Group versus individual discrimination among young workers: a distributional approach

Donata Favaro*

Stefano Magrini§

Università Ca' Foscari di Venezia – Dipartimento di Scienze Economiche

Nota di Lavoro 02/2005

February 2005

Abstract

We evaluate the gender wage gap and the unexplained gender wage differential for workers 15-29 year old during the period 1990-1997, using a particularly rich set of data from the Italian Social Security System covering all individuals in the labour markets of two Italian provinces.

We estimate separate earnings functions for men and women correcting for endogeneity of education and we evaluate gender discrimination by studying the entire distribution of the unexplained wage gap as suggested by Jenkins (1994). We evaluate discrimination against females by means of bivariate density functions. This innovation makes it possible to condition the density distribution on the marginal distribution of any characteristic and to evaluate more precisely the existence of group and individual discrimination. Our analysis suggests that discrimination is not evenly distributed among women, in relation to their characteristics; in particular, there is evidence of lower discrimination against highly educated females. Moreover in 1997, compared to 1990, discrimination increased in a appreciable way, affecting human capital rich females more significantly.

While our work is based in a very local context the richness of the data and the methodological innovation give the results a wider application.

The research forms part of a project on "Employment and Unemployment in Italy: Measurement and Modelling Issues" supported by the Italian Ministry for Education and Scientific Research.

* University of Padua, Department of Economics, Via del Santo 33, 35123 Padua, Italy.
Email: donata.favaro@unipd.it

§ University Ca' Foscari of Venice, Fondamenta S.Giobbe, Cannaregio 873, 30121, Venice, Italy.
Email: s.magrini@unive.it

1. Introduction

In this paper we study wage differentials between young male and female workers.

Only a few previous studies deal with the analysis of the wage gap among young people. In Wood et al. (1993), the gender wage gap between graduates of the University of Michigan Law School classes of 1972-1975 was estimated to be equal to 40 percent of the male wage 15 years from entering the labour market. This result was partly due to different choices in the hours of work and in worker qualifications by gender; but even controlling for these factors, women were found to earn 20 percent less than men. Several other studies on young workers estimate lower levels of wage differentials; however, many of these studies were not able to control for actual experience because of lack of information and results are therefore biased. Loprest (1992), using a sample of 18-25 year old women and men drawn from the U.S. National Labour Survey, estimated an entry wage gap of 11 percentage points. Dolton and Makepeace (1987) found a 7 percent wage gap for a sample of U.K. graduates in 1970. The research on the gender pay gap in the U.S. generally explains the wage differential in terms of a lower female level of effective work experience compared to men and relatively higher returns to work experience for men. Even job changes appear to favour men compared to women in the first years of their careers [Loprest (1992), Polachek and Robst (2001)]. Kunze (2002), employing an administrative dataset for Germany, shows that the wage gap among young skilled full-time workers was around 23% across the early years of a career.

The analysis of discrimination generally adopts the Oaxaca (1973) and Blinder (1973) methodology. Although recent extensions of this approach[1] control for some distributional aspects, the methodology is fundamentally based on the analysis of the 'average' level of discrimination. Following Jenkins (1994) we propose analysing wage differentials using the complete distribution of earnings estimated at the individual level. The distributional approach makes it possible to disentangle discrimination into its group component and its individual dimension due to differences in the characteristics that employers observe as a measure of productivity when setting wages.

Our analysis is based on a sample of 15 to 29 year olds working in two provinces – Treviso and Vicenza – of the Veneto region (North-Eastern Italy). Focusing on young workers is particularly interesting since it allows us to test whether there is any emerging discrimination even during the early years of a person's career, when there is only a little difference of experience between the sexes. The analysis is of particular interest in the case of our sample age group since female participation and employment rates among these younger workers are particularly high and the educational level of women is similar to that of their male counterparts. So entry level wages or wages in the early years of one's career should only be expected to show differentials if employers have information on people's productivity which is unmeasured in the data set or if they discriminate against women or other groups.

We evaluate the gender wage gap on the universe of 15-29 year old female and male private workers drawn from the administrative data set of the National Social Security

[1] This methodology has been extended in two different directions. Juhn, Murphy and Pierce (1991) have extended the Oaxaca decomposition by taking into account the residual distribution, useful when analysing the gender gap in a comparative setting or throughout a long temporal interval. Brown, Moon and Zoloth (1980) have extended the approach to endogenize the distribution of women across occupations.

