf{I}} X_i' X_i \right)^{-1} \sum_{i \in \mathbf{I}} X_i' \Omega_i X_i \left( \sum_{i \in \mathbf{I}} X_i' X_i \right)^{-1}$$

The structure of the data is similar to a balanced panel data problem, where the ad $i$ represents the individual, and the candidate $j$ the time dimension. Arellano (1987) proposed the following estimator:

$$\hat{\mathrm{V}}_A(\hat{c}) = \left( \sum_{i \in \mathbf{I}} X_i' X_i \right)^{-1} \sum_{i \in \mathbf{I}} X_i' \hat{U}_i \hat{U}_i' X_i \left( \sum_{i \in \mathbf{I}} X_i' X_i \right)^{-1}$$

In order to get the paired Student test, we just need to add the following correction for the degrees of freedom:

$$\hat{\mathrm{V}}(\hat{c}) = \frac{\mathrm{I}}{\mathrm{I}-1} \hat{\mathrm{V}}_A(\hat{c}).$$

Using :

$$\hat{U}_i = \begin{pmatrix} d_{Ai} - \bar{d}_A \\ d_{Bi} - \bar{d}_B \end{pmatrix}$$

we get:

$$\sum_{i \in \mathbf{I}} X_i' \hat{U}_i \hat{U}_i' X_i = (\mathrm{I} - 1) \times \begin{pmatrix} \hat{\mathrm{V}}(d_A) + 2\hat{\mathrm{Cov}}(d_A, d_B) + \hat{\mathrm{V}}(d_B) & \hat{\mathrm{V}}(d_A) + \hat{\mathrm{Cov}}(d_A, d_B) \\ \hat{\mathrm{V}}(d_A) + \hat{\mathrm{Cov}}(d_A, d_B) & \hat{\mathrm{V}}(d_A) \end{pmatrix}$$

and

$$\left(\sum_{i\in\mathbf{I}} X_i' X_i\right)^{-1} = \mathrm{I}^{-1} \times \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

so that

$$\begin{aligned} \hat{\mathrm{V}}(\hat{c}) &= \begin{pmatrix} \hat{\mathrm{V}}(\bar{d}_B) & \hat{\mathrm{Cov}}(\bar{d}_B, \bar{d}_A - \bar{d}_B) \\ \hat{\mathrm{Cov}}(\bar{d}_B, \bar{d}_A - \bar{d}_B) & \hat{\mathrm{V}}(\bar{d}_A - \bar{d}_B) \end{pmatrix} \\ &= \begin{pmatrix} \hat{\mathrm{V}}(\bar{d}_B) & \hat{\mathrm{Cov}}(\bar{d}_B, \bar{\delta}) \\ \hat{\mathrm{Cov}}(\bar{d}_B, \bar{\delta}) & \hat{\mathrm{V}}(\bar{\delta}) \end{pmatrix}. \end{aligned}$$

Therefore, one just need to multiply the Arellano Student by $\sqrt{(I-1)/I}$ in order to get the paired t statistic. This correction is negligible on large samples.

### 1.3 Regression models

The previous models may be augmented by a list of explanative variables. Indeed, the tester can choose the characteristics of the candidates, but s-he cannot choose the characteristics of the firm. This opens the possibility to study conditional discrimination, which is a discrimination that occurs only on some labour contracts. For instance, a worker may be discriminated on a long term contract but not on a short term contract. In order to test for this assumption, we should add the contract term into the model and evaluate its impact on the difference of treatment between the candidates. There are two main ways to extend the comparison tests: paired regressions and stacked regressions. With paired regressions, the relevant variable is the callback difference between two candidates. Any significant variable may reveal conditional discrimination. Let us consider a recruiter that chooses between two candidates, $j = A, B$. We will assume that each candidates generates a utility level, and that a candidate will be invited each time a threshold is crossed. The utility of recruiter $i$ for candidate $j$ is denoted:

$$v_{ji}^* = X_i b_j + \alpha_i + \varepsilon_{ji} \tag{1}$$

where $X_i$ are the observable firm characteristics, $b_j$ the regression coefficient for the candidate $j$. The $\alpha_i$ term is the job ad correlated effect (or "fixed effect"), since the same recruiter replies to all the candidates and $\varepsilon_{ji}$ is the idiosyncratic error term, typically a white noise. We observe the callback dummy variable:

$$v_{ji} = \begin{cases} 1 & \text{if } v_{ji}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where 0 is the normalized reservation utility level without loss of generality provided the model includes a constant term. Notice that it may also depend on each recruiter $i$ through the unobservable term $\alpha_i$. The real assumption here is that all the candidates face the same threshold, a point that may not always be true.[7] Comparing two candidates $A$ and $B$, we can consider their utility difference for the recruiter:

$$\delta_i^* = v_{Ai}^* - v_{Bi}^* \tag{2}$$

[7] Some recruiters may impose a higher standard to discriminated candidates (Heckman, 1998).

three outcomes are possible :

$$\delta_i = \begin{cases} 1 & \text{A invited alone, if } \delta_i^* \geq c_A \\ 0 & \text{equal treatment, if } c_B \leq \delta_i^* < c_A \\ -1 & \text{B invited alone, if } \delta_i^* < c_B \end{cases}$$

where $c_A$ and $c_B$ are unknown thresholds. With our notations these three events will occur with respective probabilities $p_{10}$, $p_{00} + p_{11}$ and $p_{10}$. In other words, candidates with similar utility levels will be treated equally, and discrimination will occur when the recruiter perceives too strong a difference between them. We also notice that the expectation of this variable equals:

$$\mathrm{E}(\delta) = 1 \times p_{10} + 0 \times (p_{00} + p_{11}) - 1 \times p_{10} = D(A,B)$$

the discrimination coefficient we have already seen. The most straightforward extension to regression is obtained by combining (1) and (2). We get:

$$\begin{aligned} \delta_i^* &= X_i b_A + \alpha_i + \varepsilon_{Ai} - (X_i b_B + \alpha_i + \varepsilon_{Bi}) \\ &= X_i \beta + \eta_i. \end{aligned}$$

with $\beta = b_A - b_B$ and $\eta_i = \varepsilon_{Ai} - \varepsilon_{Bi}$. The specific form of the model will depend on the distributional assumption about $\eta$. The simplest case is the difference of two linear probability models, which gives a linear probability model. The $\delta$ parameter will measure discrimination since we explain a difference of treatment. In order to simplify the exposition, consider that we run a regression on the centred variables:

$$\delta_i = \beta_0 + (x_i - \bar{x})\beta_1 + \eta_i$$

Taking the expectation we get $\mathrm{E}(\delta_i) = \beta_0$, the constant term of the model. This is the unconditional discrimination coefficient at the mean point of the sample ($x_i = \bar{x}$). When $\beta_1 = 0$, it is equal to $D(A,B)$. It does not depend on the firms' characteristics or on the characteristics of the labour contract.[8] The other discrimination term $\beta_1$, on the contrary, acts in interaction with the ad $i$'s characteristics. This is the vector of the conditional discrimination coefficients. A positive coefficient will indicate a discrimination source against candidate $B$, and a negative coefficient against candidate $A$. This equation is easily estimated by OLS and does not need a clusterization of the variances because the left-hand variable is taken in differences.

**Ordered models.** Although convenient, the linear probability model has the following default: it could happen that the predictions get outside the $[-1,1]$ range which is compulsory for $\delta_i$ since it is the difference of two binary variables.[9] In this case, one may prefer making another assumption about the distribution of $\eta_i$. Assuming normality, $\eta_i \rightsquigarrow \mathrm{N}(0,\sigma^2)$ will give an ordered Probit model, with:[10]

[8] Notice that the constant term of this regression does not have the usual interpretation because the left-hand variable is a difference.

[9] One can check directly the predictions of the model.

[10] Ordered Logit models have also been used. It is possible to choose the model with a Vuong test (Vuong, 1989), like in Duguet and Petit (2005).

$$
\begin{aligned}
\Pr(d_i = 1) &= \Pr[X_i b + \eta_i \geq c_A] \\
&= 1 - \Phi\left(\frac{c_A - X_i\beta}{\sigma}\right), \\
\Pr(d_i = 0) &= \Pr[c_B \leq X_i\beta + \eta_i < c_A] \\
&= \Phi\left(\frac{c_A - X_i\beta}{\sigma}\right) - \Phi\left(\frac{c_B - X_i\beta}{\sigma}\right) \\
\Pr(d_i = -1) &= \Pr[X_i\beta + \eta_i \leq c_B] \\
&= \Phi\left(\frac{c_B - X_i\beta}{\sigma}\right).
\end{aligned}
$$

With this type of model, one should be careful that the thresholds will absorb the constant term. For instance, consider the last probability, with $\beta' = (\beta_0, \beta_1')$ and $X_i = (1, x_i - \bar{x})$:

$$
\Phi\left(\frac{c_B - X_i\beta}{\sigma}\right) = \Phi\left(\frac{c_B - \beta_0 - \beta_1(x_i - \bar{x})}{\sigma}\right)
$$

so that we define a new set of parameters: $\gamma_B = (c_B - \beta_0)/\sigma$ and $\gamma_1 = \beta_1/\sigma$ and get the probability:

$$
\Phi\left(\gamma_B - \gamma_1(x_i - \bar{x})\right).
$$

Performing a similar operation for the first probability, we define $\gamma_A = (c_A - \beta_0)/\sigma$, so that overall:

$$
\begin{aligned}
\Pr(d_i = 1) &= 1 - \Phi\left(\gamma_A - \gamma_1(x_i - \bar{x})\right), \\
\Pr(d_i = 0) &= \Phi\left(\gamma_A - \gamma_1(x_i - \bar{x})\right) - \Phi\left(\gamma_B - \gamma_1(x_i - \bar{x})\right) \\
\Pr(d_i = -1) &= \Phi\left(\gamma_B - \gamma_1(x_i - \bar{x})\right)
\end{aligned}
$$

The software will typically give an estimate of $(\gamma_A, \gamma_B, \gamma_1)$. When the right-hand variables are centred, the discrimination coefficient at the average point of the sample $(x_i = \bar{x})$ will be given by:

$$
\begin{aligned}
\bar{D}(A, B) &= \Pr(d_i = 1) - \Pr(d_i = -1) \\
&= 1 - \Phi(\gamma_A) - \Phi(\gamma_B)
\end{aligned}
$$

and its asymptotic variance can be estimated by the delta method. Letting $g(\gamma) = 1 - \Phi(\gamma_A) - \Phi(\gamma_B)$, we get:

$$
\hat{\mathrm{V}}\hat{\gamma}) = \frac{\delta g}{\delta \gamma'}(\hat{\gamma})\hat{\Omega}_\gamma \left(\frac{\delta g}{\delta \gamma'}(\hat{\gamma})\right)'
$$

where $\hat{\Omega}_\gamma$ is the estimated asymptotic covariance matrix of $\hat{\gamma}$ provided by the software. It is also possible to estimate an effect for each observation of $d_i$, and to compute their average or to draw their density.

