



Università degli Studi di Modena e Reggio Emilia  
Dipartimento di Economia Politica



## Materiali di discussione

\\ 564 \\

### The Distribution of the Gender Wage Gap in Italy: Does Education Matter?

di

Tindara Addabbo<sup>1</sup>  
Donata Favaro<sup>2</sup>  
Stefano Magrini<sup>3</sup>

September 2007

Università degli Studi di Modena e Reggio Emilia  
Dipartimento di Economia Politica  
Via Berengario, 51  
41100 Modena, Italy

<sup>1</sup> e-mail: [tindara.addabbo@unimore.it](mailto:tindara.addabbo@unimore.it)

<sup>2</sup> Università degli Studi di Padova Italy  
e-mail: [donata.favaro@unimore.it](mailto:donata.favaro@unimore.it)

<sup>3</sup> Università di Venezia Ca' Foscari Italy  
e-mail: [smagrini@unive.it](mailto:smagrini@unive.it)



# The distribution of the gender wage gap in Italy: does education matter?

Tindara Addabbo, Donata Favaro and Stefano Magrini

## Tindara ADDABBO

Department of Economics, University of Modena and Reggio Emilia, Italy.

**Affiliations:** CAPP (Centro di Analisi delle Politiche Pubbliche)

<http://www.capp.unimo.it>; Child (Centre for Household, Income, Labour and Demographic Economics) <http://www.child-centre.it/>. RECent (Center for Economic Research),

<http://www.economia.unimo.it/userfile/21/recent.pdf>

Email: tindara.addabbo@unimore.it

## Donata FAVARO

Department of Economics, University of Padova, Italy.

**Affiliations:** Child (Centre for Household, Income, Labour and Demographic Economics), University of Torino, <http://www.child-centre.it/>.

Email: donata.favaro@unipd.it

## Stefano MAGRINI

Department of Economics, University of Venice 'Ca' Foscari', Italy.

**Affiliations:** SSAV (School for Advanced Studies in Venice), Venice International University. Email: smagrini@unive.it

September 2007

## **Abstract**

In this paper we evaluate the wage gap due to differences in rewards to characteristics by studying the entire distribution of the individual unexplained wage gap. We use quantile regressions and an adaptation of the procedure suggested by Machado and Mata (2005) to derive the marginal distributions of predicted and counterfactual female wages. Then, we estimate probability distributions of unexplained wage gap conditional to observed characteristics. The methodology allows to evaluate the probability of women with different characteristics to experience any level of discrimination. The main focus of the paper is on the relationship between human capital characteristics and outcomes in differences in pay. In particular, we focus our attention on different educational levels. Under the hypothesis that women invest in education to signal their productivity, we should detect a lower wage gap –due to differences in rewards to characteristics– among high educated females. Our analysis suggests that education can be a good signal but not for all females. We also show that education interacts differently with other human capital characteristics, such as general experience acquired in the labour market. The analysis is carried out on Italian data drawn from the last available cross-section of the European Community Household Panel (2001)

JEL classification: J31, C14

Keywords: Gender wage gap, distributive analysis, human capital.

## 1. Introduction

Despite the attempt of institutions and international organizations to raise the awareness on the relevance of gender wage differentials, research on the gender pay gap in Italy has been relatively scant, although increasing in recent years especially because of a surge in interest in comparative studies among European countries (Arulampalam et al. 2005, Beblo et al. 2003b, Olivetti and Petrongolo 2005). In particular, some effort at the national level has been made in the last few years through the contributions of Addabbo and Favaro (2007), Favaro and Magrini (*in press*), Comitato Nazionale Parità e Pari Opportunità (2001), Rustichelli (2005) and Villa (2005).

The international literature on the topic has been paying special attention on discussing the methodological aspects of the analysis, suggesting new analytical schemes, often complementary, that enrich the set of methodological instruments for measuring the incidence of wage differentials and provide alternative ways and more sophisticated tools than the traditional methodology suggested by Oaxaca and Blinder (1973).

The main contribution of the traditional Oaxaca-Blinder approach can be summarised with the idea that the raw wage gap can be decomposed into two terms: a first term representing productivity differences explained by individual characteristics and a second term explaining earnings gaps in terms of differences in the remuneration of the characteristics; this second term is then usually interpreted as the discriminatory component of the wage gap. On the other hand, the decomposition relies on the preliminary evaluation of the rewards to individual characteristics, which is carried out by applying OLS estimates to gender-separate ‘earnings functions’; consequently, the two components of the raw wage gap are calculated in average terms.

This methodology dominated the empirical literature on wage differentials throughout the 1980s and the 1990s and is still largely applied by political and scientific institutions to evaluate the two components of the pay gap. The European Commission (2002) estimates an average gap between

female and male gross hourly earnings in the European area of 16.2 percentage points<sup>1</sup>, of which less than 3% appears to be caused by average differences in characteristics while more than 13% is attributed to differences in remuneration. When disaggregating data by country, a more composite picture comes forward where United Kingdom, the Netherlands, Austria, Ireland, Germany and Spain account for the highest raw gaps; on the contrary, Greece, France, Denmark, Italy, Belgium and Portugal experience the lowest gender wage gaps. In addition to this, the study shows that the share of gender pay gaps due to differences in the characteristics' remuneration varies significantly among countries. Italy performs rather well, having one of the lowest wage gap in the area (around 9%); however, the differential is almost totally explained by differences in the remuneration of the characteristics.

Inter-country analyses that make use of comparable data, however, depict a gender pay gap in Italy which appears to be more optimistic than that emerging from national sources. On the other hand, Italy lacks good national sources of data on earnings and worked hours and most of the research has to rely on approximations of hourly wages or on monthly/yearly earnings, probably implying an overevaluation of the gap. Flabbi (2001), using the Bank of Italy dataset, estimates a monthly earning gap equal, on average, to 19%, 5% of which is attributed to differences in characteristics and 13.7% to differences in the rewards. More recently, Rustichelli (2005) evaluates a gender gap between female and male daily wages of, on average, 39 percentage points, whereby differences in the rewards are responsible for most of the gap (precisely, 26.9 points).

The Oaxaca-Blinder's approach is recognised as a fundamental contribution to the research on gender differences in payment; nevertheless, as recalled at the outset, a number of recent contributions have introduced important methodological improvements to the traditional framework. In primis, it has to be mentioned the development set up by Juhn, Murphy and Pierce (1991), who decompose the wage gap by taking into account also differences in non-observable characteristics.

---

<sup>1</sup> Estimates based on data from the European Community Household Panel (ECHP), wave 5 (year 1998).

A more recent advancement is owed to Machado and Mata (2005), according to which female marginal distributions of theoretical and counterfactual wages are derived using coefficients estimated at different points of male and female distributions, using quantile regression analysis (Koenker and Bassett, 1978; Buchinsky, 1998).

All these papers, however, still suffer from an important drawback: the analysis is based on the comparison, quantile by quantile, between the distribution of estimated wages and the distribution of counterfactual wages. As a result, these methodologies are unable to shed light on the individual dimension of the wage gap due to characteristics and may produce misleading results as the differences between the two distributions do not depend exclusively on discrimination suffered by individuals but also on the mobility across quantiles introduced when generating the counterfactual wage distribution. In other words, these recent methodologies are not actually able to analyse the extent of the gap at different points of wage distribution.

In this paper we study wage differentials between male and female workers in Italy by employing a method recently developed (Favaro and Magrini, *in press*), which allows to exactly evaluate the extension of the gap, due to rewards, experienced by each female worker. The method consists in measuring the incidence of the differential at the individual level, i.e., female by female, and then in estimating the conditional bivariate density functions defined on the individual levels of the gap and the corresponding individual characteristics. In this way, it is possible to identify which categories of female workers are affected by low rewards the most and to assign a probability of occurrence to any interval of “unexplained” wage gap, also conditional on any characteristic or factor of interest. The method, while providing a contribution to the evaluation of the individual incidence of the wage gap and its components, still relies on estimating first gender-specific predicted and counterfactual wages. Differently from Favaro and Magrini (forthcoming), in this article we depart from the OLS estimation procedure and employ the quantile regression analysis (Koenker and Bassett, 1978; Buchinsky, 1998) and a simplified version of Machado and Mata (2005) to derive marginal distributions from estimated values.

Based on the literature showing that investing in education is a signal of productivity, we explore the Italian component of the wage gap due to differences in rewards to characteristics conditional on different educational levels, expecting a lower component along the wage distribution of most educated females. Actually, our study will show that, although the “unexplained” part of the wage gap is proportionally lower among highly than lowly educated women, its extent tends to broaden at the highest wage levels. Conversely, among lowly educated women the probability to experience higher differences in rewards raises at the lowest earnings levels.

Our analysis is based on the sample of employed workers, 15-65 year olds, selected from the last available wave (year 2001) of the European Community Household Panel (ECHP). We focus exclusively on employed people primarily because of the lack of satisfactory information on self-employed workers (especially on earnings and hours of work), which makes rather difficult any comparison or unified treatment with employed workers.

The ECHP allows to control for many demographic and socioeconomic characteristics of the individuals and to derive different measures of human capital; in addition to providing information on the starting year of activity of every individual, which allows to determine her potential working experience, the dataset includes data on the number of years of working activity with the present firm. Therefore, we can separately evaluate the incidence on wages of general experience and firm-specific experience. Furthermore, the survey provides detailed information on professions and some subjective evaluations on the degree of responsibility involved in the occupation. Finally, other characteristics related to working activity can also be observed, such as the type of contract, the size of the firm and the economic sector of activity.

The paper is organised as follows. In Section 2 we present a brief review of the literature on the evaluation of the gender pay gap in Italy, with special attention to the latest contributions. In Section 3 we describe the distributive methodological approach to the analysis of the unexplained wage gap component of the gender wage gap. In Section 4 we describe the dataset and provide

some descriptive statistics. Earnings function estimates and the distributive analysis of the unexplained wage gap are discussed in Section 5. Section 6 concludes.

## **2. The wage gap due to differences in the rewards to characteristics. The Italian case**

Gender wage differentials have not been arousing in Italy as much interest among researchers as in other European countries, although institutions –at the national and European level– have been taking a great deal of attention to the issue and calling for strong initiatives at different levels. Only recently a few attempts to evaluate the Italian gender wage gap have been carried out; however, most of these studies relied on data which are not up to date and without taking proper account of the recent methodological developments in the field.

Most of the literature on the Italian case has been relying on the traditional Oaxaca-Blinder method to decompose the raw wage gap and to evaluate the average of the component due to differences in rewards, that is labelled the “unexplained” part of the wage gap –because it is not justified by differences in characteristics–. Some of the most recent examples are Flabbi (1997 and 2001), ISTAT (2005) and Comitato Nazionale Parità e Pari Opportunità (2001). However, the results of these studies strongly depend on the type of data employed in the analysis. Indeed, most Italian surveys do not provide comparable information on wages; some collect data on gross rather than net earnings; some allow to indirectly construct hourly wages but many of them do not provide the necessary information to calculate the wage rate and estimations have to rely on total earnings. As a consequence, results can be very different from one study to another and some caution is needed while making comparisons and drawing general conclusions.

Addis and Waldmann (1996), using the 1989 cross-section from the Bank of Italy Survey on Household, Income and Wealth (SHIW), estimate an average level of unexplained wage differentials equal to 13%; the differential is evaluated with respect to yearly earnings net of taxes and social security contributions. By using the same survey, but on year 1991, Flabbi (1997)



evaluates a wage gap due to differences in rewards that ranges between 8 and 12 percentage points, depending on the population of reference and on the econometric method employed –either OLS or Instrumental Variables. The same author estimates in 1995 that differences in the rewards to characteristics are responsible for a gap of 15% in net yearly earnings. He also shows how the percentage of the gap due to differences in the returns has not been changing across time: in 1977, out of a raw gap of 29.4 percentage points, 16% was attributed to differences in rewards; in 1995, despite a much lower raw wage gap than in the late 1970s (19.9%), the proportion attributed to differences in the rewards did decreased in an analogous way and still mounted to 15 percentage points.