System for the two provinces in two different years (1990 and 1997). The full data set contains the universe of workers employed in the private sector for the period 1976-1997. The availability of a time-series and the structure of the data give us precise information on work experience, firm-specific human capital (tenure in the actual firm) and make it possible to control for individual and occupational characteristics, economic sectors and local labour markets.

Our dataset allows the match of information on individual characteristics to information on the size and economic sector of firms. Then we are able to estimate earnings functions where we control for different human capital characteristics as the general experience accumulated in the labour marker and the specific experience acquired with the present employer; in addition we construct a proxy for the individual's educational level and we instrument that variable with a measure of the supply-side of the educational system. We also control for a firm's local labour market location.

In Section 2 we present the methodological approach based on the distributional analysis. In Section 3 we describe the dataset and some descriptive statistics. Earnings function estimates by gender are discussed in Section 4, while in Section 5 the distributional analysis of the unexplained wage gap is then illustrated and implemented. Section 6 concludes.

2. The Oaxaca-Blinder approach: remarks and suggestions

The empirical research on discrimination widely adopts the approach of Oaxaca-Blinder (1973)[2]. Gender discrimination is defined as the difference between the average wage women earn and the average wage they would earn if they were men, given the same characteristics. The methodology consists in estimating gender-separate 'earnings functions' explaining worker income as a function of individual and occupational characteristics. Then, the total estimated difference is decomposed into two terms: a first term representing productivity differences explained by individual characteristics and a second term explaining any earnings gaps in terms of differences in the remuneration of the characteristics. This second term represents the discriminatory component of the wage gap. Traditionally, the methodology then suggests the calculation of an index of discrimination defined as the average of the difference between the average reference wage – e.g. the wage that women would be estimated to earn if their characteristics were paid as much as men's – and the average estimated female earnings.

Although the aim of the Oaxaca-Blinder's analysis is to estimate the extent of discrimination throughout the distribution of individuals' characteristics, the evaluation of the wage differential is actually made in average terms. Therefore, the estimate of discrimination is effectively reduced to the average and the analysis is simplified to an analysis of group discrimination. The comparison of average wages between minority and majority groups makes the model equivalent to exact or non-stochastic models.

The neoclassical theories of discrimination can help us to understand the meaning of the empirical analysis of discrimination better. The theoretical approaches suggest distinguishing between group (or deterministic) discrimination and individual (or stochastic) discrimination (Cain, 1986). In both cases, discrimination may arise even

[2] For a presentation of the analysis of discrimination, we suggest the surveys of Antonji and Blank (1999) and Kunze (2002).

though workers are equally productive and have equal tastes for work. Group discrimination affects all workers belonging to the minority group equally. In contrast stochastic (or individual) discrimination allows for individual deviations within the particular group. The models of individual discrimination assume that employers do not have perfect information and do not know the true productivity of each worker; they set wages relying on the characteristics they observe, which although related to productivity do not perfectly represent productivity. The stochastic model does not predict group discrimination if employers set wages equal to the expected value of productivity and individuals are equally productive (or have equal productive capacity). Individual discrimination can arise either because of the different reliability of the indicator that employers observe for the two groups or because men have a higher probability of market work compared with women (Thurow 1975).

The deterministic and stochastic models have always been considered as complementary as they do not predict the same type of discrimination. Actually, the stochastic model can be generalised supposing that employers' behaviour on average, discriminates against the minority group; this can happen if employers, although setting wages on the base of the expected productivity, discriminate against the minority group by setting wages equal to the expected productivity less a discriminatory amount. This generalisation allows stochastic models to 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. Summing up the two components of the wage differential, we can observe whether individuals from the minority group are discriminated against equally or whether, depending on the distribution of the individual characteristics, the market differently discriminates between workers within the minority group.

The Oaxaca-Blinder methodology substantially limits the explanatory power of the regression analysis, which potentially allows us to identify the existence of individual discrimination along the distribution of the characteristics that employers use as indicators to set the level of wages. The first critical remarks relating to this approach go back to Dolton and Makepeace (1985) and Munroe (1988) who identified the shortcomings of the Oaxaca-Blinder methodology and the need to use the entire set of information contained in the distribution of earnings. Building on these contributions, Jenkins (1994) developed a method for analysing discrimination that made use of the complete information contained in the distributions of estimated and reference female earnings. Briefly, the method he proposed entails a comparison of the relative position of the Generalised Lorenz Curve (GLC) with respect to the Generalised Concentrations Curve (GCC)[3], 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

[3] In order to construct the GLC, after ordering female workers by ascending observed wage, one plots the cumulative estimated wage per capita $\sum_{i=1}^{k} \hat{y}_i / n_w$, against the cumulative sample share $p \equiv k / n_w$, for each k = 1, …, n_w (n_w is the total number of femal workers). Similarly, to summarise the no-discrimnation distribution using a GCC, one plots the cumulative reference wage per capita $\sum_{i=1}^{k} \hat{r}_i / n_w$ against p, for each k = 1, …, n_w, ensuring women are ordered exactly as for the GLC.