**Stacked regressions.** When the data is stacked, the estimation raises specific issues. Consider the equation (1). Considering two candidates for the same ad, we have:

$$
\begin{aligned}
v_{Ai}^* &= X_i b_A + u_{Ai} \\
v_{Bi}^* &= X_i b_B + u_{Bi}
\end{aligned}
$$

with $u_{ji} = \alpha_i + \epsilon_{ji}$. Letting $\mathrm{V}(\alpha_i) = \sigma_\alpha^2$, $\mathrm{V}(\varepsilon_{ji}) = \sigma_j$ and $u_i' = (u_{Ai}, u_{Bi})$ we conclude that there is an ad-level block correlation in the model:

$$\mathrm{V}(u_i) = \begin{pmatrix} \sigma_\alpha^2 + \sigma_A^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_B^2 \end{pmatrix}$$

so that a clusterization of the covariance matrix will be needed. Also notice that the model is heteroskedastic between the two candidates, like in Neumark (2012). An additional issue is raised when the ad fixed effect $\alpha_i$ is correlated with the right-hand variables: the estimates may not be consistent any more.[11]

Furthermore, even in the favourable case where the effect in not correlated, the model should allow for $b_A \neq b_B$. The only solution is to add the cross products of a group dummy with the $X_i$ variables. In order to develop the argument, let us rewrite the estimation problem. The individual will be indexed by $n = 1, \ldots, 2I$ since there are two candidates for the $I$ ads. We introduce a dummy variable, $A_n$, equal to 1 if the individual $n$ is in the potentially favoured group ($A$), 0 otherwise. Let us also assume that the data is stacked at the ad level, so that all the odd indices refer to the $A$ candidates and all the even indices to the $B$ candidates.[12]The regression equation should be written:

$$v_n^* = X_n b_A + A_n X_n (b_A - b_B) + u_n$$

where $u_n = u_{A,(n+1)/2}$ if $n$ is odd and $u_n = u_{B,n/2}$ if $n$ is even. When $A_n = 1$, we get $X_i b_A + u_{Ai}$ and when $A_n = 0$ we get $X_i b_B + u_{Bi}$. Notice that the $X$ variables include the constant term, so that the group dummy is among the regressors. A widespread practice consists in estimating a model with $X$ and $A$ alone. It will be valid as long as $b_A = b_B$ for all the variables but the intercept. To make it more sensible, let $X_n = (1, x_n)$ and $b_j' = (b_{0j}, b_{1j})$ for $j \in \{A, B\}$. We get:

$$v_n^* = b_{0A} + b_{1A} x_n + A_n (b_{0A} - b_{0B}) + A_n x_n (b_{1A} - b_{1B})$$

and it is clear that the model with $x_n$ and $A_n$ alone can only be valid if $b_{1A} = b_{1B}$.

Overall the stacked model raises two problems. First, the fixed effect problem. When there is no significant right-hand variable, ignoring the issue is not problematic (as we saw), because there is no variable in the model that is susceptible to be correlated with the fixed effect. But this result does not extend to models with explanative variables, since a correlation with the fixed effect leads to inconsistent estimates. A second problem happens when the coefficients of the explanative variables differ in the two groups of workers. One should include the cross products of the explanative variables with the group dummy among the regressors. Forgetting this cross products will create a missing variable bias, because these missing variables are, by definition, correlated with the regressors.

In conclusion, the paired ordered models, based on a difference of treatment between two candidates seem to be a better tool because they eliminate the fixed effect and allow for a differentiated effect of the explanative variables among the applicants to a job. They are also more easy to implement, since they do not require a heteroskedastic Probit estimation.

[11]The issue is similar to the estimation of panel data models. When the individual effect is not correlated, one must account for the covariance matrix between the disturbances. When the effect is correlated, one should difference out the individual effects.

[12]The indices of the $A$ candidates are given by $2i - 1$, for $i = 1, \ldots, I$, and the indices of $B$ candidates by $2i$.

## 1.4 Component models

Some correspondence test have a more complicated structure than the basic tests that we have studied so far. There may be several characteristics studied together (e.g, gender and age) so that several candidates may be affected by the same type of discrimination. Young women and older women may both be discriminated because they are women. Another example happens when all the candidates are not send to all the job adds, but only a part of them with a rotation. All the paired comparisons cannot be made on all the job adds, but only a part of them. In This case, we need to combine the information from all the possible combinations in a coherent and efficient manner (Duguet et al. (2018)). Several types of models can be used: linear or transformed. In the first case, the estimation will proceed from callback rates differences of difference in differences. It may be inconvenient when the callback rates are close to zero and the model leads to prediction outside the unit interval. In the second type of model, the predictions are bounded, and one has to work with concepts like odds ratios rather than rates. Overall, both models allow to control for unobserved heterogeneity. These models are also useful in overidentified cases. They allow for testing some restrictions in the structure of discrimination.

**A rationalization of differences method.** A first way to use component model is for determining the right differences to use in the comparisons and to deal with the case where there are several ways to compute a given coefficient. Consider a situation with two types of discrimination: gender and origin. We send four candidates on each offer:[13] the local origin man ($j = 1$), the foreign origin man ($j = 2$), the local origin woman ($j = 3$) and the foreign origin woman ($j = 4$). Let $\delta_O < 0$ measure discrimination against the foreign origin candidates and $\delta_G < 0$ measure gender discrimination. The first candidate should not be discriminated and his callback probability should be of the form:

$$m_1 = \theta$$

where $\theta$ is labour market tightness. For the local origin woman, we can add a gender discrimination component :

$$m_2 = \theta + \delta_G$$

and for the foreign orign man, we should add an origin discrimination component:

$$m_3 = \theta + \delta_O.$$

For the last candidate, we could think of adding the two discriminatory components and to add a joint component, know as intersectionality in the literature (Tariq and Syed, 2017). Denote it $\delta_{OG}$. If $\delta_{OG} = 0$, discrimination is additive, a foreign origin woman will have a discrimination coefficient equal to the sum of the gender discrimination coefficient and of the foreign origin discrimination coefficient. If the coefficient is negative, the discrimination will be stronger, discrimination is said to be superadditive and if $\delta_{OG} > 0$ the discrimination of the sum will be less strong that the sum of discriminations and is said to be subadditive.

$$m_4 = \theta + \delta_G + \delta_O + \delta_{OG}.$$

[13]The method can be adapted to any number of candidates, provided that at least two candidates are sent on each offer.

The $\theta$ parameter does not measure discrimination so that its estimation is not important in this context. We will focus on the $\delta$ parameters. Taking the model in differences, we get:

$$\begin{aligned} m_2 - m_1 &= \delta_G \\ m_3 - m_1 &= \delta_O \\ m_4 - m_3 - (m_2 - m_1) &= \delta_{OG} \end{aligned}$$

we get gender discrimination by comparing the two local candidates since they do not face origin discrimination, the origin discrimination is obtained by comparing the two men since they do not face gender discrimination, and the joint discrimination by comparing gender discrimination among the foreign origin candidates with gender discrimination among the local origin candidates. Writing a components model allow to determine the right comparisons easily. It also works when there are less parameters to estimate than probability differences, as we show later in the chapter.

**A discrete choice model.** We model the probability of a callback. For any candidate $j$ on job ad $i$ we let $v^*_{ji}$ be the recruiter's gain associated with a callback:

$$v^*_{ji} = m_{ji} + \alpha_i + \varepsilon_{ji} \tag{3}$$

where $m_{ji}$ is the model for candidate $j$ on job $i$. Its form depends on each experiment and includes the discriminatory components that we wish to estimate. The $\alpha_i$ term is the job ad correlated effect (or "fixed effect"), since the same recruiter replies to all the candidates and $\varepsilon_{ji}$ is the idiosyncratic error term, typically a white noise. We observe the callback dummy :

$$v_{ji} = \begin{cases} 1 & \text{if } v^*_{ji} > 0 \\ 0 & \text{otherwise} \end{cases}$$

It equals 1 when recruiter calls the candidate back, and zero otherwise. We wish to estimate the model from a sample of dummy variables and the characteristics of the candidate, chosen by the researchers who ran the experiment. First, we need to eliminate the unobserved heterogeneity term $\alpha_i$. Let $F_\varepsilon$ be the c.d.f. of $\varepsilon$, we get the theoretical callback probability:

$$P_{ji} = \Pr(v_{ji} = 1) = \Pr(v^*_{ji} > 0) = 1 - F_\varepsilon\left(-(m_{ji} + \alpha_i)\right).$$

These probabilities have empirical counterparts and, with an assumption on $F_\varepsilon$, we can estimate the model components. Notice that the fit of several distributions can be compared with our method. In order to eliminate the $\alpha_i$ terms, we need to compare the answers to two candidates on the same job ad. Let $j = 1$ be a freely chosen reference candidate, with no loss of generality, we eliminate $\alpha_i$ with the following differencing:

$$D_i(j,1) = F_\varepsilon^{-1}(1 - P_{1i}) - F_\varepsilon^{-1}(1 - P_{ji}) = m_{ji} - m_{1i}.$$

By definition of the callback probabilities, the difference $m_{ji} - m_{1i}$ term contains the discrimination terms that we wish to estimate. Simplification occurs when $\varepsilon$ is assumed to have a symmetric distribution. In this case we get:[14]

$$P_{ji} = F_\varepsilon\left(m_{ji} + \alpha_i\right)$$

[14] The method can be applied without this assumption.

and we can take the difference:

$$\Delta_i(j,1) = F_\varepsilon^{-1}(P_{ji}) - F_\varepsilon^{-1}(P_{1i}) = m_{ji} - m_{1i}.$$

Three well-known cases are worth commenting. First, the default case of correspondence studies is the linear probability model, which leads to a direct comparison of the callback probabilities. Assuming a uniform distribution, $F_\varepsilon(\varepsilon) = \varepsilon$, we get:

$$\Delta_i(j,1) = P_{ji} - P_{1i}.$$

and the coefficients can be interpreted as percentage points. Another case encountered is the Logit model. It has the advantage to constrain the estimated probabilities in the $[0,1]$ interval. Assuming a logistic distribution, $F_\varepsilon(\varepsilon) = 1/(1+\exp(-\varepsilon))$, we must take the difference of the log odds ratios of the two candidates:

$$\Delta_i(j,1) = \ln\frac{P_{ji}}{1-P_{ji}} - \ln\frac{P_{1i}}{1-P_{1i}}.$$

and the coefficients are to be interpreted as log-odds ratios. Finally, with the Normit/Probit model, we get :

$$\Delta_i(j,1) = \Phi^{-1}(P_{ji}) - \Phi^{-1}(P_{1i})$$

where $\Phi$ is the cdf of the standard normal distribution, and the coefficients are more difficult to interpret than in the two previous cases. Now that the unobserved heterogeneity term has been eliminated, we can discuss the identification of the discriminatory components.

Consider the example of gender and origin. We would like to compare the four candidates of our previous example : local origin man ($j=1$), local origin woman ($j=2$), foreign origin man ($j=3$), foreign origin women ($j=4$). We had :

$$\begin{aligned}
m_1 &= \theta\\
m_2 &= \theta + \delta_G\\
m_3 &= \theta + \delta_O\\
m_4 &= \theta + \delta_G + \delta_O.
\end{aligned}$$

For the clarity of the exposition, we impose $\delta_{OG} = 0$ in order to get an overidentified case (i.e., more probability differences that discrimination parameters). Overidentified cases happen easily when several discrimination items are allowed for. The consequence of it is that there are now several ways to estimate the discrimination parameters. We get the three following differences:

$$\begin{aligned}
\Delta(2,1) &= m_2 - m_1 = \delta_G\\
\Delta(3,1) &= m_3 - m_1 = \delta_O\\
\Delta(4,1) &= m_4 - m_1 = \delta_G + \delta_O
\end{aligned}$$

Since the left-hand of this system has an empirical counterpart, we should estimate our two parameters from three statistics. We system is overidentified, there are more statistics than needed. In order to proceed to the estimation, we should use a minimum distance estimation method and test the validity of the restriction: $\Delta(2,1) + \Delta(3,1) =$

$\Delta(4,1)$. It is an additivity property. Let us rewrite our constraints:

$$\underbrace{\begin{pmatrix} \Delta(2,1) \\ \Delta(3,1) \\ \Delta(4,1) \end{pmatrix}}_{\pi} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} \delta_G \\ \delta_O \end{pmatrix}}_{\beta}$$

or

$$\pi = C\beta$$

The auxiliary parameter $\pi$ is easily estimated from the data, $\beta$ is the interest parameter, and is not directly observable. In order to estimate $\beta$, we need to replace $\pi$ with an estimate $\hat{\pi}$. Let:

$$\hat{\pi} = \pi + \omega$$

where $\omega$ is the estimation error on the auxiliary parameter. Substituting into the identification constraints, we get an equation that can be used for estimation:

$$\hat{\pi} = C\beta + \omega$$

where $\hat{\pi}$ and $C$ are observable, so that a minimum distance estimation is feasible. Let $\Omega = \mathrm{V}(\omega)$, its diagonal elements are the variances of the auxiliary parameters estimators, and the off diagonal term, the covariance between the estimators. They are correlated because the answers to all the candidates come from the same recruiter. The optimal estimator of $\beta$ is the Feasible Generalized Least Squares (FGLS) estimator:

$$\hat{\beta} = (C'\hat{\Omega}^{-1}C)^{-1}C'\hat{\Omega}^{-1}\hat{\pi}$$

It is asymptotically normal and its asymptotic covariance matrix can be estimated by the following statistic:

$$\mathrm{V}(\hat{\beta}) = (C'\hat{\Omega}^{-1}C)^{-1}$$

where $\hat{\Omega}$ is a consistent estimate of $\Omega$. The overidentification statistic, denoted $S$, is simply an estimate of the norm on the identification constraints, we get:

$$S = \hat{\omega}'\hat{\Omega}^{-1}\hat{\omega}$$

with $\hat{\omega} = \hat{\pi} - C\hat{\beta}$. Under the null hypothesis ($\mathrm{H}_0 : \pi = C\beta$), it is $\chi^2(1)$ distributed. More generally, for an overidentified system, the degrees of freedom equal the difference between the number of auxiliary parameters and the number of structural parameters. This statistic or its p-value can be used as a choice rule for $F_\varepsilon$. Indeed, $\pi$ depend on the callback probabilities and on the specific functional form $F_\varepsilon$. Taking the distribution with the highest p-value is therefore equivalent to take the distribution which fits the best the identification constraints.