More recently, ISTAT (2005) has evaluated the average level of “unexplained” wage gap on data from the Italian Structure of Earnings Survey (SES) for year 2002. That survey provides gross hourly earnings for individuals employed by firms with at least ten employees in the industry and service sectors. The ISTAT study estimates an average component of the “unexplained” gender gap in gross hourly earnings equal to 11 percentage points. However, in the same year Rustichelli (2005) reports rather different results. Using the administrative data of the *Istituto Nazionale della Previdenza Sociale* (National Institute of Social Security)<sup>2</sup> Rustichelli estimates a random-effect model of monthly wages controlling for the level of bargaining (national- versus regional/provincial- or firm-level bargaining) and evaluates an average component of the wage gap due to differences in rewards to characteristics equal to almost 27%, out of an estimated average gap equal to 39%.

A few years ago, the Italian Comitato Nazionale Parità e Pari Opportunità commissioned a group of Italian economists and sociologists a research on gender differences in pay. The results of this project have been published in 2001 (Comitato Nazionale Parità e Pari Opportunità, 2001). The economic section of the research has been carried out using data from the 1993 and 1995 waves of

---

<sup>2</sup> These data are representative of the population of workers employed in the private sector during the 1996-2002 period.

the European Community Household Panel; by applying the Oaxaca-Blinder methodology, the research estimates a gender wage gap, due to unequal rewards to characteristics, equal to around 20 percentage points in the first year of observation and 16 points in 1995.

Recent international contributions to the field have suggested to analyse the extent of the pay gap at different points of the wage distribution. Some examples are Juhn, Murphy and Pierce (1991), Fortin and Lemieux (1998) and Machado and Mata (2005). Other contributions propose to evaluate the level of wage differentials female-by-female to exactly determine the incidence of the unexplained wage gap at individual level and to study the relationship between each individual characteristic and the extent of wage differentials (Jenkins, 1994). However, very little work on Italian data has been taking distributional aspects into account. Addabbo and Favaro (2007) use data from the last available cross-section of the ECHP to evaluate the extension of the wage gap due to differences in rewards at different points of the distribution of female earnings, conditioning on the educational level. Using a quantile regression estimation procedure, they show that lowly educated women suffer a higher unexplained wage gap than highly educated colleagues across the whole distribution. However, while the authors find lowly educated females to be affected by a marked sticky-floor effect, they show some evidence of a glass-ceiling pattern among highly educated females.

Favaro and Magrini (*in press*) study the distribution of the wage gap among young workers in North-eastern Italy. Following the methodological suggestions of Jenkins (1994), they develop a non-parametric estimation of bivariate density functions to evaluate the probability distribution of the wage gap for female workers. As will be clarified later, this method differs from previous ones inasmuch it allows to determine the range of the gap relative to any characteristic and to assign a probability to any level<sup>3</sup>. Their results show that the component of the wage gap due to differences in rewards has been increasing through the 1990s across the whole distribution, but it has been more accentuated for females earning the highest wages. Therefore, a glass ceiling effect seems to have

---

<sup>3</sup> The same method is applied in the present article and it will be discussed in the methodological Section of the paper.

been emerging through the last decade: highly educated women suffer, in general, lower levels of differences in returns to characteristics than lowly educated females but experience much higher increases in the gaps on moving towards the top of the distribution. In addition, the accumulation of other human capital characteristics, such as experience and tenure in the firm, does not help women to close up the wage gap.

The international literature has been contributing to enrich the analysis of the Italian gender wage gap; in fact, some comparative analyses have recently provided additional empirical results on the distribution of the gender pay differential in Italy. Arulampalam *et al.* (2005), for example, present quantile regression analyses on a sample of eleven European countries and focus on the wage gap in the private and public sector, separately. They confirm, for the Italian case, an unequal incidence –along the distribution of wages– of the proportion of the raw wage gap that is explained by differences in the returns to characteristics. Moreover, this component behaves differently between private and public sectors: a higher proportion of wage gap due to differences in rewards to characteristics is estimated at both extremes of the private sector wage distribution; in contrast, female employees working in the public sector experience higher differences in rewards only at the top of the distribution. This result is common to most of the analysed countries.

Beblo *et al.* (2003) apply Juhn, Murphy and Pierce decomposition method on wages estimated through the Lewbel two-step procedure and show that, for employees 25-55 years old who work at least 8 hours per week, the raw wage gap is almost stable across different deciles of the distribution. On the contrary, the part of the gap due to differences in remunerations of characteristics slowly decreases up till the median value and sharply declines thereafter.

Olivetti and Petrongolo (2005) suggest a methodology to impute wages to individuals not in work, by making assumptions on the position of the imputed wage observations with respect to the median. By means of this procedure, Olivetti and Petrongolo are able to detect the impact of selection on gender wage gaps by comparing estimated wage gaps on the base sample with those obtained on the sample enlarged with wage imputation. Using the ECHP and applying the Oaxaca-

Blinder decomposition method, they estimate an average gender pay gap due to differences in characteristics of around 12 percentage points.

The study of the extent of the wage gap along the wage distribution has been drawing much attention in other European countries. So, Albrecht *et al.* (2003) have analysed the Swedish context, Albrecht *et al.* (2004) have applied the methodology to the Netherlands, and different analyses have been proposed on Spanish data [Garcia *et al.* (2001), Gardeazabal and Ugidos (2005), Del Río *et al.* (2006), De la Rica *et al.* (2005)].

### **3. A distributive approach applied to the estimated marginal female wage functions**

The empirical research on gender wage differentials has been developing since the initial years of the 1970s, starting from the fundamental contribution of the works by Oaxaca (1973) and Blinder (1973). The original suggestion of the Oaxaca-Blinder's methodology was that the observed gap between male and female wages can be decomposed into two different parts: the component due to differences in individual characteristics responsible for differences in labour productivity (the so called "explained" part of the gap), and the part due to gender differences in the rewards to those characteristics (the "unexplained" component of the differential); this second term is recognized as the discriminatory component of the wage gap though differences in the individual characteristics may themselves being determined by discrimination.

The Oaxaca-Blinder approach provides a method to evaluate the unexplained part of the wage gap; the evaluation, however, is made in average terms, by applying OLS regression to estimate the rewards to individual characteristics, on the basis of which an index of discrimination is constructed. As a consequence, the evaluation of discrimination is effectively reduced to an average prediction and the analysis is simplified to detect group discrimination, which means deterministic discrimination against minority groups, whose individuals are affected equally. As a consequence, the Oaxaca-Blinder methodology is not the proper instrument to evaluate stochastic (or individual)

discrimination, which allows for individual deviations within a particular group and could arise because employers do not have perfect information and do not know the true productivity of each worker: since employers are imperfectly informed, they set wages relying on the characteristics they observe, which although related to productivity, do not perfectly represent productivity. Stochastic models predict both group discrimination – when the wage difference is evenly distributed across all levels of the observed characteristics – and individual discrimination, because of a different distribution of the characteristics within the different groups.

The inadequacy of the O-B methodology to evaluate the distribution of the discriminatory component of the wage gap across individuals was first emphasized in the middle 1980s by Dolton and Makepeace (1987) and Munroe (1988) who highlighted the risk in using the Oaxaca-Blinder methodology not to detect any discrimination even though some workers were effectively discriminated against. These authors highlighted the need for evaluating both the average value of discrimination and its dispersion. Building on these contributions, Jenkins (1994) developed a method for analysing discrimination that makes use of the complete information contained in the distributions of estimated and reference female earnings. The method he proposes entails a comparison of the relative position of the Generalised Lorenz Curve (GLC) with respect to the Generalised Concentrations Curve (GCC), suggesting that discrimination exists whenever the GCC lies above the GLC. However, while supporting the call for ‘tractable methods of presenting information about the complete distribution of discrimination experience’ (Jenkins, 1994), we believe that the method developed by Jenkins falls short from being an adequate answer. The fundamental problem in Jenkins’ method is that a GCC above the GLC does not necessarily imply discrimination against women along the whole distribution: by construction, the distance between the GCC and the GLC depends on cumulated differences between estimated and reference earnings, so that it can be positive even though the marginal contribution of one more female worker is negative. The effective contribution of Jenkins proposal to overcome Oaxaca-Blinder limitations is weakened further; in fact, while proposing to compare the whole distribution of predicted and

counterfactual female wages, the method adopts OLS estimates to construct those distributions, estimates which provide average evaluations of the rewards to the characteristics.

It is only in the last few years that research on wage differentials has been able to advance with respect to the OLS methodology, by resorting to quantile regression (Koenker and Bassett, 1978; Buchinsky, 1998). The quantile regression method, applied to the estimation of wage equations, consists in evaluating the rewards to individual characteristics by allowing for different values in correspondence of any chosen point across the wage distribution. The method has been applied to the wage differential analysis by Machado and Mata (2005) who developed a procedure to obtain the marginal distributions of female predicted and counterfactual wages, once the vector of coefficients of the wage equations is estimated. The Machado and Mata methodology has improved the power to explain the pattern of wage discrimination, since the level of discrimination (defined as the difference between the predicted wage and the counterfactual wage) can be observed across the whole female distribution; this allows to detect the likely uneven incidence of discrimination across individual characteristics.

Based on these considerations, in this article we propose to combine the econometric analysis developed by Machado and Mata (2005) to build marginal distributions of theoretical and counterfactual wages, with an alternative method that evaluates the wage gap by focussing directly on the distributions of marginal estimated and reference earnings. The method we propose makes it possible to detect both group and individual components of the wage gap, and to assign a probability of occurrence to any level of wage differential at any point in the distribution with respect to a given characteristic. This methodology has been used in a previous work where we studied wage differentials among young workers in North-eastern Italy (Favaro and Magrini, *in press*).

Here, we focus only on the analysis of the component of the gap due to differences in rewards to characteristics, the so called “unexplained” part of the pay gap, and try to detect its incidence across

the distribution of wages by different educational levels. The econometric method and the distributive methodology we propose for doing so can be described as follows.

The quantile regression method consists in estimating wage equations at different points of the relative wage distribution. Given the covariates vector  $z$ , quantile regression allows to estimate  $Q_\theta(\omega|z)$ , corresponding to the  $\theta$ -th quantile of the distribution of the log wage ( $\omega$ ), at any  $\theta \in (0,1)$ . The quantile regression model is assumed to be linear:

$$\omega = z'\beta_\theta + u_\theta \quad (1)$$

Where  $\omega$  is the log of wages and  $\beta_\theta$  is a vector of coefficients, the quantile regression coefficients. The distribution of the error term  $u_\theta$  is unspecified and it is simply assumed that  $Q_\theta(u_\theta|z) = 0$ .

The estimated values of the  $\theta$ -th quantile of the log wages, conditioned to covariates  $z$ , is equal to:  $Q_\theta(\omega|z) = z'\hat{\beta}_\theta$ .

For any  $\theta \in (0,1)$ ,  $\hat{\beta}_\theta$  can be estimated by minimising in  $\hat{\beta}_\theta$  the following expression:

$$n^{-1} \sum_{i=1}^n \tilde{\rho}_\theta(u_i - z_i'\hat{\beta}_\theta) \quad (2)$$

where:

$$\tilde{\rho}_\theta(u_i) = \begin{cases} \theta u_i & \text{for } u_i \geq 0 \\ (\theta - 1)u_i & \text{for } u_i < 0 \end{cases} \quad (3)$$

The vector of coefficients  $\hat{\beta}_\theta$  can be obtained by estimating each equation separately or simultaneously. The simultaneous procedure allows to obtain an estimate of the entire variance-covariance matrix of the estimated coefficients, which is necessary to implement the testing analysis of inter-quantile difference of coefficients<sup>4</sup>. Following the above described procedure, we estimate

---

<sup>4</sup> The bootstrapping procedure allows us to test whether coefficients of different quantile regressions are significantly different.

the rewards to worker characteristics, by specifying different models for females and males; thus, we obtain a vector of estimated quantile coefficients for female workers,  $\hat{\mathbf{a}}_{\hat{\epsilon}_f}$  and a vector of estimated quantile coefficients for male workers,  $\hat{\mathbf{a}}_{\hat{\epsilon}_m}$ .