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.

Based on these considerations, we propose an alternative method that evaluates discrimination by focussing directly on the distributions of estimated and reference earnings, resorting to the estimate of stochastic kernels to describe their relationship; this method makes it possible to detect both group and individual components of the unexplained wage gap, and to assign a probability of occurrence to any level of discrimination at any point in the distribution with respect to a given characteristic.

There is a clear analogy to be made between the alternative approach we propose and the distributional approach to the analysis of cross-sectional economic convergence developed by Quah (1996, 1997). Within the convergence literature, the evolution over time of the cross-sectional distribution of per capita income is viewed as a stochastic process whose law of motion is studied by resorting to a stochastic kernel[4]. However, the definition of a stochastic kernel does not require the distributions under study to relate exclusively to per capita income nor, as Quah (1997) makes clear, to be sequential in time and can thus be used to relate any two distributions.

The approach can be summarised as follows. First, as in the traditional analysis, we estimate log wage equations for men and women as a function of their characteristics – X_m for men and X_f for women – and an error term:

$$log(W_{m_i}) = X_{m_i}\beta_m + \varepsilon_i \tag{1}$$

$$log(W_{f_i}) = X_{f_i}\beta_f + \varepsilon_i \tag{2}$$

Second, we consider the series of estimated female earnings:

$$\hat{y}_i = \exp(X_i^f b^f) \tag{3}$$

and the series of reference earnings, defined as the wages women would receive were they paid the same as men for their characteristics:

$$\hat{r}_i = \exp(X_i^f b^m) \tag{4}$$

for each individual *i* within the set of female workers *F*.

It is at this stage that our approach departs from the traditional one. The latter would evaluate discrimination as the average of the difference between estimated and reference

[4] Other examples of the use of stochastic processes in economics include stochastic growth models (see e.g. Stokey and Lucas (with Prescott), 1989) and models of income distribution dynamics (see e.g. Loury (1981) who use a stochastic process to model the dynamics of the earnings distribution among successive generations of workers.

earnings, expressed in log transformation: $\overline{X}_f(b_m - b_f)$. In contrast, to evaluate the incidence of discrimination we focus directly on the relationship between the cross-individual distributions of reference and estimated earnings. For this purpose, let *R* and *Y* denote respectively the distributions of reference and estimated earnings across female workers. Associated with these distributions there are two probability measures: ρ corresponds to the distribution of reference earnings and υ to the distribution of estimated earnings. Next, we introduce a relationship between the two distributions. The simplest way of modelling this relationship is the following:

$$\rho = T^*(\upsilon, u) = T_u^*(\upsilon) \tag{5}$$

where u is a sequence of disturbances, T^* an operator that maps the Cartesian product of the probability measure υ and the disturbances, and T_u^* absorbs the disturbance into the definition of the operator and encodes information on the relationship between the two distributions, i.e. on discrimination[5].

Notice that when the earnings space is discrete, i.e. the earnings are grouped into a set of non-overlapping classes, the measures ρ and υ can be represented by probability vectors and the operator in equation (5) simplifies into a probability matrix, *P*, whose rows and columns are indexed by the element of the discretisation, and whose generic element p_{ij} reports the probability of finding reference earnings that belong to class *j*, given that estimated earnings are in class *i*. More generally, when the continuous set of earnings values is retained, the operator T_u^* can be interpreted as a stochastic kernel, giving the probability density function (PDF) of $R = \hat{r}$ conditional on $Y = \hat{y}$, i.e. $f\,(\hat{r} \mid \hat{y})$.

From an operational point of view, the procedure to produce an estimate of the conditional distribution $f\,(\hat{r} \mid \hat{y})$ can be described as follows. First, we estimate non-parametrically the joint distribution of reference and estimated earnings using a two-dimensional kernel:

$$\hat{f}(\hat{r}, \hat{y}) = \frac{1}{n h_{\hat{r}} h_{\hat{y}}} \sum_{i=1}^{n} K\left(\frac{\hat{r} - \hat{r}_i}{h_{\hat{r}}}, \frac{\hat{y} - \hat{y}_i}{h_{\hat{y}}} \right) \tag{6}$$

where d_i = ($\hat{r}_i, \hat{y}_i$) is the *i*-th observation, d = ($\hat{r}, \hat{y}$) is a fixed point, *K*(.) is the kernel function, while $h_{\hat{r}}$ and $h_{\hat{y}}$ are the bandwidths[6]. Next, we estimate the marginal distribution $f\,(\hat{y})$ using a univariate kernel[7]. Finally, we obtain the estimate of $f\,(\hat{r} \mid \hat{y})$, the

[5] For a rigorous presentation of the underlying mathematics see Durlauf and Quah (1999) and references given there.