**Rotating candidates.** In order to avoid detection, it is possible to send only a part of the candidate each time. For the simplicity of exposition, suppose that we send the benchmark candidate ($j = 1$) and only one of the other candidates. We could still compute estimates of the $\Delta_j$ and apply the method. Notice that we do not even need that the number of ads is the same for all the candidate. In fact, we do not even need to send the benchmark candidate all the time, but just two candidates, because what matters is to relate the auxiliary parameters to the structural parameters. With four candidates, there are 6 possible differences :

$$\begin{aligned}
\Delta(2,1) &= m_2 - m_1 = \delta_G \\
\Delta(3,1) &= m_3 - m_1 = \delta_O \\
\Delta(4,1) &= m_4 - m_1 = \delta_G + \delta_O \\
\Delta(3,2) &= m_3 - m_2 = \delta_O - \delta_G \\
\Delta(4,2) &= m_4 - m_2 = \delta_O \\
\Delta(4,3) &= m_4 - m_3 = \delta_G
\end{aligned}$$

and any combination involving the four candidates will provide an estimate provided that the $A$ matrix is adapted. For instance, if we use $\Delta(2,1)$, $\Delta(3,2)$ and $\Delta(4,3)$, we simply need to write :

$$\underbrace{\begin{pmatrix} \Delta(2,1) \\ \Delta(3,2) \\ \Delta(4,3) \end{pmatrix}}_{\pi} = \underbrace{\begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 1 & 0 \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} \delta_G \\ \delta_O \end{pmatrix}}_{\beta}$$

and proceed as before.

**Ad dependent model.** In some cases, the model may depend on an ad characteristics, like the contract length. The method will be the same but this time, there will be one equation for each candidate, depending on they reply to short term or long term contracts. The same method is applied with more equations.

**Backward selection.** When a component is not significant, a new estimation should be made in order to reduce the variance of the remaining estimators. The variance is lower because more observations are used to estimate the remaining parameters. In order to illustrate this point, consider the following system:

$$\begin{aligned}
\Delta(2,1) &= m_2 - m_1 = \delta_G \\
\Delta(4,1) &= m_4 - m_1 = \delta_G + \delta_O
\end{aligned}$$

If, say, $\delta_O = 0$, the system becomes:

$$\Delta(2,1) = m_2 - m_1 = \delta_G \tag{4}$$

$$\Delta(4,1) = m_4 - m_1 = \delta_G \tag{5}$$

and, assuming that each candidate is sent on all ads, we have twice more observations to estimate $\delta_G$. Now, we can use both $\Delta(2,1)$ and $\Delta(4,1)$. A standard solution is:[15]

$$\frac{1}{2}\Delta(2,1) + \frac{1}{2}\Delta(4,1) = \delta_G \tag{6}$$

and the estimation proceeds as usual, with a change of the left-hand variable. This also shows that process of backward selection is very different from the standard OLS case. For a more detailed presentation, see Duguet et al. (2018).

[15] More generally, the weight of each difference is proportional to its number of observations.

**Rotation of the candidates.** A rotation of the candidate simply reduces the number of observations available for each difference. Therefore, the empirical probabilities are computed as usual, and the method can be applied without any change. This remark is also valid when the rotation is not balanced among the candidates. The only constraint is to send at least two candidates on each ad, in order to compute a difference of treatment.

## 1.5 Structural estimation

The previous analyses are often thought of as subject to the Heckman critique (Heckman, 1998). First, all audit studies adjust as much characteristics as they can. However, there may be other characteristics, related to productivity and with a different distribution in the two groups. In this case, the results can be biased. Second, recruiters may apply different rules to different candidates. People from a discriminated group may be applied a higher standard for hiring. Neumark (2012) proposes an answer to the first Heckman critique. Consider two groups $j = A, B$ where the group $B$ is subject to discrimination. Let us consider that the productivity of the workers depend on two sets of variables $\tilde{X} = (X, \varepsilon)$, where $X$ is controlled in the experiment and $\varepsilon$ is not. Finally, let $\alpha$ be a summary of firm-level characteristics. The productivity of a worker is denoted $y(\tilde{X}, \alpha)$. Let $T^*$ be a function denoting the outcome of the worker in the labour market, like a callback rate (the "treatment" of the worker). There is discrimination when two identical workers are not treated equally:

$$T^*(y(\tilde{X}, \alpha), A) \neq T^*(y(\tilde{X}, \alpha), B)$$

Assuming a linear model, we can write:

$$y(X, \alpha) = Xb + \varepsilon + \alpha$$

and the treatment:

$$T^*(y(X, \alpha), d_B) = y + d \times d_B$$

with $d_B = 1$ then the worker belongs to group $B$ and 0 when s-he belongs to group $A$. The parameter $d < 0$ measures discrimination since it applies only to the $B$ group, while it is not related to $\tilde{X}$ or $\alpha$. The expected productivities in the two groups are denoted $y_j^*$. The experiment send two candidates $j = A, B$ and records their treatment by the recruiter:

$$T^*(y_B^*, 1) - T^*(y_A^*, 0) = y_B^* - y_A^* + d.$$

The goal of correspondence tests is to set $y_A^* = y_B^*$, so that the treatment difference reflects discrimination, measured by $d$. It can be obtained by an OLS regression of $T^*$ on $d_B$ and a constant term. Now consider the detailed list of variables influencing productivity $\tilde{X} = (X, \varepsilon)$. The first variable $X$ is controlled in the experement, but the second variable $\varepsilon$ is not. Let $\tilde{X}_j$ denote their values for the candidate $j$. A tested pair of candidates $(A, B)$ should verify:

$$y_A^* = X_A b + \varepsilon_A + \alpha$$
$$y_B^* = X_B b + \varepsilon_B + \alpha$$

so that the difference of treatment equals:

$$\begin{aligned} T^*(y_B^*, 1) - T^*(y_A^*, 0) &= X_B b + \varepsilon_B + \alpha + d - (X_A b + \varepsilon_A + \alpha) \\ &= \varepsilon_B - \varepsilon_A + d. \end{aligned}$$

when the correspondence test imposes $X_A = X_B$. The hypothesis $\mathrm{E}(\varepsilon_A) = \mathrm{E}(\varepsilon_B)$ guarantees an unbiased estimation of $d$. But this is not the end of the argument, since there is also a variance issue.[16]

Let us assume that a callback will be made if the expected productivity (or utility) crosses a given threshold. A recruiter may favour the group with the higher variance because it has a larger probability to cross the threshold $c$. More precisely, consider the binary treatment $T$:

$$T(y(\tilde{X}), d_B) = \begin{cases} 1 & \text{if } T^*(y(\tilde{X}), d_B) > c \\ 0 & \text{otherwise} \end{cases}$$

Consider a correspondence test with $X = X_A = X_B$, the discriminated group will receive a callback when $Xb + \varepsilon_B + d + \alpha > c$ while the non discriminated group will receive a callback when $Xb + \varepsilon_A + \alpha > c$. Assuming that $\alpha \sim \mathrm{N}(0, s_\alpha^2)$ and $\varepsilon_j \sim \mathrm{N}(0, \sigma_j^2)$ are independent, then $u_j = \alpha + \varepsilon_j \rightsquigarrow \mathrm{N}(0, \sigma_j^2)$ with $\sigma_j^2 = s_j^2 + \sigma_\alpha^2$. We get the following callback probabilities:

$$\begin{aligned} p_A &= \Pr\left[T^*(y(X, u_j), 0) > c\right] \\ &= \Pr\left[Xb + u_A > c\right] \\ &= \Phi\left(\frac{Xb - c}{\sigma_A}\right) \end{aligned} \tag{7}$$

where $\Phi$ is the cdf of the standard normal distribution, and similarly :

$$\begin{aligned} p_B &= \Pr\left[T^*(y(X, u_j), 1) > c\right] \\ &= \Pr\left[Xb + u_B + d > c\right] \\ &= \Phi\left(\frac{Xb + d - c}{\sigma_B}\right) \end{aligned} \tag{8}$$

We get the following callback difference, which is currently used for measuring discrimination:

$$p_A - p_B = \Phi\left(\frac{Xb - c}{\sigma_A}\right) - \Phi\left(\frac{Xb + d - c}{\sigma_B}\right)$$

Therefore, even in the absence of discrimination, $d = 0$, we can have different callback probabilities when $\sigma_A \neq \sigma_B$. Assuming $\sigma_A < \sigma_B$ would imply $p_A > p_B$ in the absence of discrimination. In order to detect discrimination, we would like to estimate the parameter $d$. If we estimate the callback probability models in the two groups, we would get the following:[17]

$$\begin{aligned} p_A &= \Phi(c_A + X b_A) \\ p_B &= \Phi(c_B + X b_B) \end{aligned} \tag{9}$$

[16] The focus on the two first moments is related to the normal distribution used later, since a normal distribution is fully defined by its two first moments.

[17] It is also possible to estimate a heteroskedastic Probit model, like in Neumark (2012). Notice that separate Probit estimates insures heteroskedasticity but deserves additional computations in order to get the joint covariance matrix of both candidates estimates.

the previous parameters $\pi' = (c_A, b'_A, c_B, b'_B)$ are called the auxiliary parameters because they do not always have a direct interpretation. We simply obtain them directly when we estimate the callback probabilities. Their value is to get the interest parameters. An obvious example of interest parameter is the $d$ parameter, which measures discrimination. Another parameter is the ratio of the standard errors $\psi = \sigma_B/\sigma_A$. We let $\beta' = (\psi, d)$ be the vector of interest parameters. Since the Probit model coefficients are identified up to a positive number, we will set $\sigma_A = 1$ without loss of generality. Using Using (7), (8) and (9), we get the following constraints:

$$c_A = \psi c_B - d, \; b_A = \psi b_B.$$

that we can rewrite in the following manner:

$$\underbrace{\begin{pmatrix} c_A \\ b_A \end{pmatrix}}_{\pi_A} = \underbrace{\begin{pmatrix} c_B & -1 \\ b_B & 0 \end{pmatrix}}_{C(\pi_B)} \underbrace{\begin{pmatrix} \psi \\ d \end{pmatrix}}_{\beta}$$

and the estimation can proceed with asymptotic least squares (ALS), which was originally developed in Chamberlain (1982), Chamberlain (1984) and Gouriéroux et al. (1985). More precisely, $\pi_A$ and $A(\pi_B)$ have empirical counterparts, so that a consistent and asymptotically normal estimator can be obtained for $\beta$. Also notice that $b_A$ and $b_B$ may be vectors so that the estimation problem will generally be overidentified. The implicit assumption with this method is that there should be at least one variable in $X$ that has a monotonic effect on the callback rate (positive or negative). Replacing $\pi$ by an estimate $\hat{\pi}$, we get:

$$\hat{\pi}_A = C(\hat{\pi}_B)\beta + \omega$$

where $\omega$ is an error term created by the replacement of $\pi$ by $\hat{\pi}$. We obtain a first step estimator by performing OLS on this relationship and get:

$$\hat{\beta} = (C(\hat{\pi}_B)' C(\hat{\pi}_B))^{-1} C'(\hat{\pi}_B)\hat{\pi}_A$$