Given the estimated coefficients, we derive the marginal distributions of the predicted (theoretical) and the counterfactual female wages by applying the methodology adopted by Albrecht *et al.* (2003)<sup>5</sup>. Female predicted wages are theoretical wages that females can earn given their characteristics and the rewards recognised to female workers; on the other way, female counterfactual wages are wages that women would be paid if female characteristics were rewarded at male returns. In order to construct predicted and counterfactual distributions of wages, we need to simulate a random distribution of characteristics. We proceed as follows:

- We take a draw from the female database and construct a predicted wage by multiplying characteristics  $z_f$  of the selected individual by the relative estimated coefficients,  $\hat{\mathbf{a}}_{\hat{\epsilon}_f}$ , for a given quantile  $\hat{\epsilon}$ . We repeat that operation  $N=100$  times with respect to every quantile, ending up with the estimated marginal distribution of female predicted wages.
- We repeat the procedure described above, but this time we apply male coefficients,  $\hat{\mathbf{a}}_{\hat{\epsilon}_m}$ , to female characteristics,  $z_f$ . Then, we obtain the marginal distribution of female counterfactual wages.
- The difference between the constructed marginal distributions represents the “unexplained” component of the wage gap, that means the part of the wage differential not justified by gender differences in the characteristics, which is the component of the wage gap we have planned to analyse and that we will simply call “wage gap” (or “wage differential” or “pay gap”), for simplifying the reading.

---

<sup>5</sup> Albrecht *et al.* (2003) adopt a simplified version of the methodology proposed by J.A.F. Machado and J. Mata in a mimeo that was later published in the Journal of Applied Econometrics (Machado and Mata, 2005).



Having constructed the marginal distributions of cross-female predicted and counterfactual wages, we then proceed in analysing the incidence of the wage gap conditional on estimated earnings and to different observed characteristics –mainly human capital characteristics– by estimating non-parametrically the probability distribution of wage differentials conditional on the distribution of estimated earnings and of relevant factors or characteristics.

Let us indicate with  $d_i$  the estimated level of wage differential for observation (i.e., female worker)  $i$  and with  $x_i$  the corresponding level of any relevant factor or characteristic we want to analyse jointly to the wage gap. Then denote by  $F(d)$  the distribution of the wage differential and with  $F(x)$  the distribution of the variable  $x$ . Next, suppose these distributions can be described by density function, which can be indicated with  $f(d)$  and  $f(x)$  respectively. What we are interested in is the relationship between the two distributions; this can be simply written as

$$f(d) = \int_0^{\infty} f(d|x)f(x)dx \quad (4)$$

where  $f(d|x)$  is the density of the wage gap conditional on the level  $x$  for the factor or characteristic of interest.<sup>6</sup>

From an operational point of view, we obtain an estimate of  $f(d|x)$  in three steps. First, we estimate non-parametrically the joint distribution of  $d$  and  $x$ , using a bivariate product kernel density estimator:

$$\hat{f}(d,x) = \frac{1}{nh_d h_x} \sum_{i=1}^n K\left(\frac{d-d_i}{h_d}, \frac{x-x_i}{h_x}\right) \quad (5)$$

where  $K(\cdot)$  is the kernel function, while  $h_d$  and  $h_x$  are the kernel bandwidths.<sup>7</sup> Next, we obtain an estimate of the marginal distribution  $f(x)$  by numerically integrating the joint distribution with

---

<sup>6</sup> For a more general description of the method see Favaro and Magrini (*in press*).

<sup>7</sup> To estimate the joint distribution, we use a Gaussian product kernel with bandwidths chosen optimally according to Silverman (1986).

respect to the wage gap<sup>8</sup>. Finally, we obtain the estimate of  $f(d|x)$ , the distribution of the pay differential conditional on estimated earnings or any characteristic of interest, by dividing the joint distribution by the marginal one<sup>9</sup>:

$$\hat{f}(d|x) = \frac{\hat{f}(d,x)}{\hat{f}(x)} \quad (6)$$

The incidence and direction of the wage gap can thus be studied by analysing directly the shape of the three-dimensional plot of the conditional distribution estimate and of the corresponding two-dimensional contour plot. In particular, the actual way in which this can be accomplished depends on whether the conditioning characteristics can be measured on a continuous space, as in the case of estimated earnings or of general experience accumulated during the individual's working history, or they should rather be represented as categorical or dummy variables, as in the case of the education level or of experience accumulated inside the firm<sup>10</sup>.

As for the interpretation of the results in the case in which worker's characteristics are measured on a continuous space (and, therefore, also in the case we focus on the relationship between wage differentials and estimated earnings), the absence of differences in rewards to characteristics is represented by a concentration of the probability mass along the line running parallel to the characteristic's axis and in correspondence to a wage gap equal to zero. As a consequence, evidence of differences in rewards against (in favour of) female workers is signalled by a probability mass lying above (below) this horizontal line. In contrast, when the characteristic of interest presents  $l$  levels so that the individuals can thus be divided into  $l$  mutually exclusive subsets, we estimate  $l$  different stochastic kernels as in equation (6). Each of these stochastic kernels therefore shows the distribution of the wage gap conditional on estimated earnings, for a given level of the

---

<sup>8</sup> In this, we follow the procedure originally suggested by Quah (1996). As an alternative, the marginal distribution is often estimated directly using a univariate kernel. However, as pointed out by Overman and Ioannides (2001), the two estimators have identical asymptotic statistical properties and produce very similar results in practice.

<sup>9</sup> Under regularity conditions, this represents a consistent estimator for the conditional distribution (Rosenblatt, 1971; Silverman, 1986; Quah, 1996; Chen, Linton and Robinson, 2001).

<sup>10</sup> As we are going to explain later in the paper, ECHP data do not contain a continuous variable measuring tenure and we are obligated to include experience accumulated inside a firm as three dummy variables.

characteristic. Moreover, direct comparisons between the estimates for different levels of the characteristic indicate how the conditional distribution is affected by changes in the level of the characteristic.

#### 4. The dataset

The analysis is carried out on a sample of employed workers aged 15 to 65 selected from the 8<sup>th</sup> wave<sup>11</sup> of the European Community Household Panel (ECHP); we do not include the group of self-employed workers due to the unsatisfactory level of information on their earnings and hours of work, which makes the comparison with employed workers rather difficult.

The model we estimate assumes that the wage level is affected by individual characteristics and other characteristics linked to the demand side of the labour market, such as the size of the firm, the sector of activity, the type of contract and the reference regional context. Regarding individual characteristics, the ECHP provides information on several factors of significant interest to evaluate individual human capital endowment; in particular, we can rely on data on education, the starting year of working activity, the number of years of experience in the actual firm, the level of supervisory responsibility in current job and the professional category. In addition, the survey supplies some key demographic and socioeconomic characteristics of the individuals, such as age, sex, marital status and family composition.

The focus of the paper is on wage differentials conditional on educational levels; therefore, we analyse two different subsamples of workers. The first includes workers with an educational level equal or higher than a “second stage of secondary level education (ISCED 3 or ISCED 5-7)”, which in Italy is equivalent to having at least a diploma of “*Scuola secondaria superiore*”<sup>12</sup>; we label this group of workers as “highly educated”. It is important to note that, when working on this subsample, we are actually studying individuals with either a diploma of “*Scuola secondaria*

---

<sup>11</sup> This is the most recent available wave, referring to year 2001.

<sup>12</sup> Upper-secondary school corresponds to post- compulsory school. Individuals are asked to choose whether to continue studying when they are 13 years old.

*superiore*” or a university degree. Consequently, in the estimation of the earning functions for this subsample, we take the group of individuals with a diploma of “*Scuola secondaria superiore*” as the reference category and add among the regressors a dummy variable reflecting the achievement of a university degree. The second subsample of workers comprises those who completed a primary stage of education and we label them as “lowly educated”.

In addition to education, we are able to control for human capital characteristics acquired in the labour market. The ECHP collects information on the year of individual first entrance in the labour market; by using that information, we construct the number of years of potential experience any worker could have accumulated as since her first working experience. Some caution is generally needed when using such a “theoretical” measure of experience in analyses on wages and gender differentials: the measure may not correspond to the effective years spent in the labour market because it does not take into account periods of absence from the labour market, owing to unemployment, inactivity, or simply illness or parenthood. If this were the case, potential experience would overestimate the real number of years of working activity. In general, such a measurement problem arises both for males and females; however, as the empirical evidence suggests, the problem is more serious for females, due to the interruption connected with maternity. We try to address the issue by adding, among the explicatory variables, the interaction of potential experience with the number of children. If it is true that having children implies some penalty in terms of experience, we should detect a negative impact of that variable on the level of wages and to solve, at least partially, the problem.

We complete the information on individual human capital endowment by taking into account the period of permanence in the current firm, that we call “tenure”. Unfortunately, the European survey does not provide the precise number of years of tenure for all workers, but only for those that have been working in the same firm for less than fifteen years. As a consequence, we are not able to know the exact period of permanence when workers have been in the present firm for more than fifteen years. So, in order to normalize information, we are forced to use dummy variables that

capture the effect of different periods of time: we construct four different intervals corresponding to a period of tenure shorter or equal to five years, longer than five but shorter or equal to ten years, between eleven and fifteen years and longer than fifteen years.

Human capital characteristics are expected to positively affect the level of wages. However, the extent of their effect can reasonably be correlated with other job characteristics, such as the occupation category and the degree of responsibility concerning that occupation; it is therefore necessary to control for these factors to avoid either overestimation or underestimation (depending on the occupation) of the returns to human capital. Given that the ECHP survey provides a rather detailed classification of occupational categories, we include among our variables a dummy for each type of occupation listed in the dataset<sup>13</sup>. Moreover, since individuals in paid employment are asked whether they have “*a supervisory role*”, “*some intermediate supervisory role*” or a “*no supervisory role*”, we construct two dummies to capture the effect on the level of wages of the level of supervisory responsibility. A positive sign is obviously expected as a reward to a higher responsibility.

The literature on earnings and wage differentials provides good evidence of the impact of the sector of activity and contract conditions on wage levels: working in service activities generally guarantees a higher average wage than does working in the industry and agriculture sector. On the other hand, it is widely shown that earnings of Italian female workers employed in the public sector are normally higher than those of private employees. Therefore, in the econometric analysis we add dummies for macrosectors of activity and for public sector workers.

As a matter of fact, Italian wage rates seem to be correlated to the weekly amount of hours of work and to the length of the contract. Consequently, we try to capture the incidence of these

---

<sup>13</sup> We control for the following professional categories: physical, mathematical, engineering, life science and health professionals, teaching professionals, other professionals; physical, mathematical, engineering, life science and health associate professionals, teaching and other technical professionals; office and customer services clerks; personal and protective services workers; models, salespersons and demonstrators; skilled agricultural and fishery workers; craft and related trades, and extraction and building trade workers; metal, machinery, precision, handicraft, craft printing and related trades workers. These occupational dummies have been included among the regressors of the econometric model while the reference category is “Sales and services elementary occupations”.

contractual conditions by adding controls for the length of the contract –distinguishing permanent employment from fixed-term, short-term contracts and from other types of employment contracts<sup>14</sup>– and for part-time work<sup>15</sup>. Finally, we account for the size of the firm and for regional effects.

Table A1 in the Appendix summarizes some information about the sample. In general, female workers, either highly- or lowly educated, work a lower number of hours than their male colleagues. As for human capital characteristics, we observe a rather different composition by sex of the subsample of highly educated: the proportion of women with university education is almost 2 percentage points lower than the proportion of males with the same level of education. On the other hand, we do not observe any relevant difference by sex in accumulated general or specific experience. As expected, highly educated workers independently of sex, display a shorter average period of work than lowly educated employees.

Accessing supervisory positions is rather uncommon among lowly educated employees as only 3% of females and 4.7% of men have a high supervisory role in their occupations. The proportion is slightly higher (7.6% for females and 10% for males) when considering only an average level of supervision. The difference by sex in accessing responsibility positions is substantial when we look at highly educated employees<sup>16</sup>: only 6% of women, compared to 16.4% of men, exercise a significant supervisory role while 12.4% of women and 20.5% of men have an average level of supervision responsibility.

As expected, we observe a higher concentration of women in part-time work, especially for those with a low educational level. On the other hand, compared to a much higher concentration of females than males in the public sector conditional on highly-education, we surprisingly detect a rather similar proportion of lowly educated employees in public employment, independently of sex.

---

<sup>14</sup> We summarise in the category “other type of contract” the categories defined by the ECHP as “casual work with no contract” and “other arrangement”.

<sup>15</sup> We summarise in the category “other type of contract” the categories defined by the ECHP as “casual work with no contract” and “other arrangement”.

<sup>16</sup> This is consistent with the existence of vertical occupational segregation by gender in Italy (Rosti, 2006).

## 5. The distributional analysis

Before moving to the main focus of our study –the distributional analysis– we spend a few words on the estimation results that are reported in the Appendix<sup>17</sup>.

According to these results, there appear to be interesting differences between the rewards to the characteristics of highly educated (Table A2a) and lowly educated (Table A2b) females, particularly with reference to human capital attributes. General experience, for example, does have a significant effect on wages of lowly educated females but only at very high levels; on the contrary, highly educated women obtain a significant reward to experience accumulated in the labour market only if their wages are lower than the median value. In addition, the estimated coefficient of the interaction between “experience” and “number of children”, capturing the measure of penalty that female workers would suffer when having children, is not significant for less-educated women. It becomes significant and negative for highly educated women up till the median level.

Returns to tenure are generally statistically not significant in the sample of lowly educated females; on the contrary returns are significant for low-earning highly educated women and the rate increases as wages raise from the lowest decile to the median level.

Having some supervisory role positively affects the level of wages in both cases, but not along the whole wage distribution; indeed, lowly educated women have some advantage only if their wages are not too high, precisely lower than or equal to the median value. Highly educated females, on the contrary, gain independently of the wage level. However, the reward is higher in the sample of lowly educated women, amounting to twice the return estimated in the highly educated female sample.

---

<sup>17</sup> For an extensive discussion of the estimates, see Addabbo and Favaro (2007).

On the other hand, having a relevant supervisory role does not, in general, guarantee any economic advantage to lowly educated females, with the exception of those earning very high wages (at the highest decile of the distribution); differently, highly educated women with high supervisory roles can earn higher wages along the whole distribution and the gain increases as the wage rises.

Moreover, we observe some fundamental differences between female and male coefficients, both in the level of statistical significance and in the extent of their incidence on wages<sup>18</sup>.

As for highly educated workers (Tables A2a and A3a), it is a matter of fact that having a university educational level implies higher wages, independently of sex; besides, the highest the wage the larger the effect. However, rewards to university education are much lower in the female than in the male sample, with the only exception of the highest decile of the distribution; in that case, highly educated females are rewarded as much as their male colleagues. Females, however, show some advantage to acquire either general or specific experience. The permanence in the labour market has a significant and positive incidence only on the left-hand part of both female and male distributions; however, we detect higher rewards to females. On the other hand, specific experience accumulated in the present firm does not appear to significantly affect male wages. Indeed, the opposite is true for females: the longer they work in the same firm, the higher the wage, in particular for high-earning ones.

As we showed in Section 4, females accessing supervisory positions are rather uncommon. In addition to that, females who achieve such responsibilities do not obtain a compensation as high as males do. The disadvantage is particularly significant if we compare men and women with a high supervisory role.

Looking at the coefficients of lowly educated workers (Tables A2b and A3b), we see noticeable differences in comparison with highly educated employees. General experience turns out to be insignificant for female workers, except when they reach very high wage levels (at the highest

---

<sup>18</sup> See Addabbo and Favaro (2007) for tests on the statistical difference between male and female coefficients.



decile of the distribution); in contrast, the positive effect of acquiring general skills does significantly affect the whole distribution of male wages, but the last quantile. Also with respect to tenure we find completely different results according to the level of education: for lowly educated workers, specific experience acquired in the present firm does not affect female wages, at any level of observation, but positively influences –with a larger impact in correspondence to a longer tenure– the wages of males.

We can now move to the investigation of the wage gap based on the distribution approach. Before proceeding, two things must be noted in order to ease the interpretation of the results. First of all, the “unexplained” wage gap, calculated as the difference between females’ estimated and counterfactual earnings, is here expressed in percentage terms with respect to estimated earnings. Secondly, the lines reported in all contour plots are percentage contour lines. In particular, the value adjacent to each line indicates the percentage of the density volume that lays above (on the vertical axis of the three-dimensional plot) the line itself. So, areas enclosed within a low-value line are in fact associated with a high conditional probability level and thus enable us to identify the peaks of the conditional probability density.

The first step is represented by the study of the probability of the wage gap conditional on estimated earnings for different levels of education. In particular, Figure 1 shows both the three-dimensional and the contour plots<sup>19</sup> of the kernel estimates corresponding to low– (upper panes) and highly– (lower panes) educated females. Several important features appear to emerge.

First, all estimates suggest the existence of substantial wage differences against female workers as a large share of the conditional density mass is positioned above the horizontal line for both levels of education. Second, the extent of the gap decreases as estimated earnings increase for female workers with a low education level, as it can be inferred from the fact that the corresponding density mass is downward sloping. In contrast, such a negative relationship between differences in rewards

---

<sup>19</sup> The contour lines reported in all contour plots are percentage contour lines. In particular, the value adjacent to each line indicates the percentage of the density volume above (on the vertical axis of the three-dimensional plot) the line itself.

to characteristics and estimated earnings cannot be found for highly educated workers. Third, there appear to be substantial differences among education levels with respect to the variability of the phenomenon along the range of estimated earnings. To see this, let us focus on the 0.9 contour line, starting from lowly educated workers. In this case, we can notice that the variability of the wage gap is extremely high for low estimated earnings and decreases as estimated earnings increase. In contrast, the extent of the relative wage gap for highly-educated workers appears to be more homogeneously distributed along the earnings range, being substantially between lower levels.

A final interesting feature that emerges from Figure 1 is that the overall conditional probability of experiencing wage differentials against appears to be higher for females with high education levels. A first indication of this feature can be obtained by looking at the position of the peaks of the conditional probability densities as identified by the 0.1 contour lines: while the peaks for lowly educated workers lay underneath the horizontal line, the peaks for highly educated are positioned clearly above it. But, the significance of these feature is also confirmed by the calculations reported in Table 1. Indeed, along with the visual inspection of the kernel estimates, an assessment of the incidence of the wage gap can be performed by calculating the share of the volume of the estimated conditional density that lies above the horizontal line. Such a measure can thus be interpreted as the cumulative conditional probability of wage differences against female workers. Hence, a value higher (lower) than 50% can be seen as evidence of wage gaps against (in favour of) female workers. Additionally, making use of horizontal lines with a positive intercept, we can decompose this measure according to the incidence of the gap relative to estimated earnings. Looking now at the results reported in the table we can then see that the cumulative probability of wage gaps against women is equal to 81.66% for highly educated and to 66.08% for lowly educated workers. Moreover, while the cumulative probability of wage differentials against women in excess of 10% of their estimated earnings is rather similar among the two groups (with values around 50%), the cumulative probability of pay gaps against women between 0 and 10% of their estimated earnings is

equal to only 17.05% for low education workers, compared to a value of 29.20% for high education workers.

**[Figure 1 and Table 1 around here]**

In the methodological Section we pointed out that, whenever an individual characteristic can be measured on a continuous space, we can estimate directly the probability of wage differentials conditional to the values of the chosen characteristic. So, Figure 2 reports the estimated conditional density functions of the pay gap with respect to the years of potential experience accumulated in the labour market prior to the present occupation, again distinguishing between low and high education levels. While the conditional densities are predominantly positioned in the part of the  $(\hat{d}, \hat{x})$  plane corresponding to wage differences against female workers both for high and low levels of education, the variability of the gap appears to be sensibly wider for lowly educated females. However, looking at the position of the densities' crests in the three-dimensional plots, it appears that a wage gap of approximately 10% of the estimated earnings is the most likely outcome at all levels of experience and education. However, the slope of these crests and the positions of the peaks suggest that, for high education workers, the probability of suffering this level of wage gap tends to increase with the level of experience.

**[Figure 2 around here]**

Finally, we can see the results that can be drawn from the analysis of the relation between wage gaps and workers' tenure. However, before looking at the estimates, it must be noted that here we have chosen to follow a different way of partitioning the experience accumulated inside a firm with respect to the one adopted while estimating the wage equations. On that occasion, we constructed four separate dummy variables – corresponding to tenure levels between 0 and 5 years, between 6

and 10 years, between 11 and 15 years, and of more than 15 years – and excluded the first from the analysis. Here, in order to make use of all available observations and to distribute them as homogeneously as possible, we grouped the dummies into two simple categories. So, while Figure 3 reports the estimated density function of the wage gap conditional to estimated earnings for female workers with a tenure of 10 years or less, Figure 4 reports the corresponding estimate for workers with a tenure of 11 years or more.

The general picture that emerges from these Figures is quite unambiguous. Once more, we find significant evidence of a substantial degree of pay differences against female workers given that all reported densities lay well above the horizontal line. Besides, through these estimates we can further qualify one of the features noticed from the analysis of Figure 1. There, we noted the existence of a negative relationship between the extent of the wage gap and the level of estimated earnings for lowly educated females. Looking at the upper panes of Figures 3 and 4, we can see that this feature is substantially confirmed for both tenure categories. At the same time, the lower panes of the figures also confirm the absence of such a negative relationship for high education workers.

However, other interesting features can be recognised through a more detailed comparison between the two tenure categories. Concentrating on the estimates for low education levels, we can observe that, for workers with a tenure period that exceeds 11 years, the differences in the behaviour of the wage gap along the range of estimated earnings are more evident. In particular, we can clearly notice the presence of significant gaps in their favour for relatively high earnings (and, in particular, for earnings in excess of approximately 38 thousand euros), given that both peaks in the conditional density estimate are positioned below the horizontal line (Figure 4, upper panes). In contrast, one of the two corresponding peaks in the density estimate for workers with shorter tenure periods (Figure 3, upper panes) remains above the horizontal line. Moving now to the estimates for high education levels, we can notice that, while female workers with tenure periods of 11 years or more are characterised by a total absence of pay gaps in their favour for relatively high levels of

estimated earnings (Figure 4, lower panes), this phenomenon is instead significantly present for workers with shorter tenure periods.

**[Figures 3 and 4 around here]**

## **6. Conclusions**

In this paper we evaluate the gender pay gap due to differences in returns to characteristics by suggesting a distributional approach. In particular, the method assigns a probability of occurrence to any level of discrimination conditional to any level of a given factor or characteristic.

The analysis shows that differences in pay due to differences in rewards to characteristics between Italian men and women are not evenly distributed among workers with different educational endowments. Women achieving the highest educational levels experience lower pay gaps compared to their colleagues with lower education; more precisely, the probability distribution of the pay gap conditional on the distribution of earnings of highly educated females is concentrated on lower wage differentials than the distribution concerning lowly educated females. However, the variability of the wage gap along the range of estimated earnings is different: for lowly educated females, the gap is extremely high in correspondence of low earning levels and decreases as earnings increase. On the other hand, for highly educated females the variability decreases only at very high wage levels; still, in correspondence of these higher wages, the probability distribution shifts upwards, showing that gender differences in the returns to characteristics are higher and more likely to occur in correspondence of highly educated women reaching the top of earnings. These results confirm that in Italy, as elsewhere, there is a different pattern of wage discrimination affecting female workers in line with their educational endowment. Females with an educational level lower than an upper secondary-school diploma experience some sticky-floor pattern; on the contrary, highly-educated female wages are affected by some glass-ceiling pattern.