Letting $\omega = g(\hat{\pi}, \beta) = \hat{\pi}_A - C(\hat{\pi}_B)\beta$, its asymptotic variance equals:

$$\begin{aligned} \Omega(\hat{\pi}, \beta) &= \mathrm{V}(\omega) \\ &= \frac{\partial g}{\partial \pi'}(\hat{\pi}, \beta)\mathrm{V}(\hat{\pi})\left(\frac{\partial g}{\partial \pi'}(\hat{\pi}, \beta)\right)' \end{aligned}$$

with

$$\frac{\partial g}{\partial \pi'}(\pi, \beta) = \begin{pmatrix} 1 & 0_{(1,k)} & -\psi & 0_{(1,k)} \\ 0_{(k,1)} & \mathrm{Id}_k & 0_{(k,1)} & -\psi \mathrm{Id}_k \end{pmatrix}$$

where $k$ is the number of regressors in $X$ (excluding the constant term). Replacing $\psi$ by the $\hat{\psi}$ obtained in the first stage estimation. We get the FGLS estimator:

$$\beta^* = \left(C(\hat{\pi}_B)\Omega(\hat{\beta})^{-1} C(\hat{\pi}_B)\right)^{-1} C(\hat{\pi}_B)\Omega(\hat{\beta})^{-1}\hat{\pi}_A$$

with the asymptotic covariance matrix:

$$\hat{\mathrm{V}}(\beta^*) = \left(C(\hat{\pi}_B)\Omega(\beta^*)^{-1} C(\hat{\pi}_B)\right)^{-1}.$$

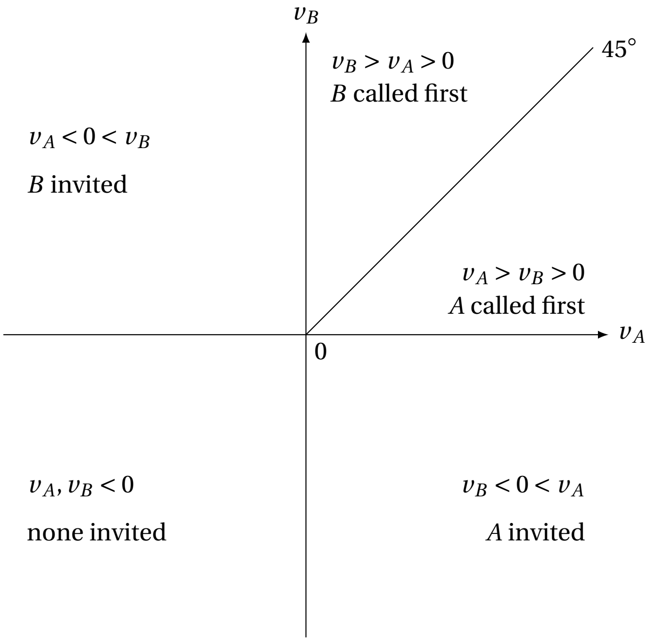


Figure 1: Callback decisions

### 1.6 Evaluation with ranks

When the data are collected not only about the answers to the candidates but also about their order of reception, it is possible to test for the existence of stronger forms of discrimination (Duguet et al., 2015). Consider a recruiter with preferences for the candidates $A$ and $B$ represented by the utilities $U_A$ and $U_B$. These utilities are specific to each recruiter and result from pre-conceptions about the candidates, because the candidates are equally productive by construction of the correspondence test experiment. Each recruiter has a reservation utility level $U_R$ above which the candidates are invited to an interview. We can define the relative utility levels $v_A = U_A - U_R$ and $v_B = U_B - U_R$. The four potential answering cases can be represented in the following way. If $v_A < 0$ and $v_B < 0$, no candidate is invited to an interview. When $v_A < 0 < v_B$, only candidate $B$ is invited; when $v_B < 0 < v_A$, only candidate $A$ is invited. Finally, when $v_A > 0$ and $v_B > 0$, both candidates are. These cases are illustrated in Figure 1. The standard measure of discrimination against candidate $B$, used in the literature, considers only cases in which only one of the two candidates is invited. These cases are illustrated by the North-West and the South-East quadrants of Figure 1. We denote this discrimination coefficient as $D(A,B)$:

$$\begin{aligned} D(A,B) &= \Pr[v_B < 0 < v_A] - \Pr[v_A < 0 < v_B] \\ &= \Pr[\text{A invited,B uninvited}] - \Pr[\text{B invited,A uninvited}] \end{aligned}$$

According to this measure, there is no discrimination when both candidates have equal chances to be invited, and a positive number indicates that candidate $A$ is, on average, preferred to candidate $B$.

It is possible to extend the standard measure of discrimination $D(A,B)$ to the ranking of the candidates when both are invited, which is equivalent to consider all the quadrants in Figure 1. In order to compare the rankings of two candidates, we use the concept of first order stochastic dominance. Suppose that we have J candidates, ranked according to the recruiter's utilities. The candidates that have not been invited satisfy the condition that $v_j = U_j - U_R < 0$. The ranking of the candidates (from 1st to Jth) results from the recruiter's tastes for the candidates. The highest utility corresponds to the candidate ranked first, and negative utilities correspond to the candidates that have not been invited. In order to perform our analysis, we need to separate the candidates that have not been invited from the others by creating a rank $\mathrm{J}+1$. This additional rank is required because the candidates that have not been called cannot be ranked between themselves. We only know that the uninvited candidates' utilities are below the recruiter's reservation utility levels and therefore that they are ranked behind the candidates that have been invited. Consider first the case for which all the candidates have been invited. Using the order statistic, denoted (), we obtain the ranking of the utilities of the candidates $\{j_1, j_2, \ldots, j_\mathrm{J}\}$, $0 < v_{(j_1)} \leq v_{(j_2)} \leq \cdots \leq v_{(j_\mathrm{J})}$, that corresponds to the ranking $\mathrm{J}, \mathrm{J}-1, \ldots, 1$. When only $k$ candidates are invited, we have the ranking $v_{(j_1)} \leq \ldots v_{(j_{\mathrm{J}-k})} < 0 < v_{(j_{\mathrm{J}-k+1})} \leq \cdots \leq v_{(j_\mathrm{J})}$ that corresponds to the ranking $\mathrm{J}+1, \ldots, \mathrm{J}+1, k, \ldots, 1$. The first stochastic dominance (henceforth, FOSD) of candidate $j_1$ over candidate $j_2$ is defined as:

$$\begin{aligned}&\Pr[v_{j_1} \geq v] \geq \Pr[v_{j_2} \geq v] \ \forall\, v,\\ &\qquad\qquad \text{and } \exists\ \bar{v} \text{ such that } \Pr[v_{j_1} \geq \bar{v}] > \Pr[v_{j_2} \geq \bar{v}],\end{aligned}$$

which means that candidate $j_1$ has a higher probability to reach a given utility level than candidate $j_2$, whatever the utility level is. This relationship is especially easy to interpret when $v$ is set at the reservation utility level of the recruiter, since it means that candidate $j_1$ has a higher probability to be invited to the interview than candidate $j_2$. This is the standard discrimination measure. We also see that FOSD covers more cases than the standard discrimination measure because it makes use of all possible utility thresholds.

For practical reasons we work with the ranks, since they are observable while the utilities are not. We just need to reverse the inequalities inside the probabilities, since the higher the utility the lower the rank (rank 1 for the most preferred candidate with utility $v_{(j_\mathrm{J})}$) :

$$\begin{aligned}&\Pr[r_{j_1} \leq r] \geq \Pr[r_{j_2} \leq r] \ \forall\, r \in \{1, \ldots, J+1\}\\ &\qquad\qquad \text{and } \exists\ \bar{r} \text{ such that } \Pr[r_{j_1} \leq \bar{r}] > \Pr[r_{j_2} \geq \bar{r}]. \qquad (10)\end{aligned}$$

Consider the case $r = 1$. Then $\Pr[r_j \leq 1] = \Pr[r_j = 1]$, which gives the probability to be ranked first. If the inequality (10) holds, the candidate $j_1$ has a higher probability to be ranked first than the candidate $j_2$. Now set $r = 2$. We conclude that candidate $j_1$ also has a higher probability to be ranked among the two first candidates than candidate $j_2$. Performing the comparisons up to $r = \mathrm{J}$ , $\Pr[r_j \leq \mathrm{J}]$ is the probability that candidate $j$ is invited to the interview. Therefore candidate $j_1$ has a higher probability to be invited to the interview than candidate $j_2$. In summary, when $j_1$ FOSD $j_2$, the candidate $j_1$ always has a higher probability to be in the preferred group than the candidate $j_2$, whatever the definition of the preferred group is. This definition is especially relevant for the measurement of discrimination, and this is what motivates the use of FOSD. Graphically, FOSD means that the CDF of candidate $j_1$ – defined on ranks– stands above the CDF of candidate $j_2$.

The implementation of the test is rather straightforward since the correspondence tests provide the empirical CDF directly. Consider the empirical distributions of the ranks for candidate $j$:

$$\hat{p}' = (\hat{p}_1(j), \ldots, \hat{p}_J(j))$$

and the corresponding empirical covariance matrix $\hat{\Sigma} = \hat{\mathrm{V}}(\hat{p})$. The CDF at rank $r$ is given by $\hat{P}_r = M_r' \hat{p}$, with

$$\underbrace{M_r}_{(1,J)} = (\underbrace{1, \ldots, 1}_{r \text{ times}}, \underbrace{0, \ldots, 0}_{J-r \text{ times}})$$

For example, with three ranks

$$\hat{p} = \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \end{pmatrix}$$

the cdf is given by:

$$\begin{aligned} M_1 \hat{p} &= (1,0,0)\,\hat{p} = \hat{p}_1 \\ M_2 \hat{p} &= (1,1,0)\,\hat{p} = \hat{p}_1 + \hat{p}_2 \\ M_3 \hat{p} &= (1,1,1)\,\hat{p} = \hat{p}_1 + \hat{p}_2 + \hat{p}_3 = 1 \end{aligned}$$

and the covariances are given by $\mathrm{V}(M_r \hat{p}) = M_r \hat{\Sigma} M_r'$. Now, we consider the distributions of two candidates, $j_1$ and $j_2$, and stack them in a $2 \times J$ vector $\hat{p}' = (\hat{p}'_{j_1}, \hat{p}'_{j_2})$, $\hat{\Sigma}$ be the associated joint $2J \times 2J$ covariance matrix, which accounts for the correlation between the ranks. Let $D_r = (M_r, -M_r)$, then the difference between the cumulative distribution functions of candidates $j_1$ and $j_2$ equals:

$$\begin{aligned} \Delta \hat{P}_r = D_r \hat{p} &= (M_r, -M_r) \begin{pmatrix} \hat{p}_{j_1} \\ \hat{p}_{j_2} \end{pmatrix} \\ &= M_r' \hat{p}_{j_1} - M_r' \hat{p}_{j_2} \\ &= \hat{P}_{j_1} - \hat{P}_{j_2} \end{aligned}$$

and the covariance matrix is obtained by $\mathrm{V}(D_r \hat{p}) = D_r \hat{\Sigma} D_r'$. Therefore, we can perform a Wald test at each point $r$ of the CDF. The following statistics is chi-squared distributed, with one degree of freedom under the null hypothesis $P_{j_1} = P_{j_2}$:

$$S_r = D_r \hat{p} (D_r \hat{\Sigma} D_r')^{-1} \hat{p}' D_r'.$$

# 2 The evaluation of wage discrimination

## 2.1 The wage regressions

**The Mincer equation** Many estimation method start with the wage equation (Mincer, 1958). Restricted to the people at work, for a sample of size $\mathbf{I}$, it can be written:

$$w_i = X_i b_1 + u_i,\ i \in \mathbf{I} \tag{11}$$

where $w_i = \ln W_i$ is the log-wage of worker $i$, $X_i$ the observable individual data, often called the *endowments*, like tenure, the years of schooling or the occupation, and $u_i$ a zero-mean disturbance, without loss of generality provided the model includes a constant term. Under standard assumptions this wage equation can be estimated by OLS. This regression will be used on the datasets where no information is available on the jobless people. It will be the case, for instance, with matched employer-employee data, since workers are sampled inside firms.