Our study also suggests interesting results when interacting education with other human capital characteristics. In summary, we find that lowly educated women can suffer lower gender pay gaps due to differences in rewards to characteristics as they achieve longer labour market experience and longer permanence in the firm. In contrast, the gender gap in rewards for highly educated females worsens up as their experience, either general or specific, increases.

Although our method clearly differs from those recently applied by other authors to several European countries and that rely on the comparison between predicted and counterfactual female marginal wage distributions, our findings on the Italian case have some points in common with theirs: indeed, we detect some patterns of sticky floor in the sample of lowly educated females and some glass ceiling pattern among highly educated females.

## References

- Addabbo, T. and Favaro D., 2007, "Education and wage differentials by gender in Italy", CHILD Working Papers N. 04/2007, Centre for Household, Income, Labour and Demographic Economics, University of Torino.
- Addis, E., and Waldmann, R., 1996, "Struttura salariale e differenziale per sesso in Italia", *Economia e lavoro*, 30(1-2), 87-103.
- Albrecht, J., Bjorklund, A. and Vroman, S., 2003, "Is there a glass ceiling in Sweden?", *Journal of Labor Economics*, 21 (1), 145-177.
- Albrecht, J., Van Vuuren, A. and Vroman, S., 2004, "Decomposing the gender wage gap in the Netherlands with sample selection adjustments", *IZA DP 1400*, Bonn.
- Arulampalam, W., Booth, A.,L., and Bryan, M.L., 2005, "Is there a glass ceiling over Europe? Exploring the gender pay gap across the wages distribution", *ISER Working Paper 2005-25*. Colchester: University of Essex.
- Beblo, M., Beninger, D., Heinze, A. and Laisney, F., 2003a, "Methodological issues related to the analysis of gender gaps in employment, earnings and career progression", *Final Report*, European Commission, Employment and social affairs DG.
- Beblo, M., Beninger, D., Heinze, A. and Laisney, F., 2003b, "Measuring selectivity corrected gender wage gaps in the EU", *ZEW Discussion paper No.03-74*.
- Blinder, A. S., 1973, "Wage discrimination: reduced forms and structural estimates", *The Journal of Human Resources*, 8 (4), 436-55.
- Buchinsky M., 1998, "Recent advances in quantile regression models: a practical guideline for empirical research", *The Journal of Human Resources*, 33, 88-126.
- Chen, X., Linton, O. and Robinson, P.M., 2001, "The estimation of conditional densities", *The Suntory Centre discussion paper no. EM/01/415*, LSE, London.
- Comitato Nazionale di Parità e Pari Opportunità, (eds.) 2001, *I differenziali salariali per sesso in Italia*, Roma, Istituto Poligrafico dello Stato.
- De La Rica, S., Dolado, J.J. and Llorens, V., 2005, "Ceiling and Floors: gender wage gaps by education in Spain", *IZA DP 1483*, Bonn.
- Del Rio, C., Gradìn, C. and Cantò, O., 2006, "The measurement of gender wage discrimination: the distributional approach revisited", *Ecineq WP 2006-25*.
- Dolton P.J. and G.H. Makepeace, 1987, Marital status, child rearing and earnings

- differentials in the graduate labour market, *Economic Journal*, 97, pp. 897-922.
- Durlauf, S.N. and Quah, D.T., 1999, "The new empirics of economic growth", in: Taylor J., Woodford M. (Eds.), *Handbook of Macroeconomics*, Vol. 1. North-Holland, Amsterdam, 235-308.
- European Commission, 2002, *Employment in Europe*, Chapter 1.
- Favaro, D. and Magrini, S., (*in press*), "Group versus individual discrimination among young workers: A distributional approach", *The Journal of Socio-Economics* (2007), doi: 10.1016/j.socec.2007.08.009.
- Flabbi, L., 1997, "Discriminazione di genere e rendimenti dell'istruzione: un'analisi su microdati", *Rivista di Politica Economica*, 12, 187-234.
- Flabbi, L., 2001, "La discriminazione: evidenza empirica e teoria economica" in Brucchi Luchino, 2001, *Manuale di Economia del Lavoro*, Chapter 17, Bologna, Il Mulino.
- Fortin N. M. and Lemieux T., 1998, "Rank regressions, wage distributions, and the gender wage gap", *Journal of Human Resources*, 33(3), 610-643.
- García, J., Hernández, P.J. and López-Nicolàs, A., 2001, "How wide is the gap? An investigation of gender wage differences using quantile regression", *Empirical Economics*, 26, 149-167.
- Gardeazàbal, J. and Ugidos, A., 2005, "Gender wage discrimination at quantiles", *Journal of population economics*, 18, 165-179.
- ISTAT, 2005, *Rapporto Annuale 2004*, Roma, Istat.
- Jenkins S., 1994, Earnings discrimination measurement: A distributional approach, *Journal of Econometrics*, 61, 1, 81-102.
- Juhn, C., Murphy, K.M. and Pierce, B., 1991, "Accounting for the slowdown in black-white wage convergence", in Kosters M. (ed.), *Workers and their wages*, AEI Press, Washington DC.
- Koenker R. and Bassett G., 1978, "Regression quantiles", *Econometrica*, 46, 33-50.
- Lewbel, A., 2001 "Selection model and conditional treatment effects, including endogenous regressors", mimeo, Boston College.
- Machado, J.A.F. and Mata, J., 2005, "Counterfactual decomposition of changes in wage distributions using quantile regression", *Journal of Applied Econometrics*, 20 (4), 445-465.
- Munroe A., 1988, The measurement of racial and other forms of discrimination, Discussion paper in economics n. 148 (University of Stirling, Stirling).



- Oaxaca, R., 1973, "Male-female wage differentials in urban labor markets", *International Economic Review*, 14 (3), 693-709.
- Olivetti, C. and Petrongolo, B., 2005, "Unequal pay or unequal employment? A cross country analysis of gender gaps", *CEP Discussion paper*, 711.
- Overman, H.G. and Ioannides, Y.M., 2001, "Cross sectional evolution of the US city size distribution", *Journal of Urban Economics*, 49, 543-566.
- Quah, D.T., 1996, "Convergence empirics across economies with (some) capital mobility", *Journal of Economic Growth*, 1, 95-124.
- Quah, D.T., 1997, "Empirics for growth and distribution: stratification, polarization and convergence clubs", *Journal of Economic Growth*, 2, 27-59.
- Rosenblatt, M., 1971, "Curve estimates", *Annals of Mathematical Statistics*, 42, 1815-1842.
- Rosti, L., 2006, "La segregazione occupazionale in Italia", in Simonazzi A. (ed.) *Questioni di genere, questioni di politica*, Roma, Carocci.
- Rustichelli E., 2005, "I differenziali retributivi di genere", in Battistoni L. (Ed.), *I numeri delle donne 2005*, Quaderni Spinn, Ministero del Lavoro, 79-106.
- Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Stokey, N.L. and Lucas, R.E. Jr. (with Prescott E.C.), 1989, *Recursive methods in economic dynamics*, Harvard University Press, Cambridge MA.
- Thurow, L.C., 1975, *Generating Inequality*, Basic Books, New York.
- Villa P., 2006, "The gender pay gap in Italy", External report for the EU Directorate-general Employment and Social Affairs, Unit G1 "Equality between women and men".

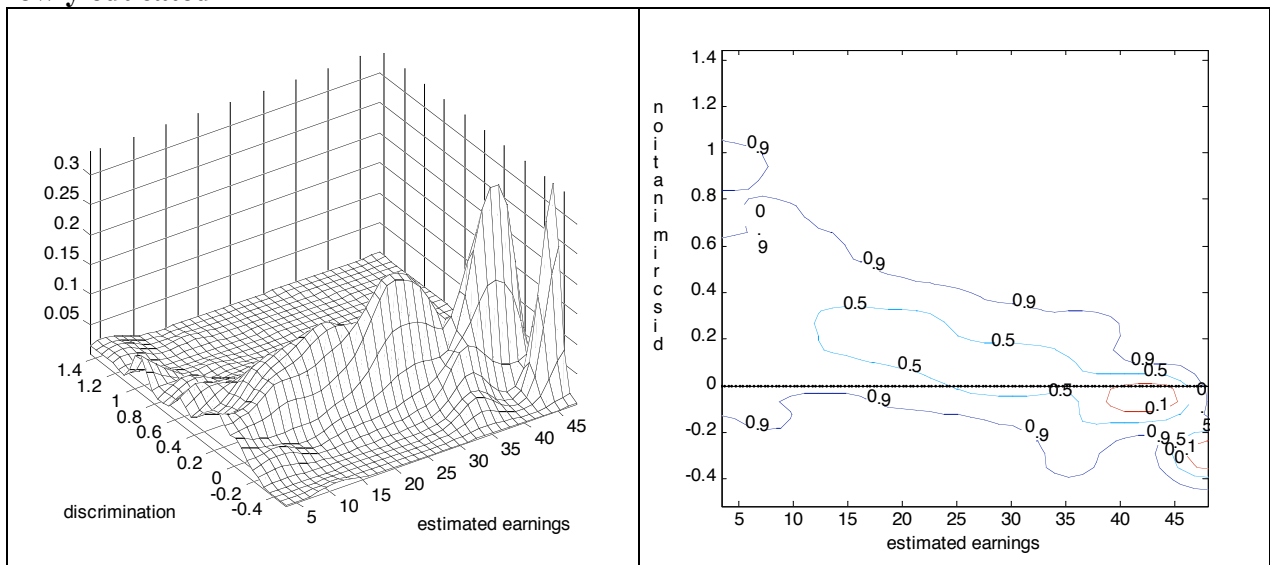
## Figures and Tables

**Table 1**  
**Incidence of women's discrimination**

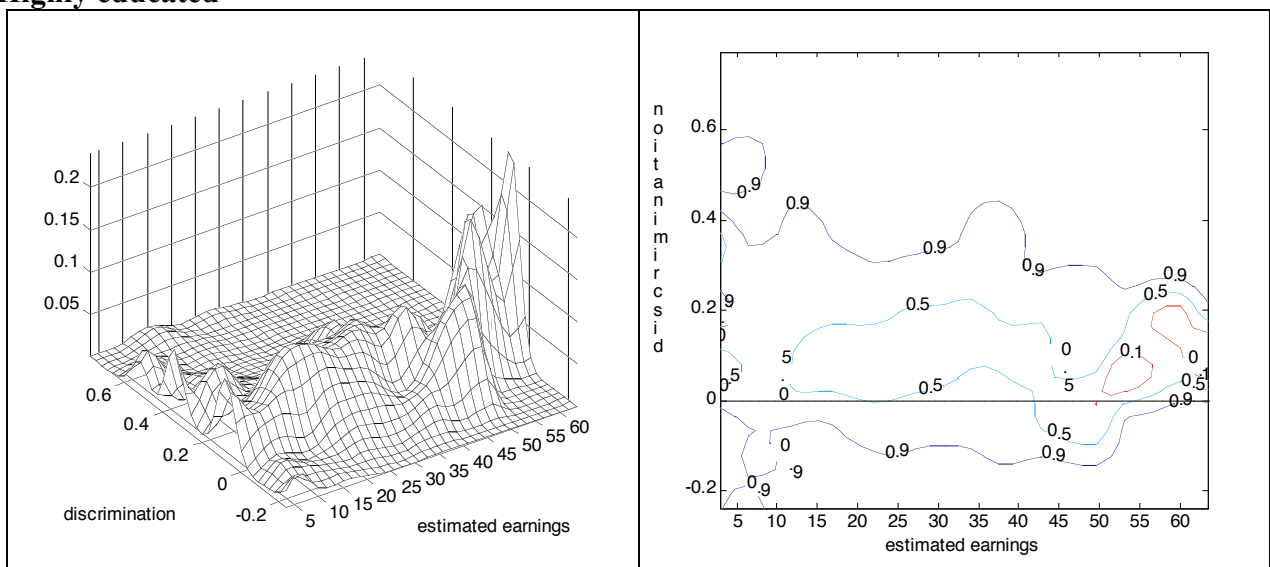
<b>Women's discrimination</b>		<b>Education Level</b>	
against / in favour	relative to estimated wage	High	Low
against	over 10%	52.47 %	48.02 %
	between 5% and 10%	16.26 %	9.11 %
	between 0 and 5%	12.94 %	8.94 %
in favour		18.34 %	33.92 %