**The Heckman selection model.** Sometimes, more data are available. The labour force surveys generally include information both about the workers, the unemployed and the inactive people. In this case, we need a two-part model (Heckman, 1976). First, there is a labour market participation equation. A straightforward motivation is the reservation wage theory. If the offered wage is above the reservation wage, the worker takes the job and we observe the wage, otherwise we don't. This participation equation can be written:

$$d_i^* = H_i \gamma + v_i,$$

where $d_i^*$ is a latent, unobservable variable which summarizes the labour participation process. One may think of it as the difference between the offered wage and the reservation wage. The right-hand variables $H$ include the determinants of participation. The household composition and the replacement revenues may be important determinants. $v_i$ is a disturbance. We observe the following binary variable, which equals 1 when $i$ participates in the labour market, 0 otherwise :

$$d_i = \begin{cases} 1 & \text{if } d_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

This equation will be estimated with a binary model (Probit or other). The important point is that we observe the value of the wage when there is a participation only. Therefore, if there is a correlation between the disturbances $(u_i, v_i)$, OLS will suffer from a selection bias. Several solutions can be proposed to this problem. The first is the joint estimation of the model by maximum likelihood. Assuming the joint normality of the disturbances:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim \mathrm{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right)$$

where $\sigma^2$ if the variance of the disturbance in the wage equation and $\rho$ is the linear correlation coefficient of the two disturbances. Notice that the variance of the participation equation has been normalised to 1, like in a Probit model. Under this assumption, we obtain a Tobit 2 model in the terminology of Amemiya (1985). SAS and R typically allow for the estimation of this two-equation model.[18] The second, more popular method, is the Heckman method (Heckman, 1976). It relies on the following conditional expectation:[19]

$$\mathrm{E}(w_i | X_i, H_i, d_i = 1) = X_i b + \sigma\rho \frac{\varphi(H_i \gamma)}{\Phi(H_i \gamma)}.$$

where $\varphi$ and $\Phi$ are respectively the pdf and the cdf of the standard normal distribution. Ignoring the sample selection $d_i = 1$ can be interpreted as a missing regressor problem.

[18] SAS with the QLIM procedure and R with the sampleSelection package.

[19] Notice that when the right-hand variables are the same in the two equations, the identification proceeds from a functional form assumption only. It is better to use an exclusion restriction, with at least one variable present in the Probit equations and not in the wage equation (see Olsen (1980) for a discussion).

The estimation can proceed in two step. First, one estimates $\gamma$ with a Probit model and computes the estimated inverse of Mills ratio:

$$\hat{\lambda}_i = \frac{\varphi(H_i \hat{\gamma})}{\Phi(H_i \hat{\gamma})}$$

where $\hat{c}$ is the maximum likelihood estimator of the Probit model. In a second step, we perform the OLS regression of $w_i$ on the variables $(X_i, \hat{\lambda}_i)$, on the selected sample $(d_i = 1)$. Let $\theta = \sigma\rho$, we get $\hat{b}$ and $\hat{\theta}$ in this second step. The covariance matrix of these estimators should be computed according to Heckman (1979) because one regressor is estimated. See also Murphy and Topel (2002). In some cases, we need the selection-corrected wage:

$$\tilde{y}_{1i} = y_{1i} - \hat{\theta}\hat{\lambda}_i$$

In practise, one can combine the two methods, and use the Tobit 2 maximum likelihood estimate of $\gamma$ instead of the Probit estimate.[20]

## 2.2 The Oaxaca-Blinder decomposition

Labour market samples typically exhibit average wage differences between groups (men and women, say). A part of these differences may be justified by objective differences between the workers. In one group, workers may be older or with a longer education. This will create wage differences we can explain. The Oaxaca-Blinder decomposition is a method for separating the average wage differences attributable to the observed characteristics of the workers, from the wage differences attributable to a group membership (Oaxaca (1973), Blinder (1973)).

**Presentation.** Let us consider two groups of workers, $A$ and $B$. Following Becker (Becker (1971), p. 22), we can define a market discrimination coefficient (MDC) by comparing the observed wage $(W_A, W_B)$ and the theoretical wages without discrimination $(W_A^*, W_B^*)$:

$$\text{MDC} = \frac{W_A}{W_B} - \frac{W_A^*}{W_B^*}$$

In the previous expression, $W_A^*/W_B^*$ represents the wage difference in the absence of discrimination. This difference comes from the productivity differences of the workers and are therefore justified. The expression $W_A/W_B$ represents the real world wage difference. It includes both the justified wage difference and the unjustified one, which we refer to as wage discrimination. Since the Mincer equations refer to a wage in logarithm, we will prefer to work with the log-wage, denoted $w_j = \ln W_j,\ j \in \{A, B\}$. We will use the following approximation:[21]

$$\begin{aligned}
\text{MDC} &= \frac{W_A - W_B}{W_B} - \frac{W_A^* - W_B^*}{W_B^*} \\
&\simeq \ln\left(1 + \frac{W_A - W_B}{W_B}\right) - \ln\left(1 + \frac{W_A^* - W_B^*}{W_B^*}\right) \\
&= \ln W_A - \ln W_B - (\ln W_A^* - \ln W_B^*) \\
&= w_A - w_B - (w_A^* - w_B^*)
\end{aligned}$$

[20] The Tobit 2 estimator should reaches the Fréchet-Darmois-Cramer-Rao asymptotic variance lower bound when the distribution is bivariate normal.

[21] $\ln(1 + x) \simeq x$, for $x \simeq 0$.

Suppose that we have two groups of workers, $A$ and $B$, with respective wages $w_A$ and $w_B$. The difference of their wages may reflect any difference between them, including but not restricted to, wage discrimination. Suppose that we find a method to compute the non discriminatory wages, we denote them $w_A^*$ and $w_B^*$. Then we should measure wage discrimination against group $B$ by the following quantity:[22]

$$\text{MDC} = w_A - w_B - (w_A^* - w_B^*).$$

The justification is the following. Since $(w_A^*, w_B^*)$ are not discriminatory by assumption, the difference $w_A^* - w_B^*$ represent the *justified* wage difference. It comes from the productivity differences between the workers. Since the distribution of workers' productivity is not the same in the groups $A$ and $B$, there is a average wage difference. Now, consider $(w_A, w_B)$, they are the real-world wage distributions. By definition, they include the effect of all the differences among the two groups of workers, both justified and unjustified. Therefore the difference $D(A,B)$ measure the unjustified wage difference that the workers in the group $A$ benefit from (compared to the worker in the group $B$). We obtain this quantity as a "residual", that is as a *difference*. But this is not the definition of the residual that is used in econometrics. It is rather an accounting residual. The methods that we will present in this section simply differ by the definition that they take for $(w_A^*, w_B^*)$. Once a choice is set, a decomposition method follows.

**Method.** Let $b^*$ be the regression coefficient for the non discriminatory wage equation, we have the theoretical wage equation:

$$w^* = Xb^*$$

and all the differences of wages between the groups $A$ and $B$ come from differences in productive characteristics $X$. If the average characteristics of group $j$ is $X_j$, with $j \in \{A, B\}$, we get :

$$w_A^* - w_B^* = (X_A - X_B)b^*$$

applying this result at the mean point of the sample, we obtain the average wage difference that we should observe in the absence of discrimination:

$$\bar{w}_A^* - \bar{w}_B^* = (\bar{X}_A - \bar{X}_B)b^*.$$

Now let us consider discrimination. We clearly need the regression coefficient to differ between the two groups. What does this mean? That the same characteristics will not have the same return in the two groups any more. For instance, one year of experience will be less paid in group $B$ than in group $A$ for discriminatory reasons. This simply state that the average wage difference between two groups of identical workers must originate in a difference of the way their productive characteristics are paid. Now consider the two real-world regressions, obtained separately from the $A$ and $B$ groups. The observed wage is equal, by definition, to the sum of the predicted wage and the residual. We can write:

$$w_j = X_j \hat{b}_j + \hat{u}_j$$

[22] By convention, a positive difference will indicate wage discrimination. The use the log-wages because it is the common practice in the literature.

with $\hat{b}_j = (X_j' X_j)^{-1} X_j' w_j$. Another property is that the mean of the residual is zero provided there is a constant term among the regressors. This allows for writing, with $j = A, B$:

$$\bar{w}_j = \bar{X}_j \hat{b}_j \text{ with } \bar{w}_j = \frac{1}{\mathbf{I}} \sum_{i \in \mathbf{I}} w_{j,i},\ \bar{X}_j = \frac{1}{\mathbf{I}} \sum_{i \in \mathbf{I}} X_{j,i}.$$

Therefore the difference between the groups $A$ and $B$ equals:

$$\bar{w}_A - \bar{w}_B = \bar{X}_A \hat{b}_A - \bar{X}_B \hat{b}_B$$

The discrimination coefficient is defined by:

$$\begin{aligned} D(A,B) &= \bar{w}_A - \bar{w}_B - (\bar{w}_A^* - \bar{w}_B^*) \\ &= \bar{X}_A \hat{b}_A - \bar{X}_B \hat{b}_B - (\bar{X}_A - \bar{X}_B) b^* \\ &= \bar{X}_A (\hat{b}_A - b^*) + \bar{X}_B (b^* - \hat{b}_B) \end{aligned}$$

and we get the definition in Oaxaca and Ransom (1994). This is equivalent to use the following decomposition of the average wage:

$$\bar{w}_A - \bar{w}_B = \bar{X}_A (\hat{b}_A - b^*) + \bar{X}_B (b^* - \hat{b}_B) + (\bar{X}_A - \bar{X}_B) b^* \tag{12}$$

Consider the case where $B$ is discriminated against. The first term $\bar{X}_A(\hat{b}_A - b^*)$ measures the nepotism in favour of group $A$ since it is an extra wage compared to the non discriminatory case. The term $\bar{X}_B(b^* - \hat{b}_B)$ represents, on the contrary, the revenue lost caused by discrimination. The last term represents the wage difference justified by the difference of observable characteristics. There remains to choose $b^*$, this leads to the "index number problem" (Oaxaca (1973)).

## 2.3 The index number problem

In the original paper, Oaxaca (1973) makes an application with $b^* = \hat{b}_A$ (the men wage structure, denoted $A$) and $b^* = \hat{b}_B$ (the women wage structure, denoted $B$), suggesting that this will give an interval for the non discriminatory wage structure. Choosing $\hat{b}_A$ is equivalent to assume that the group $A$, which is not discriminated, should be a good proxy for the non discriminatory wage structure. But, Theoretically, the situation if more complex since the suppression of discrimination in the labour market may change the whole wage structure, including the wage of the favoured group. The solution may lie somewhere between $\hat{b}_A$ and $\hat{b}_B$. For example, using the (men) $A$ wage structure in equation (12), we get the following decomposition :

$$\bar{w}_A - \bar{w}_B = \bar{X}_B (\hat{b}_A - \hat{b}_B) + (\bar{X}_A - \bar{X}_B) \hat{b}_A \tag{13}$$

and discrimination is measured by :

$$\text{MDC} = \bar{X}_B (\hat{b}_A - \hat{b}_B).$$

At the other end, we could have taken the women wage structure as a reference ($B$), it would give (with $b^* = \hat{b}_B$):[23]

$$\text{MDC} = \bar{X}_A (\hat{b}_A - \hat{b}_B).$$

[23]The reader should notice that the occupations are often gendered, so that women wages may reflect the non discriminatory wage structures in the predominantly female occupations.

the same coefficient difference is used, but they are now weighted according to women characteristics. There are obviously many ways to set $b^*$ and the index number problem can be interpreted as finding a sensible way of fixing $b^*$. The solutions adopted in the literature are surveyed in Oaxaca and Ransom (1994).

**Weighting.** A first way to fix the problem is to weight the two estimators. Reimers (1983) uses the median value, by taking $b^* = \frac{1}{2}(\hat{b}_A + \hat{b}_B)$, we get:

$$\text{MDC} = \frac{\bar{X}_A + \bar{X}_B}{2}(\hat{b}_A - \hat{b}_B)$$

so that the average of the mean characteristics in two groups in now used as a benchmark. This clearly leads to the idea of weighting the means according to the size of the populations $A$ and $B$, so as to get the sample mean of all workers as the benchmark. This was proposed in Cotton (1988). Let $\pi_A$ be the share of $A$ workers in the sample (and $\pi_B = 1 - \pi_A$ the share of B workers), we let $b^* = \pi_A \hat{b}_A + \pi_B \hat{b}_B$, so that:

$$\text{MDC} = \tilde{X}(\hat{b}_A - \hat{b}_B) \tag{14}$$

with

$$\tilde{X} = (1 - \pi_A)\bar{X}_A + \pi_A \bar{X}_B,\ 0 < f_A < 1.$$

Here the reader should notice that the weights are inverted: the weight $\pi_A$ is used to weight the group $B$ average. This is consistent with (13).

What weight should we use? Neumark (1988) proposes an elegant solution using the approach of Becker (1971). Let us assume that the employers maximise their utility. It depends on the profits, but not only. The composition of the labour force also matters for them. Let the utility function be:

$$U\left(\Pi(L_A, L_B), L_A, L_B\right) \text{ with } \Pi(L_A, L_B) = pf(L_A + L_B) - w_A L_A - w_B L_B$$

where $p$ is the output price and $f$ the production function. Notice that the sum of labour inputs only matters for production. This means that the two labour types have an identical productivity and should be considered as equivalent as far as production is concerned. Suppose that the employer likes group $A$ and not group $B$, we would typically have $\partial U / \partial L_A \geq 0$ and $\partial U / \partial L_B \leq 0$. Utility maximisation implies that:

$$\frac{\partial U}{\partial \Pi}\left(pf'(L) - w_j\right) + \frac{\partial U}{\partial L_j} = 0,\ j = A, B \tag{15}$$

which implies the standard result:

$$W_A - W_B = \left(\frac{\partial U}{\partial \Pi}\right)^{-1}\left(\frac{\partial U}{\partial L_A} - \frac{\partial U}{\partial L_B}\right) > 0.$$

The $A$ group will be better paid than the $B$ group when the employers have a taste for the $A$ group ($\partial U / \partial L_A \geq 0$) and a distaste for the $B$ group ($\partial U / \partial L_B \leq 0$). We also see that the discrimination coefficient is decreasing with the taste for profit ($\partial U / \partial \Pi > 0$). These equations can also be used to determine the non discriminatory wage, provided that we accept an additional assumption.

The identifying assumption is simply that the employer should value the proportion of $A$ workers, not their absolute number. This does not seem unreasonable. Technically, this means that the utility function is homogeneous of degree 0 in $(L_A, L_B)$.[24] Under this assumptions, we can apply the Euler theorem:[25]

$$L_A \frac{\partial U}{\partial L_A} + L_B \frac{\partial U}{\partial L_B} = 0.$$

Using this expression with (15) gives:

$$W^* = \frac{L_A}{L_A + L_B} \times W_A + \frac{L_B}{L_A + L_B} \times W_B$$

where $W^* = pf'(L)$ is the nominal productivity, which defines the non discriminatory nominal wage. In order to apply the standard methodology, we need $\ln(W^*)$ rather than $W^*$. One can use the following shortcut.[26] First, notice that:

$$W^* = W_B \left(1 + \pi_A \frac{W_A - W_B}{W_B}\right)$$

with $\pi_A = L_A/(L_A + L_B)$ the share of group $A$ workers. Taking the logarithm, we get:[27]

$$\ln W^* \simeq \ln W_B + \pi_A \frac{W_A - W_B}{W_B}$$

next, using $(W_A - W_B)/W_B \simeq \ln W_A - \ln W_B$, we obtain:

$$\ln W^* \simeq \pi_A \ln W_A + (1 - \pi_A) \ln W_B$$
$$\Leftrightarrow w^* \simeq \pi_A w_A + (1 - \pi_A) w_B$$

Inserting the wage equations in this expression, we get (14). Here, the reader should notice that one can use either the number of workers or the number of hours worked, the latter being better in a production function. In the case when there is more than one labour qualification, Neumark (1988) also proposes to estimate $w^*$ as the pooled OLS estimator, also proposed in Oaxaca and Ransom (1988). Stacking the wage regressions of groups $A$ and $B$, we get the following model:

$$\begin{pmatrix} w_A \\ w_B \end{pmatrix} = \begin{pmatrix} X_A \\ X_B \end{pmatrix} b + u$$

or $w = Xb + u$. Applying OLS, we get:[28]

$$\hat{b} = \Omega_A \hat{b}_A + (\mathrm{Id} - \Omega_A) \hat{b}_B$$

with $\Omega_A = (X'X)^{-1} X'_A X_A$. Notice that this estimator generally differs from Cotton (1988), since the weighting is based on the second order sample moments of the explanative variables rather than on first order sample moments.

[24] A function $U(L_A, L_B)$ is homogeneous of degree $k$ if $U(mL_A, mL_B) = m^k U(L_A, L_B)$. With $k = 0$, we get $U(mL_A, mL_B) = U(L_A, L_B)$. Letting $m = 1/(L_A + L_B)$ and $\pi_A = L_A/(L_A + L_B)$, we get $U(L_A, L_B) = U(\pi_A, 1 - \pi_A)$, so that the utility depends on the proportions of $A$ workers only.

[25] If $U(L_A, L_B)$ is homogeneous of degree $k$: $Ł_A \partial U/\partial L_A + L_B \partial U/\partial B = kU(L_A, L_B)$.

[26] For a discussion, see Neumark (1988).

[27] We use the approximation $\ln(1 + x) \simeq x$ for $x \simeq 0$. In the applications, $0 < \pi_A < 1$ and $(W_A - W_B)/W_B$ is relatively low so that the approximation is good.

[28] Here, we assume that $A$ and $B$ workers do not work in the same firms, so that $X'_A X_B = 0$.

## 2.4 Accounting for labour market participation

**Wage correction.** The previous estimators are applied on the observed wage data. The problem is that these estimates may suffer from a selection bias. This point has been studied in Duncan and Leigh (1980) and Reimers (1983). With this approach, the wage difference is simply corrected for the selection biases (one in each wage equation), according to the Heckman method. Considering the group $j = A, B$, we estimate the Probit parts of the two wage equations:[29]

$$d^*_{ji} = H_{ji}\gamma_j + v_{ji},\ d_{ji} = \begin{cases} 1 & \text{if } d^*_{ji} > 0 \\ 0 & \text{otherwise} \end{cases}$$

and get the coefficients $\hat{\gamma}_j$. This enables us to compute the inverses of the Mills ratios, $\hat{\lambda}_{ji}$ with :

$$\hat{\lambda}_{ji} = \frac{\varphi(H_{ji}\hat{\gamma}_j)}{\Phi(H_{ji}\hat{\gamma}_j)}$$

Then, we run the second stage regressions on the observed wages only ($d_{ji} = 1$):

$$w_{ji} = X_{ji} b_j + \theta_j \lambda_{ji} + u_{ji}.$$

Replacing $\lambda_{ji}$ with $\hat{\lambda}_{ji}$ and denoting the mean value as $\bar{\hat{\lambda}}_j = \mathrm{I}_j^{-1} \sum_{i \in \mathbf{I}_j} \hat{\lambda}_{ji}$, we get the regression coefficients $(\hat{b}_j, \hat{\theta}_j)$ and compute the corrected wages' difference as:

$$\bar{w}_A - \hat{\theta}_A \bar{\hat{\lambda}}_A - \left(\bar{w}_B - \hat{\theta}_B \bar{\hat{\lambda}}_B\right) = \bar{w}_A - \bar{w}_B - \left(\hat{\theta}_A \bar{\hat{\lambda}}_A - \hat{\theta}_B \bar{\hat{\lambda}}_B\right)$$

this difference is then decomposed according to the previous methods. However, the reader should notice that this quantity is not equal to the average wage difference so that the next approach may be preferred.

**Extended decompositions.** In fact, the selection term themselves could be given a discrimination interpretation since they are a part of the average wage decomposition (Neuman and Oaxaca, 2004). We just present one decomposition, which uses the group $A$ wage structure as a reference. An important point to notice is that the $\theta_j$ coefficients are considered as *not* reflecting discrimination and can, therefore, be kept different in the two groups when defining the non discriminatory wages. Let use define the function equal to the inverse Mills' ratio:

$$\lambda(x) = \frac{\varphi(x)}{\Phi(x)},$$

The average observed wages should be:

$$\bar{w}_j = \bar{X}_j \hat{b}_j + \hat{\theta}_j \overline{\lambda(H_{ji}\hat{\gamma}_j)}$$

where the bar denotes the sample mean, and the non discriminatory estimated wages are obtained using $b^*_j = \hat{b}_A$ and $\gamma^*_j = \hat{\gamma}_A$:

$$\bar{w}^*_j = \bar{X}_j \hat{b}_A + \hat{\theta}_j \overline{\lambda(H_{ji}\hat{\gamma}_A)}.$$

[29] It is also possible to estimate the from the Probit parts of Tobit 2 models.

Using the definition of the MDC, we get:

$$\begin{aligned}\text{MDC} &= \bar{w}_A - \bar{w}_B - \left(\bar{w}_A^* - \bar{w}_B^*\right) \\ &= \bar{X}_B(\hat{b}_A - \hat{b}_B) + \hat{\theta}_B\left(\overline{\lambda(H_{Bi}\hat{\gamma_A})} - \overline{\lambda(H_{Bi}\hat{\gamma_B})}\right)\end{aligned}$$

Discrimination has now two components: the classic difference of returns on the endowments, and the differences in average wages coming from the difference in participation, assumed to originate in discrimination. Two other components intervene in the wage decomposition: the endowments differences and the selection difference caused by the difference in the $\theta_j$'s. The decomposition can be rewritten:

$$\bar{w}_A - \bar{w}_B = \text{MDC} + \bar{w}_A^* - \bar{w}_B^*$$

so that we just need to decompose:

$$\bar{w}_A^* - \bar{w}_B^* = \underbrace{(\bar{X}_A - \bar{X}_B)\hat{b}_A + \hat{\theta}_A\left(\overline{\lambda(H_{Ai}\hat{\gamma_A})} - \overline{\lambda(H_{Bi}\hat{\gamma_A})}\right)}_{\text{endowments}} + \underbrace{(\hat{\theta}_A - \hat{\theta}_B)\overline{\lambda(H_{Bi}\hat{\gamma_A})}}_{\text{selection}}$$

We get a decomposition in three parts, where the Heckman correction has been allocated to discrimination, endowments' differences and selection's difference. It relies on the assumption that the group $A$ has the non discriminatory participation and wage structure, and that the $\theta_j$ terms do not reflect discrimination. One can easily change the reference wage structure if needed. The issue is different for the $\theta_j$ terms since it implies to choose a definition of what is discriminatory.

Neuman and Oaxaca (2004) develops this analysis, by considering the cases where the selection term includes some discriminatory components itself. Indeed, $\theta_j = \sigma_j \times \rho_j$, so that differences in wage variances or in correlation coefficients could be interpreted as discriminatory. With a maximum likelihood estimation of the Tobit 2 model, it is possible to have separate estimates of $(b_j, \sigma_j, \rho_j)$ and to use them to propose new decompositions. But this relies on sometimes strong assumption about what is discriminatory. [30]

## 2.5 The productivity approach

The previous approaches rely on the assumptions that the workers have the same productivity. Therefore, one can improve on the evaluations by including independent measures of productivity in the analysis. It is the contribution of Hellerstein et al. (1999). In this approach, the authors estimate a production function which depend on the workers' demographic characteristics (gender, race etc.) and compare their wages with their productivity.[31] The method requires additional data: one needs plant (or firm) level data in order to estimate a production function, along with the standard wage data. Consider first the production function. The output $Y$ is a function of capital $C$, materials $M$ and a labour aggregate $\tilde{L}$, which summarizes the effect of all the

[30] With the two-step method, one first get an estimate of the product $\sigma_j \rho_j$, and another step is needed to get an estimate of $\sigma_j$ from the residual variance. Using a Tobit 2 software is more convenient than the two-step approach in this case.