**Figure 1**  
**Probability density functions of discrimination conditional to estimated earnings**

**Lowly educated**



**Highly educated**

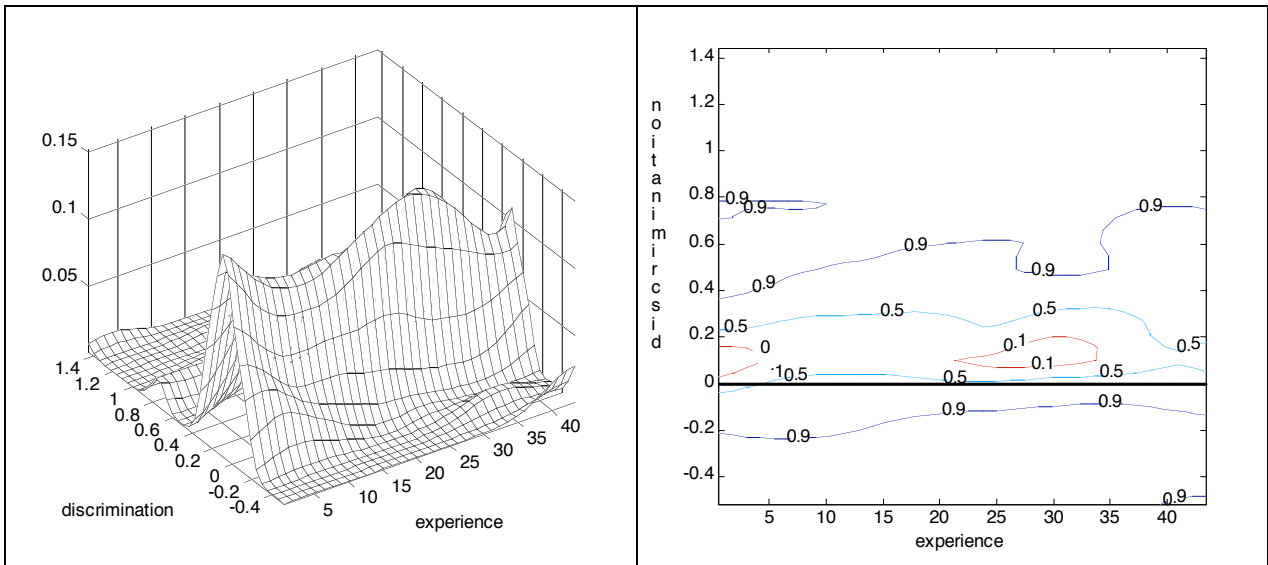



---

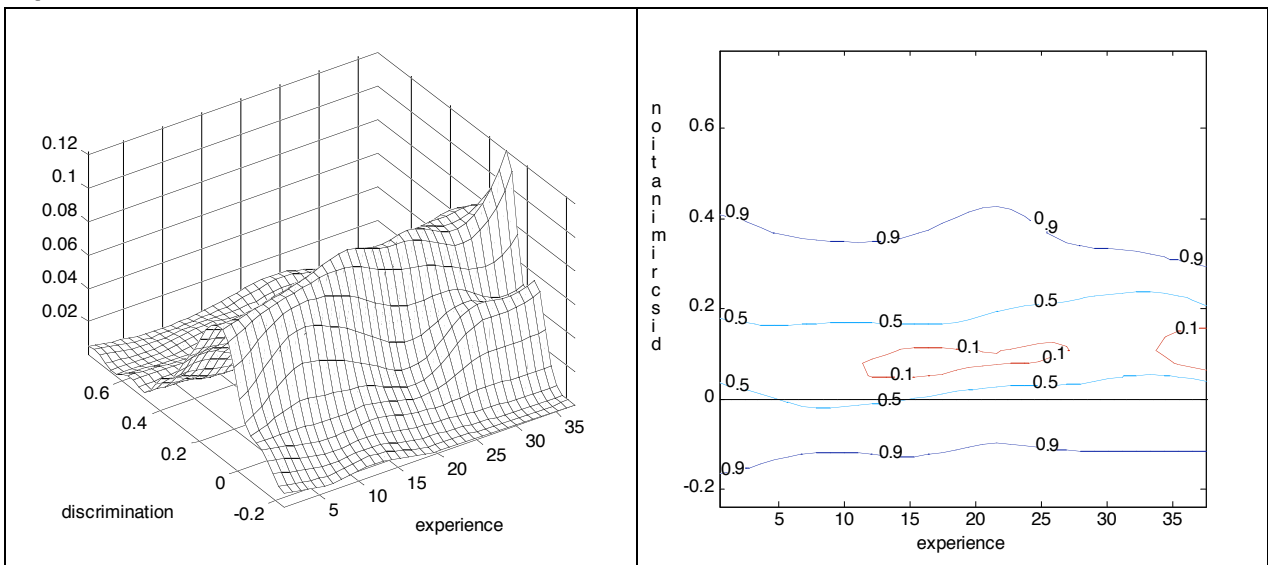
Notes:        Estimated earnings are expressed in thousand euros.  
               Discrimination is expressed in percentage terms with respect to estimated earnings.  
               Estimates use a Gaussian kernel with bandwidth chosen optimally (Silverman 1986).

**Figure 2**  
**Probability density functions of discrimination conditional to (potential) working experience**

**Lowly educated**



**Highly educated**

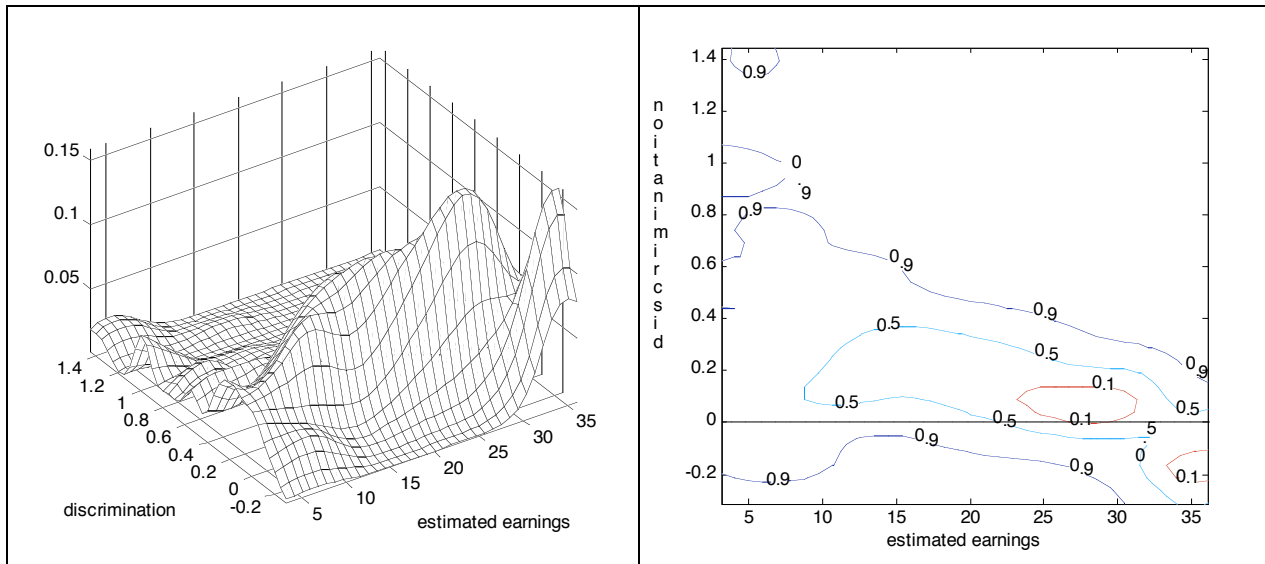


Notes: Discrimination is expressed in percentage terms with respect to estimated earnings.

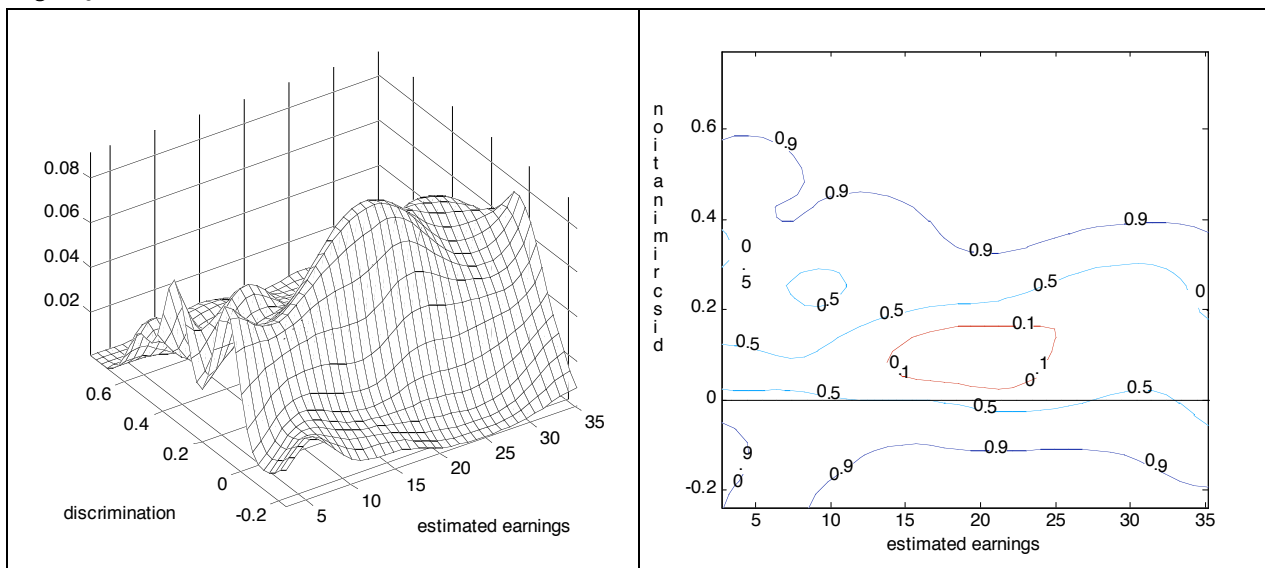
Estimates use a Gaussian kernel with bandwidth chosen optimally (Silverman 1986).