[31] In the case where the access to specific jobs is discriminatory, productivity differences may come from discrimination itself. Therefore, the method measures wage discrimination in a narrow sense.

labour types on production:

$$Y = f(C, M, \tilde{L}).$$

In the previous approach, all the labour types were considered as equivalent, so that $\tilde{L}$ was the sum of labour hours. Here, we accept that the labour productivity may differ. Let $L_j$ be the labour input of group $j \in \mathbf{J}$, $L = \sum_j L_j$ be the total labour input, $\pi_j = L_j / L$ the share of the labour input of group $j$ and $\phi_j$ be the productivity of group $j$, the labour aggregate is:

$$\tilde{L} = \sum_{j \in \mathbf{J}} \phi_j L_j = L \sum_{j \in \mathbf{J}} \phi_j \pi_j.$$

Restricting ourselves to two groups of workers, $j \in \{A, B\}$, we get:

$$\tilde{L} = L\left(\phi_A \pi_A + \phi_B \pi_B\right)$$

using $\pi_A = 1 - \pi_B$ and using the group $A$ productivity as the benchmark, $\phi_A = 1$, without loss of generality, we get the following expression:

$$\tilde{L} = L\left(1 + (\phi_B - 1)\pi_B\right)$$

and the production function:

$$Y = f\left(C, M, L(1 + (\phi_B - 1)\pi_B)\right).$$

with which we can estimate $\phi_B$ the (relative) marginal productivity of group $B$. Hellerstein et al. (1999) use a translog production function. In order to simplify the exposition, we will retain a Cobb-Douglas. Let:

$$Y = AC^{\alpha} M^{\gamma} \tilde{L}^{\beta} \tag{16}$$

where $A$ is the total factor productivity. Using (16), we can write:

$$\begin{aligned} \ln Y &= \ln A + \alpha \ln C + \gamma \ln M + \beta \ln\left(L(1 + (\phi_B - 1)\pi_B)\right) \\ &\simeq \ln A + \alpha \ln C + \gamma \ln M + \beta \ln L + \beta(\phi_B - 1)\pi_B \end{aligned}$$

and the estimation of this equation will provide a consistent estimate of $\phi_B$.[32] This productivity differential estimate will be compared to a wage differential obtained in the following way. Consider a plant-level equation. The plant are denoted $k \in \mathbf{K} = \{1, \ldots, \mathrm{K}\}$. Let $d_{ijk}$ be a dummy variable equal to 1 if the worker $i$ belongs to the group $j$ in plant $k$. The individual-level wage equation is defined as:

$$W_{ik} = \sum_{j \in \mathbf{J}} d_{ijk} W_j,$$

summing this equation for all the $N_k$ workers in the plant $k$ gives:

$$\begin{aligned} \sum_{i=1}^{N_k} W_{ik} &= \sum_{i=1}^{N_k} \sum_{j \in \mathbf{J}} d_{ijk} W_j \\ &= \sum_{j \in \mathbf{J}} W_j \sum_{i=1}^{N_k} d_{ijk} \\ &= \sum_{j \in \mathbf{J}} W_j L_{jk} \end{aligned}$$

[32] The use of the delta method is necessary to get a separate estimate of $\phi_B$.

where $L_{jk} = \sum_{i=1}^{N_k} d_{ijk}$ is the number of group $j$ workers in plant $k$. This equation is definitional: it says that the total wage bill in plant $k$ is the sum of the wages bills in the J groups of workers. Considering two groups $j \in \{A, B\}$, the total wage bill in plant $k$ is:

$$W_k = W_A L_{Ak} + W_B L_{Bk}$$

and the average wage in plant $k$ is obtained by dividing by $L_k$, the total number of workers in plant $k$:

$$\frac{W_k}{L_k} = W_A \pi_{Ak} + W_B \pi_{Bk}$$

where $\pi_{jk} = L_{jk}/L_k$. Denoting $\mu_B = W_B/W_A$ the wage ratio, we get:

$$\begin{aligned}\frac{W_k}{L_k} &= W_A \pi_{Ak} + W_B \pi_{Bk} \\ &= W_A \left(1 + (\mu_B - 1)\phi_{Bk}\right)\end{aligned}$$

taking logarithms:

$$\begin{aligned} w_k &= w_A + \ln\left(1 + (\mu_B - 1)\phi_{Bk}\right) \\ &\simeq w_A + (\mu_B - 1)\phi_{Bk}\end{aligned}$$

with $w_k = \ln(W_k/L_k)$. Since the productivity and wage equations are defined at the plant level, they can be estimated together on consistent employer-employee data. A comparison of $\phi_j$ and $\mu_j$ may provide some evidence of discrimination. Clearly, if $\phi_j > \mu_j$ there is wage discrimination, while the reverse case may indicate nepotism in favour of the group $j$ workers.

# References


AMEMIYA, T. (1985): Advanced econometrics, Harvard university press. 27

ARELLANO, M. (1987): "Practitioners' Corner: Computing robust standard errors for within-groups estimators," Oxford bulletin of Economics and Statistics, 49, 431–434. 5, 11

BAERT, S. (2017): "Hiring discrimination: An overview of (almost) all correspondence experiments since 2005," GLO discussion paper. 2

BECKER, G. S. (1957): The economics of discrimination, University of Chicago press. 2

——— (1971): The economics of discrimination, University of Chicago press, 2nd edition. 28, 31

BERTRAND, M. AND E. DUFLO (2017): Field experiments on discrimination, North Holland, chap. 8. 2

BLINDER, A. S. (1973): "Wage discrimination: reduced form and structural estimates," Journal of Human resources, 436–455. 28

CHAMBERLAIN, G. (1982): "Multivariate regression models for panel data," Journal of Econometrics, 18, 5–46. 23

——— (1984): "Panel data," in Handbook of Econometrics, ed. by Z. Griliches and M. D. Intriligator, North-Holland, vol. 2, 1247–1318. 23

COTTON, J. (1988): "On the decomposition of wage differentials," The review of economics and statistics, 236–243. 31, 32

DUGUET, E., L. DU PARQUET, Y. L'HORTY, AND P. PETIT (2015): "New Evidence of Ethnic and Gender discriminations in the French Labor Market using experimental data: A ranking extension of responses from correspondence tests," Annals of Economics and Statistics/Annales d'Économie et de Statistique, 21–39. 24

DUGUET, E., R. LE GALL, Y. L'HORTY, AND P. PETIT (2018): "How does labour market history influence the access to hiring interviews?" International Journal of Manpower, 39, 519–533. 16, 20

DUGUET, E. AND P. PETIT (2005): "Hiring discrimination in the French financial sector: An econometric analysis on field experiment data," Annals of Economics and Statistics/Annales d'Économie et de Statistique, 79–102. 13

DUNCAN, G. M. AND D. E. LEIGH (1980): "Wage determination in the union and nonunion sectors: A sample selectivity approach," ILR Review, 34, 24–34. 33

GOURIÉROUX, C., A. MONFORT, AND A. TROGNON (1985): "Moindres carrés asymptotiques," in Annales de l'INSEE, JSTOR, 91–122. 23

HECKMAN, J. J. (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," in Annals of Economic and Social Measurement, Volume 5, number 4, NBER, 475–492. 27

——— (1979): "Sample selection bias as a specification error," Econometrica: Journal of the econometric society, 153–161. 28

——— (1998): "Detecting discrimination," Journal of Economic Perspectives, 12, 101–116. 2

HELLERSTEIN, J. K., D. NEUMARK, AND K. R. TROSKE (1999): "Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations," Journal of labor economics, 17, 409–446. 34, 35

MINCER, J. (1958): "Investment in human capital and personal income distribution," Journal of political economy, 66, 281–302. 26

MURPHY, K. M. AND R. H. TOPEL (2002): "Estimation and inference in two-step econometric models," Journal of Business & Economic Statistics, 20, 88–97. 28

NEUMAN, S. AND R. L. OAXACA (2004): "Wage decompositions with selectivity-corrected wage equations: A methodological note," The Journal of Economic Inequality, 2, 3–10. 33, 34

NEUMARK, D. (1988): "Employers' discriminatory behavior and the estimation of wage discrimination," Journal of Human resources, 279–295. 31, 32

——— (2012): "Detecting discrimination in audit and correspondence studies," Journal of Human Resources, 47, 1128–1157. 15

——— (2018): "Experimental research on labor market discrimination," Journal of Economic Literature, 56, 799–866. 2

OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," International economic review, 693–709. 28, 30

OAXACA, R. L. AND M. R. RANSOM (1988): "Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers." Journal of Labor Research, 9, 139. 32

——— (1994): "On discrimination and the decomposition of wage differentials," Journal of econometrics, 61, 5–21. 30, 31

OLSEN, R. J. (1980): "A least squares correction for selectivity bias," Econometrica, 1815–1820. 27

REIMERS, C. W. (1983): "Labor market discrimination against Hispanic and black men," The review of economics and statistics, 570–579. 31, 33

RIACH, P. A. AND J. RICH (2002): "Field experiments of discrimination in the market place," The Economic Journal, 112, F480–F518. 3

RICH, J. (2014): "What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000," IZA discussion paper. 2

——— (2018): "Do photos help or hinder field experiments of discrimination?" International Journal of Manpower, 39, 502–518. 3

TARIQ, M. AND J. SYED (2017): "Intersectionality at Work: South Asian Muslim Women's Experiences of Employment and Leadership in the United Kingdom," Sex Roles, 77, 510–522. 16

VUONG, Q. H. (1989): "Likelihood ratio tests for model selection and non-nested hypotheses," Econometrica, 57, 307–333. 13

## TEPP Working Papers 2020

**20-6. The effects of age on educational performances at the end of primary school: cross-sectional and regression discontinuity approach applications from Reunion Island**
Daniel Rakotomalala

**20-5. Slowdown antitrust investigations by decentralization**
Emilie Dargaud, Armel Jacques

**20-4. Is international tourism responsible for the pandemic of COVID19? A preliminary cross-country analysis with a special focus on small islands**
Jean-François Hoarau

**20-3. Does labor income react more to income tax or means tested benefit reforms?**
Michaël Sicsic

**20-2. Optimal sickness benefits in a principal-agent model**
Sébastien Ménard

**20-1. The specific role of agriculture for economic vulnerability of small island spaces**
Stéphane Blancard, Maximin Bonnet, Jean-François Hoarau

## TEPP Working Papers 2019

**19-8. The impact of benefit sanctions on equilibrium wage dispersion and job vacancies**
Sebastien Menard

**19-7. Employment fluctuations, job polarization and non-standard work: Evidence from France and the US**
Olivier Charlot, Idriss Fontaine, Thepthida Sopraseuth

**19-6. Counterproductive hiring discrimination against women: Evidence from French correspondence test**
Emmanuel Duguet, Loïc du Parquet, Yannick L'Horty, Pascale Petit

**19-5. Inefficient couples: Non-minimization of the tax burden among French cohabiting couples**
Olivier Bargain, Damien Echevin, Nicolas Moreau, Adrien Pacifico

**19-4. Seeking for tipping point in the housing market: evidence from a field experiment**
Sylvain Chareyron, Samuel Gorohouna, Yannick L'Horty, Pascale Petit, Catherine Ris

**19-3. Testing for redlining in the labor market**
Yannick L'Horty, Mathieu Bunel, Pascale Petit

**19-2. Labour market flows: Accounting for the public sector**
Idriss Fontaine, Ismael Galvez-Iniesta, Pedro Gomes, Diego Vila-Martin