**Figure 3**  
**Probability density functions of discrimination conditional to estimated earnings**  
**(for tenure of 10 years or less)**

**Lowly educated**



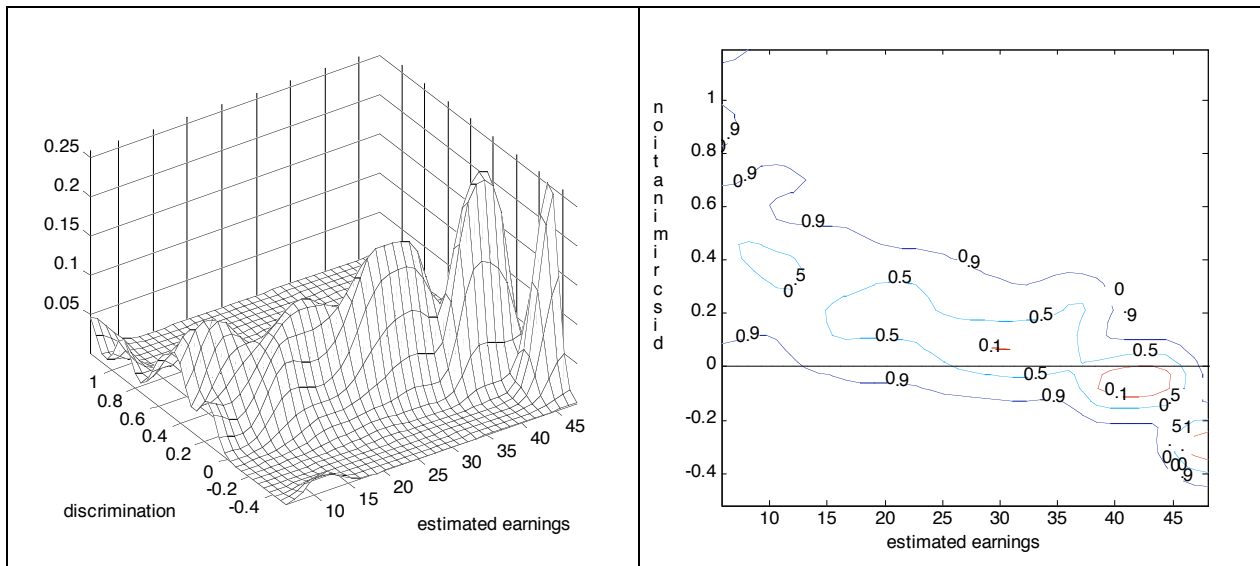
**Highly educated**



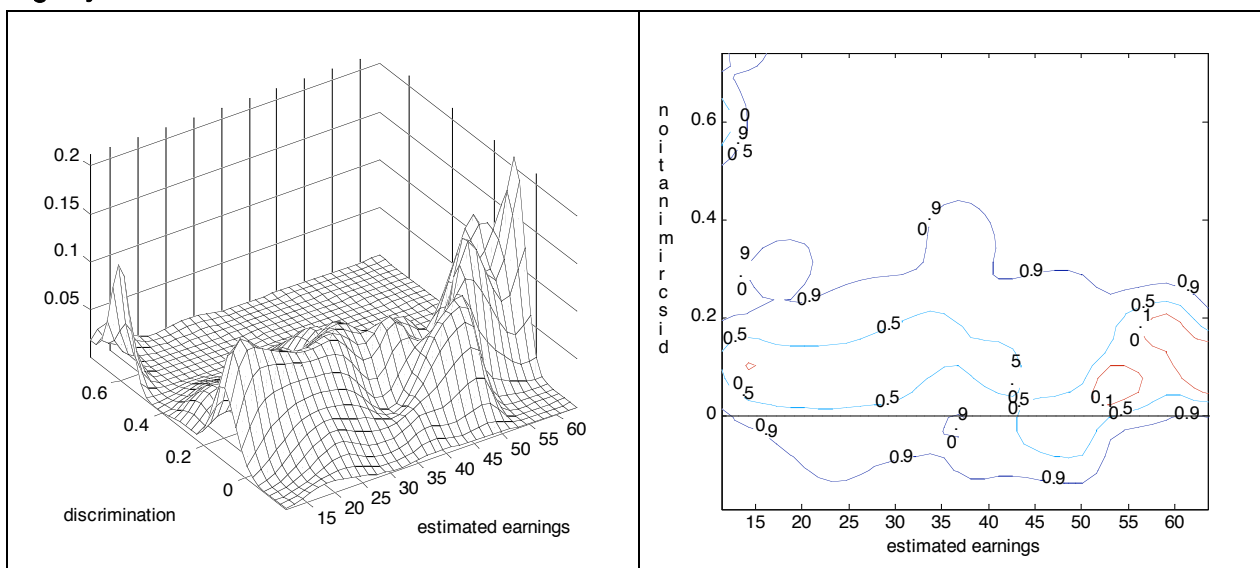
Notes: Estimated earnings are expressed in thousand euros.  
 Discrimination is expressed in percentage terms with respect to estimated earnings.  
 Estimates use a Gaussian kernel with bandwidth chosen optimally (Silverman 1986).

**Figure 4**  
**Probability density functions of discrimination conditional to estimated earnings**  
**(for tenure of 11 years or more)**

**Lowly educated**



**Highly educated**



Notes: Estimated earnings are expressed in thousand euros.

Discrimination is expressed in percentage terms with respect to estimated earnings.

Estimates use a Gaussian kernel with bandwidth chosen optimally (Silverman 1986).

## Appendix

Table A1. Sample descriptive statistics. Employees 16-65 years old.

	Highly educated				Lowly educated			
	Women		Men		Women		Men	
	Average	St.Dev.	Average	St.Dev.	Average	St.Dev.	Average	St.Dev.
Log yearlywage	9.9744	0.54	10.24218	0.52	9.6396	0.58	9.9861	0.54
Log months worked	2.3972	0.34	2.4328	0.24	2.3403	0.37	2.3923	0.33
Log hours worked	3.4853	0.30	3.6497	0.20	3.5153	0.33	3.6875	0.19
Upper-secondary education	0.8236	0.38	0.8072	0.39	-	-	-	-
University education	0.1763	0.38	0.1928	0.30	-	-	-	-
Experience	14.3949	9.96	15.3319	10.59	20.9120	12.33	19.8974	12.048
Experience squared	306.2994	353.13	347.1322	382.01	589.1104	551.62	540.9266	530.81
Experience*Children	5.5667	8.15	6.0073	9.24	5.1966	9.17	6.8674	10.31
Average supervisory level	0.1240	0.33	0.2056	0.40	0.0764	0.27	0.1013	0.30
High supervisory level	0.0664	0.25	0.1648	0.37	0.0310	0.17	0.0476	0.21
<i>Reference group:</i>								
<i>Tenure 0-5 years</i>								
Tenure 6-10 years	0.1569	0.36	0.1567	0.36	0.1288	0.33	0.1143	0.32
Tenure 11-15 years	0.1109	0.31	0.1099	0.31	0.1171	0.32	0.0952	0.29
Tenure more than 15 years	0.3261	0.47	0.3562	0.48	0.3115	0.46	0.3324	0.47
Public Sector	0.5277	0.50	0.3928	0.49	0.1832	0.38	0.2298	0.42
Industry	0.1324	0.34	0.3213	0.47	0.3175	0.47	0.4868	0.50
Services	0.8641	0.34	0.6531	0.48	0.6199	0.49	0.4429	0.50
<i>Reference group:</i>								
<i>Full-time</i>								
Part-time	0.0747	0.26	0.0176	0.13	0.1377	0.34	0.0154	0.12
Fixed-term or short-term contract	0.0770	0.27	0.0565	0.23	0.1152	0.32	0.0739	0.26
Other contract	0.0332	0.18	0.0283	0.17	0.0905	0.29	0.0800	0.27
<i>Reference group:</i>								
<i>Firm size: 5-19 employees</i>								
Firm size: 20-49 employees	0.1697	0.38	0.1889	0.39	0.1479	0.35	0.1313	0.34
Firm size 50-99 employees	0.1231	0.33	0.1192	0.32	0.0817	0.27	0.0784	0.27
Firm size 100-499 employees	0.1469	0.35	0.1687	0.37	0.1015	0.30	0.1048	0.31
Firm size: more than 500 employees	0.0862	0.28	0.1507	0.36	0.0618	0.24	0.0875	0.28
North-west	0.0984	0.30	0.0949	0.29	0.0631	0.24	0.0646	0.25
North-east	0.0984	0.30	0.0931	0.29	0.0987	0.30	0.0771	0.27
South and Islands	0.3207	0.47	0.3236	0.467	0.4115	0.49	0.4074	0.49

*Source: Descriptive statistics on ECHP 2001 sample*

Table A2a. Quantile regressions – Highly educated women 16-65 years old

Dep. variable: log income from work	Q10	Q25	Q50	Q75	Q90
Log months worked	1.191*** (6.09)	1.114*** (7.18)	1.128*** (10.82)	.992 *** (10.41)	.766*** (4.60)
Log hours worked	.339*** (2.82)	.144* (1.68)	.167*** (3.56)	.216*** (3.32)	.311*** (3.28)
Married/cohabitating	-.002 (-0.06)	-.011 (-0.43)	-.004 (-0.19)	.010 (0.35)	.007 (0.24)
University education	.033 (0.77)	.075** (2.39)	.114*** (4.49)	.143*** (4.33)	.223*** (4.36)
Experience	.026*** (2.84)	.013** (2.51)	.007* (1.76)	.000 (0.00)	.001 (0.21)
Squared experience	-.000*** (-2.31)	-.000 (-1.39)	-.000 (-0.37)	.000 (0.65)	-.000 (-0.45)
Experience*Children	-.003* (-1.69)	-.002 (-1.62)	-.002* (-1.72)	-.000 (-0.05)	-.000 (-0.03)
Average supervisory level	.094** (2.23)	.077** (2.51)	.092*** (3.91)	.063* (1.83)	.046* (1.06)
High supervisory level	.110** (2.17)	.038 (0.73)	.148*** (2.78)	.195*** (2.54)	.280*** (3.96)
Tenure 6-10 years	.099 (1.50)	.073* (1.89)	.075*** (3.08)	.048* (1.67)	.032 (0.75)
Tenure 11-15 years	.047 (0.56)	.088** (2.02)	.102*** (3.32)	.067** (2.16)	.084 (1.42)
Tenure more than 15 years	.075 (0.91)	.088** (2.28)	.059** (2.10)	.112*** (2.51)	.208*** (3.40)
Public sector	.163*** (3.01)	.097*** (3.36)	.056** (2.15)	-.018 (-0.51)	-.035 (-0.84)
Agriculture	-.611 (-1.38)	-.1101** (-2.00)	.067 (0.12)	.061 (0.12)	-.087 (-0.17)
Services	-.066 (-0.87)	-.076** (-2.04)	-.025 (-0.79)	.009 (0.19)	.004 (0.08)
Part-time	-.246** (-2.10)	-.347*** (-4.44)	-.308*** (-6.86)	-.199*** (-2.82)	-.065 (-0.61)
Fixed-term or short-term contract	-.241** (-1.95)	.001 (0.01)	-.058 (-1.62)	-.097** (-2.37)	-.109* (-1.84)
Other type of contract*	-.279 (-1.14)	-.207 (-1.03)	-.100 (-0.93)	-.080 (-0.67)	-.051 (-0.44)
Firm size: 5-19 employees	.056 (0.78)	.029 (0.83)	.046* (1.88)	.051 (1.51)	.123** (2.48)
Firm size: 20-49 employees	.101 (1.51)	.033 (0.97)	.064** (2.31)	.094** (2.10)	.138*** (2.68)
Firm size 50-99 employees	.121 (1.53)	.051 (1.26)	.079*** (2.77)	.070* (1.69)	.102** (1.98)
Firm size 100-499 employees	.053 (0.73)	.028 (0.75)	.064** (2.49)	.083* (1.87)	.130** (2.23)
Firm size: more than 500 employees	.132 (1.49)	.058 (1.20)	.122*** (3.77)	.115*** (2.51)	.215*** (3.29)
North-west	.068 (1.17)	.047 (1.16)	.058** (1.92)	.080* (1.93)	.124*** (2.65)
North-east	.131*** (3.22)	.082*** (3.23)	.030 (1.42)	.012 (0.38)	.038 (0.77)
South and Islands	-.077* (-1.84)	-.041 (-1.25)	-.035* (-1.69)	.005 (0.18)	.008 (0.23)
Constant	5.216*** (8.02)	6.250*** (11.75)	6.264*** (19.41)	6.544*** (18.58)	6.828*** (14.18)
R <sup>2</sup>	.50	.42	.34	.28	.30

Observations: 870. t-values in brackets. \*\*\* Significant 1%. \*\* Significant at 5%. \* significant at 10%