**19-1. The interaction between labour force participation of older men and their wife: lessons from France**
Idriss Fontaine

## TEPP Working Papers 2018

**18-15. Be healthy, be employed: a comparison between the US and France based on a general equilibrium model**
Xavier Fairise, François Langot, Ze Zhong Shang

**18-14. Immigrants' wage performance in the routine biased technological change era: France 1994-2012**
Catherine Laffineur, Eva Moreno-Galbis, Jeremy Tanguy, Ahmed Tritah

**18-13. Welfare cost of fluctuations when labor market search interacts with financial frictions**
Elini Iliopulos, François Langot, Thepthida Sopraseuth

**18-12. Accounting for labor gaps**
François Langot, Alessandra Pizzo

**18-11. Unemployment fluctuations over the life cycle**
Jean-Olivier Hairault, François Langot, Thepthida Sopraseuth

**18-10. Layoffs, Recalls and Experience Rating**
Julien Albertini, Xavier Fairise

**18-9. Environmental policy and health in the presence of labor market imperfections**
Xavier Pautrel

**18-8. Identity mistakes and the standard of proof**
Marie Obidzinski, Yves Oytana

**18-7. Presumption of innocence and deterrence**
Marie Obidzinski, Yves Oytana

**18-6. Ethnic Discrimination in Rental Housing Market: An Experiment in New Caledonia**
Mathieu Bunel, Samuel Gorohouna, Yannick L'Horty, Pascale Petit, Catherine Ris

**18-5. Evaluating the impact of firm tax credits. Results from the French natural experiment CICE**
Fabrice Gilles, Yannick L'Horty, Ferhat Mihoubi, Xi Yang

**18-4. Impact of type 2 diabetes on health expenditure: an estimation based on individual administrative data**
François-Olivier Baudot , Anne-Sophie Aguadé,Thomas Barnay, Christelle Gastaldi-Ménager, Anne Fargot-Campagna

**18-3. How does labour market history influence the access to hiring interviews?**
Emmanuel Duguet, Rémi Le Gall, Yannick L'Horty, Pascale Petit

**18-2. Occupational mobility and vocational training over the life cycle**
Anthony Terriau

**18-1. Retired, at last? The short-term impact of retirement on health status in France**
Thomas Barnay, Eric Defebvre

## TEPP Working Papers 2017

**17-11. Hiring discrimination against women: distinguishing taste based discrimination from statistical discrimination**
Emmanuel Duguet, Loïc du Parquet, Pascale Petit

**17-10. Pension reforms, older workers' employment and the role of job separation and finding rates in France**
Sarah Le Duigou, Pierre-Jean Messe

**17-9. Healthier when retiring earlier? Evidence from France**
Pierre-Jean Messe, François-Charles Wolff

**17-8. Revisting Hopenhayn and Nicolini's optimal unemployment insurance with job search monitoring and sanctions**
Sebastien Menard, Solenne Tanguy

**17-7. Ethnic Gaps in Educational Attainment and Labor-Market Outcomes: Evidence from France**
Gabin Langevin, David Masclet, Fabien Moizeau, Emmanuel Peterle

**17-6. Identifying preference-based discrimination in rental market: a field experiment in Paris**
Mathieu Bunel, Yannick L'Horty, Loïc du Parquet, Pascale Petit

**17-5. Chosen or Imposed? The location strategies of households**
Emilie Arnoult, Florent Sari

**17-4. Optimal income taxation with composition effects**
Laurence Jacquet, Etienne Lehmann

**17-3. Labor Market Effects of Urban Riots: an experimental assessment**
Emmanuel Duguet, David Gray, Yannick L'Horty, Loic du Parquet, Pascale Petit

**17-2. Does practicing literacy skills improve academic performance in first-year university students? Results from a randomized experiment**
Estelle Bellity, Fabrices Gilles, Yannick L'Horty

**17-1. Raising the take-up of social assistance benefits through a simple mailing: evidence from a French field experiment**
Sylvain Chareyron, David Gray, Yannick L'Horty

## TEPP Working Papers 2016

**16-8. Endogenous wage rigidities, human capital accumulation and growth**
Ahmed Tritah

**16-7. Harder, better, faster...yet stronger? Working conditions and self-declaration of chronic diseases**
Eric Defebvre

**16-6. The influence of mental health on job retention**
Thomas Barnay, Eric Defebvre

**16-5. The effects of breast cancer on individual labour market outcomes: an evaluation from an administrative panel**
Thomas Barnay, Mohamed Ali Ben Halima, Emmanuel Duguet, Christine Le Clainche, Camille Regaert

**16-4. Expectations, Loss Aversion, and Retirement Decisions in the Context of the 2009 Crisis in Europe**
Nicolas Sirven, Thomas Barnay

**16-3. How do product and labor market regulations affect aggregate employment, inequalities and job polarization? A general equilibrium approach**
Julien Albertini, Jean-Olivier Hairault, François Langot, Thepthida Sopraseuth

**16-2. Access to employment with age and gender: results of a controlled experiment**
Laetitia Challe, Florent Fremigacci, François Langot, Yannick L'Horty, Loïc Du Parquet, Pascale Petit

**16-1. An evaluation of the 1987 French Disabled Workers Act: Better paying than hiring**
Thomas Barnay, Emmanuel Duguet, Christine Le Clainche, Yann Videau

## TEPP Working Papers 2015

**15-10. Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach**
Kory Kroft, Kavan Kucko, Etienne Lehmann, Johannes Schmieder

**15-9. Search frictions and (in) efficient vocational training over the life-cycle**
Arnaud Chéron, Anthony Terriau

**15-8. Absenteeism and productivity: the experience rating applied to employer contributions to health insurance**
Sébastien Ménard, Coralia Quintero Rojas

**15-7. Take up of social assistance benefits: the case of homeless**
Sylvain Chareyron

**15-6. Spatial mismatch through local public employment agencies. Answers from a French quasi-experiment**
Mathieu Bunel, Elisabeth Tovar

**15-5. Transmission of vocational skills at the end of career: horizon effect and technological or organisational change**
Nathalie Greenan, Pierre-Jean Messe

**15-4. Protecting biodiversity by developing bio-jobs: A multi-branch analysis with an application on French data**
Jean De Beir, Céline Emond, Yannick L'Horty, Laetitia Tuffery

**15-3. Profit-Sharing and Wages: An Empirical Analysis Using French Data Between 2000 and 2007**
Noélie Delahaie, Richard Duhautois

**15-2. A meta-regression analysis on intergenerational transmission of education: publication bias and genuine empirical effect**
Nicolas Fleury, Fabrice Gilles

**15-1. Why are there so many long-term unemployed in Paris?**
Yannick L'Horty, Florent Sari

## TEPP Working Papers 2014

**14-14. Hiring discrimination based on national origin and the competition between employed and unemployed job seekers**
Guillaume Pierné

**14-13. Discrimination in Hiring: The curse of motorcycle women**
Loïc Du Parquet, Emmanuel Duguet, Yannick L'Horty, Pascale Petit

**14-12. Residential discrimination and the ethnic origin: An experimental assessment in the Paris suburbs**
Emmanuel Duguet, Yannick L'Horty, Pascale Petit

**14-11. Discrimination based on place of residence and access to employment**
Mathieu Bunel, Yannick L'Horty, Pascale Petit

**14-10. Rural Electrification and Household Labor Supply: Evidence from Nigeria**
Claire Salmon, Jeremy Tanguy

**14-9. Effects of immigration in frictional labor markets: theory and empirical evidence from EU countries**
Eva Moreno-Galbis, Ahmed Tritah

**14-8. Health, Work and Working Conditions: A Review of the European Economic Literature**
Thomas Barnay

**14-7. Labour mobility and the informal sector in Algeria: a cross-sectional comparison (2007-2012)**
Philippe Adair, Youghourta Bellache

**14-6. Does care to dependent elderly people living at home increase their mental health?**
Thomas Barnay, Sandrine Juin

**14-5. The Effect of Non-Work Related Health Events on Career Outcomes: An Evaluation in the French Labor Market**
Emmanuel Duguet, Christine le Clainche

**14-4. Retirement intentions in the presence of technological change: Theory and evidence from France**
Pierre-Jean Messe, Eva Moreno-Galbis, Francois-Charles Wolff

**14-3. Why is Old Workers' Labor Market more Volatile? Unemployment Fluctuations over the Life-Cycle**
Jean-Olivier Hairault, François Langot, Thepthida Sopraseuth

**14-2. Participation, Recruitment Selection, and the Minimum Wage**
Frédéric Gavrel

**14-1. Disparities in taking sick leave between sectors of activity in France: a longitudinal analysis of administrative data**
Thomas Barnay, Sandrine Juin, Renaud Legal

## TEPP Working Papers 2013

**13-9. An evaluation of the impact of industrial restructuring on individual human capital accumulation in France (1956-1993)**
Nicolas Fleury, Fabrice Gilles

**13-8. On the value of partial commitment for cooperative investment in buyer-supplier relationship**
José de Sousa, Xavier Fairise

**13-7. Search frictions, real wage rigidities and the optimal design of unemployment insurance**
Julien Albertini, Xavier Fairise

**13-6. Tax me if you can! Optimal nonlinear income tax between competing governments**
Etienne Lehmann, Laurent Simula, Alain Trannoy

**13-5. Beyond the labour income tax wedge: The unemployment-reducing effect of tax progressivity**
Etienne Lehmann, Claudio Lucifora, Simone Moriconi, Bruno Van Der Linden

**13-4. Discrimination based on place of residence and access to employment**
Mathieu Bunel, Emilia Ene Jones, Yannick L'Horty, Pascale Petit

**13-3. The determinants of job access channels: evidence from the youth labor market in Franc**
Jihan Ghrairi

**13-2. Capital mobility, search unemployment and labor market policies: The case of minimum wages**
Frédéric Gavrel

**13-1. Effort and monetary incentives in Nonprofit et For-Profit Organizations**
Joseph Lanfranchi, Mathieu Narcy

## The TEPP Institute

The CNRS **Institute for Theory and Evaluation of Public Policies** (the TEPP Institute, FR n°2024 CNRS) gathers together research centres specializing in economics and sociology:

- L'**Equipe de Recherche sur l'Utilisation des Données Individuelles en lien avec la Théorie Economique** (Research Team on Use of Individuals Data in connection with economic theory), **ERUDITE**, University of Paris-Est Créteil and University of Gustave Eiffel
- Le **Centre d'Etudes des Politiques Economiques de l'université d'Evry** (Research Centre focused on the analysis of economic policy and its foundations and implications), **EPEE**, University of Evry Val d'Essonne
- Le **Centre Pierre Naville** (Research on Work and Urban Policies), **CPN**, University of Evry Val d'Essonne
- Le **Groupe d'Analyse des Itinéraires et des Niveaux Salariaux** (Group on Analysis of Wage Levels and Trajectories), **GAINS**, University of Le Mans
- Le **Centre de Recherches en Economie et en Management**, (Research centre in Economics and Management), **CREM**, University of Rennes 1 et University of Caen Basse-Normandie
- Le **Groupe de Recherche ANgevin en Économie et Management** (Angevin Research Group in Economics and Management), **GRANEM**, University of Angers
- Le **Centre de Recherche en Economie et Droit** (Research centre in Economics and Law) **CRED**, University of Paris II Panthéon-Assas
- Le **Laboratoire d'Economie et de Management Nantes-Atlantique** (Laboratory of Economics and Management of Nantes-Atlantique) LEMNA, University of Nantes
- Le **Laboratoire interdisciplinaire d'étude du politique Hannah Arendt** – Paris Est, **LIPHA-PE**
- Le **Centre d'Economie et de Management de l'Océan Indien**, **CEMOI**, University of La Réunion

TEPP brings together 230 teacher-researchers and 100 doctoral students. It is both one of the main academic operators in the evaluation of public policies in France, and the largest multidisciplinary federation of research on work and employment. It responds to the demand for impact assessment of social programs using advanced technologies combining theoretical and econometric modeling, qualitative research techniques and controlled experiences.