Table A2b. Quantile regressions – Low educated women 16-65 years old

Dep. variable: log income from work	Q10	Q25	Q50	Q75	Q90
Log months worked	.823*** (6.38)	.905*** (8.20)	.948*** (9.82)	.872*** (5.96)	.780*** (4.08)
Log hours worked	.502*** (2.68)	.585*** (4.39)	.550*** (4.49)	.606*** (3.30)	.384* (1.82)
Married/cohabitating	.128* (1.89)	-.011 (-0.22)	-.019 (-0.55)	-.044 (-0.97)	-.053 (-0.94)
Experience	.014 (1.04)	.013 (1.32)	.010 (1.50)	.012 (1.27)	.022** (2.07)
Squared experience	-.000 (-1.25)	-.000 (-1.13)	-.000 (-1.24)	-.000 (-1.02)	-.000* (-1.91)
Experience*Children	-.007 (-1.41)	-.002 (-0.71)	-.001 (-0.64)	-.000 (-0.09)	-.004 (-1.27)
Average supervisory level	.200** (2.26)	.130** (2.37)	.102** (2.06)	.053 (0.68)	.116 (1.02)
High supervisory level	.227 (1.54)	.173 (1.41)	.164 (1.20)	.189 (1.11)	.397*** (2.56)
Tenure 6-10 years	.000 (0.00)	-.070 (-1.11)	-.063 (-1.02)	.017 (0.24)	-.010 (-0.12)
Tenure 11-15 years	.191* (1.74)	.012 (0.16)	-.036 (-0.62)	-.007 (-0.11)	-.074 (-0.88)
Tenure more than 15 years	.146 (1.32)	.010 (0.15)	-.007 (-0.13)	.036 (0.54)	.060 (0.74)
Public sector	.031 (0.34)	.064 (1.05)	.049 (1.05)	-.018 (-0.29)	-.061 (-0.77)
Agriculture	-.285 (-0.52)	-.142 (-0.36)	-.174 (-0.47)	-.014 (-0.06)	-.262 (-1.13)
Services	-.08 (-0.77)	.062 (0.82)	.051* (0.87)	.059 (0.84)	.088* (1.10)
Part-time	-.297* (-1.63)	-.282** (-2.20)	-.126 (-1.24)	-.069 (-0.59)	-.133 (-0.85)
Fixed-term or short-term contract	-.048 (-0.31)	-.064 (-0.58)	-.072* (-0.93)	-.086 (-0.83)	-.090 (-0.76)
Other type of contract*	-.267 (-1.49)	-.258* (-1.82)	-.328** (-2.18)	-.106 (-0.74)	-.141 (-1.11)
Firm size: 5-19 employees	.207* (1.66)	.179** (1.99)	.164*** (2.65)	.089 (1.31)	.099 (1.13)
Firm size: 20-49 employees	.312** (2.45)	.152* (1.70)	.186*** (2.80)	.099 (1.37)	.036 (0.38)
Firm size 50-99 employees	.262* (1.66)	.295*** (2.70)	.296*** (4.01)	.161** (2.05)	.145 (1.57)
Firm size 100-499 employees	.261** (1.93)	.191* (1.64)	.263*** (2.89)	.183* (1.83)	.250** (2.39)
Firm size: more than 500 employees	.381*** (2.61)	.293*** (2.77)	.284*** (4.05)	.198** (2.27)	.182* (1.63)
North-west	.035 (0.48)	.010 (0.20)	-.018 (-0.33)	.007 (0.10)	-.024 (-0.30)
North-east	.040 (0.73)	-.020 (-0.40)	-.029 (-0.63)	.028 (0.40)	.125 (1.55)
South and Islands	-.129 (-1.07)	-.170* (-1.67)	-.049 (-0.53)	.040 (0.54)	.052 (0.72)
Constant	5.360*** (6.05)	5.035*** (7.81)	5.246*** (9.13)	5.360*** (5.77)	6.448*** (6.23)
R <sup>2</sup>	.62	.55	.44	.37	.39

Observations: 318. t-values in brackets. \*\*\* Significant 1%. \*\* Significant at 5%. \* significant at 10%

Table A3a. Quantile regressions – Highly educated men 16-65 years old

Dep. variable: log income from work	Q10	Q25	Q50	Q75	Q90
Log months worked	1.363*** (4.86)	1.107*** (11.88)	1.099*** (18.91)	1.123*** (13.66)	1.061*** (7.32)
Log hours worked	.190* (1.75)	.208* (1.71)	.520*** (3.98)	.663*** (6.69)	.464*** (3.21)
Married/cohabitating	.126*** (2.85)	.074** (2.04)	.061** (1.95)	.099*** (2.83)	.111** (2.03)
University education	.184*** (4.46)	.189*** (4.73)	.216*** (5.15)	.269*** (5.11)	.241*** (3.57)
Experience	.008 (1.19)	.008* (1.30)	.009* (1.72)	.011** (1.90)	.002 (0.22)
Squared experience	-.000 (-0.20)	-.000 (-0.61)	-.000 (-0.85)	-.000 (-0.82)	.000 (0.49)
Experience*Children	.002 (1.57)	.003*** (2.79)	-.001 (0.75)	.001 (0.51)	.001 (0.59)
Average supervisory level	.095*** (3.56)	.086*** (3.74)	.092*** (3.37)	.100*** (2.94)	.082* (1.88)
High supervisory level	.126*** (2.78)	.152*** (4.05)	.234*** (5.83)	.247*** (6.26)	.289*** (4.31)
Tenure 6-10 years	.043 (0.87)	.024* (0.61)	.047 (1.41)	.084** (2.22)	.084 (1.42)
Tenure 11-15 years	.000 (0.00)	.019 (0.40)	.039 (1.00)	.047 (1.05)	.059 (0.84)
Tenure more than 15 years	.061 (1.13)	.071 (1.59)	.087** (2.08)	.069 (1.46)	.159** (2.14)
Public sector	-.023 (-0.52)	-.006 (-0.18)	.009 (0.31)	.001 (0.03)	-.017 (-0.39)
Agriculture	-.174 (-1.43)	-.100 (-0.93)	-.165** (-1.98)	-.087 (-0.92)	-.022 (-0.13)
Services	-.031 (-0.62)	-.035 (-0.99)	.032 (1.16)	.046 (1.59)	.010* (0.21)
Part-time	-.360 (-1.51)	-.224 (-0.92)	.274 (0.96)	.270 (0.99)	.301 (1.08)
Fixed-term or short-term contract	-.290** (-2.13)	-.106 (-1.16)	-.056 (-1.20)	-.038 (-0.58)	-.027 (0.21)
Other type of contract*	-.668 (-1.54)	-.118 (-0.61)	-.028 (-0.27)	.055 (0.36)	.151 (0.57)
Firm size: 5-19 employees	.132 (1.44)	.088** (1.94)	.033 (0.99)	.006 (0.13)	.043 (0.69)
Firm size: 20-49 employees	.188* (1.88)	.111** (2.22)	.047 (1.19)	.027 (0.65)	-.035 (-0.64)
Firm size 50-99 employees	.215** (2.16)	.187*** (3.57)	.071* (1.85)	.037 (0.73)	-.017 (0.25)
Firm size 100-499 employees	.205** (2.19)	.133*** (2.84)	.097*** (2.65)	.064 (1.42)	.003 (0.05)
Firm size: more than 500 employees	.184** (1.93)	.164*** (3.09)	.136*** (3.50)	.059 (1.22)	-.025 (0.43)
North-west	.014 (0.27)	-.004 (-0.10)	.007 (0.21)	-.046 (-1.07)	-.001 (-0.02)
North-east	.053** (1.30)	.047 (1.54)	.037 (1.18)	.000 (-0.01)	.013 (0.21)
South and Islands	-.043 (-1.08)	-.068** (-2.38)	-.052** (-2.03)	-.074** (-2.49)	-.103** (-2.46)
Constant	5.428*** (7.21)	6.223*** (12.44)	5.171*** (10.53)	4.672*** (11.58)	5.815*** (8.69)
R <sup>2</sup>	.41	.34	.34	.39	.44

Observations: 1035. t-values in brackets. \*\*\* Significant 1%. \*\* Significant at 5%. \* significant at 10%

Table A3b. Quantile regressions – Lowly educated men 16-65 years old

Dep. variable: log income from work	Q10	Q25	Q50	Q75	Q90
Log months worked	1.095*** (8.00)	1.164*** (11.77)	1.034*** (12.36)	1.055*** (10.08)	.830*** (2.85)
Log hours worked	.232* (1.74)	.211* (1.84)	.191* (1.71)	.318*** (3.36)	.483*** (3.89)
Married/cohabitating	.074 (0.92)	.101** (2.46)	.043 (1.43)	.052** (2.07)	.027 (0.51)
Experience	.018** (2.01)	.011** (2.10)	.008** (2.03)	.006 (1.52)	.013* (1.67)
Squared experience	-.000 (-1.51)	-.000* (-1.81)	-.000* (-1.61)	-.000 (-1.05)	-.000 (-1.32)
Experience*Children	-.000 (-0.00)	.000 (0.35)	.001 (1.00)	.003** (2.07)	.003 (1.59)
Average supervisory level	.200*** (3.28)	.104*** (2.82)	.118*** (3.34)	.070** (2.37)	.008 (0.15)
High supervisory level	.019 (0.07)	.062 (0.92)	.108* (1.70)	.114** (2.29)	.078 (1.32)
Tenure 6-10 years	.190** (2.46)	.097*** (3.19)	.069*** (2.51)	-.007 (-0.23)	-.059 (-1.31)
Tenure 11-15 years	.063 (0.65)	.090* (1.85)	.107*** (3.18)	.020 (0.60)	.004 (0.07)
Tenure more than 15 years	.110 (1.32)	.105*** (2.95)	.111*** (3.13)	.081** (2.42)	.044 (0.83)
Public sector	.055 (0.98)	.024 (0.61)	-.016 (-0.48)	-.024 (-0.61)	-.053 (-0.95)
Agriculture	.152 (0.87)	-.054 (-0.77)	-.031 (-0.60)	-.062* (-0.77)	.058 (0.33)
Services	-.001 (-0.02)	-.036 (-0.92)	.008 (0.24)	.024 (0.68)	.023 (0.44)
Part-time	-.419 (-1.50)	-.313 (-1.41)	-.036* (-1.88)	-.232 (-0.95)	.236 (0.78)
Fixed-term or short-term contract	-.202* (-1.71)	-.145 (-1.29)	-.037 (-0.58)	-.024 (-0.25)	.168 (1.19)
Other type of contract*	-.483** (-2.16)	-.186*** (-2.02)	-.164*** (-3.14)	-.173*** (-2.82)	-.180 (-0.82)
Firm size: 5-19 employees	.058 (1.19)	.100*** (2.85)	.073** (2.34)	.037 (1.42)	.019 (0.41)
Firm size: 20-49 employees	.022 (0.32)	.065 (1.52)	.051* (1.79)	.048 (1.26)	.058 (1.22)
Firm size 50-99 employees	.101 (1.26)	.140*** (3.42)	.110*** (3.13)	.088* (1.74)	.146 (1.60)
Firm size 100-499 employees	.070 (0.97)	.089* (1.84)	.158*** (3.84)	.123*** (3.23)	.143** (2.51)
Firm size: more than 500 employees	.109 (1.51)	.162*** (3.23)	.161*** (3.90)	.145*** (3.36)	.118* (1.75)
North-west	.109* (1.89)	.042 (1.00)	-.010 (-0.23)	.013 (0.27)	-.055 (0.49)
North-east	.115* (1.82)	.068 (1.58)	.087** (2.20)	.048 (1.28)	.044 (0.67)
South and Islands	-.012 (-0.20)	-.028 (-0.90)	-.011 (-0.47)	-.002 (-0.09)	-.028* (-0.81)
Constant	5.762*** (8.84)	5.938*** (11.47)	6.523*** (14.95)	6.200*** (15.11)	6.274*** (6.67)
R <sup>2</sup>	.47	.39	.32	.27	.21

Observations: 729. t-values in brackets. \*\*\* Significant 1%. \*\* Significant at 5%. \* significant at 10%