[6] To estimate the joint distribution, we have used a Gaussian kernel with bandwidth chosen optimally according to Silverman (1986).

[7] As in the bivariate case, we have used a Gaussian kernel with bandwidth chosen optimally (Silverman 1986). Note that Quah (1996) suggests estimating the marginal distribution by numerically integrating the joint distribution with respect to reference earnings. However, the asymptotic statistical properties of both estimators are identical and produce very similar results in practice (Overman and Ioannides, 2001).

distribution of reference earnings conditional on estimated earnings, by dividing the joint distribution by the marginal one[8]:

$$\hat{f}(\hat{r}|\hat{y}) = \frac{\hat{f}(\hat{y},\hat{r})}{\hat{f}(\hat{y})} \tag{8}$$

The incidence and direction of discrimination can thus be studied by analysing directly the shape of the three-dimensional plot of the stochastic kernel and of the corresponding two-dimensional contour plot. In particular, a probability mass stretching along the 45 degree line would imply the absence of discrimination as, for any level of estimated earnings, reference earnings would tend to assume the same value. In contrast, a probability mass lying consistently above the 45 degree line would indicate that, for any level of estimated earnings, reference earnings tended to exceed their estimated counterparts, thereby suggesting the existence of discrimination against female workers. Analogously, discrimination in favour of female workers would be signalled by a concentration of the probability mass below the 45 degree line, as reference earnings tended to be smaller than the corresponding levels of estimated earnings.

Finally, stochastic kernels can also be used in order to investigate the role of specific factors or worker's characteristics in shaping the distribution of discrimination. The actual way in which this can be accomplished depends on the nature of the characteristics. In general, these characteristics fall into two categories. They can be measured on a continuous space, as in the cases of the experience accumulated within the firm and of the general experience accumulated during the individual's working history or they can be represented as categorical or dummy variables, like the level of education. In the former case, we can exploit the fact that a stochastic kernel can be used to relate any two distributions, and study directly the relationship between the distribution of discrimination and the distribution of any characteristic. To do this we first calculate the level of discrimination for each female worker as the difference between reference and estimated earnings. Then, we denote the distribution of discrimination with D and the associated probability measure with δ. Similarly, we denote the distribution of the characteristic with C and its probability measure with γ. Similarly to what we did before, we model the relationship between the two distributions as

$$\delta = T^{*}(\gamma,\varepsilon) = T^{*}_{\varepsilon}(\varepsilon). \tag{9}$$

and T^{*}_{ε} is the stochastic kernel, giving the distribution of discrimination conditional on the distribution of the characteristic. Operationally, this can be estimated according to the same procedure described above. As for the interpretation of the results, the only difference is that now the absence of discrimination is represented by a concentration of the probability mass along the line running parallel to the characteristic's axis and in correspondence to a level of discrimination equal to zero.

Where the characteristics are measured by categorical or dummy variables, the conditional density function cannot be estimated as in equation (8). Suppose that the

---

[8] Under regularity conditions this represents a consistent estimator for the conditional distribution [Rosenblatt (1971), Silverman (1986), Quah (1996), Chen, Linton and Robinson (2001)].

characteristic of interest presents l levels and that the set of all individuals can thus be divided into l separate subsets. Denote with G_n the subset of individuals having a level n for the characteristic (with n = 1, …, l), and with $\hat{r}_{G_n} = \{\hat{r}_i : i \in G_n\}$ and $\hat{y}_{G_n} = \{\hat{y}_i : i \in G_n\}$ the observations on reference and estimated earnings corresponding to individuals in G_n. Then, modelling the relationship between reference and estimated earnings as in equation (5), we can estimate l different stochastic kernels. Each of these stochastic kernels shows the distribution of reference earnings conditional on the distribution of estimated earnings, for a given level of the characteristic. Moreover, direct comparisons between the estimates of the kernels for different levels of the characteristic indicate how the distribution of reference and estimated earnings is affected by changes in the level of the characteristic.

3. Data and descriptive statistics

We use administrative data made available from the National Social Security System (*Istituto Nazionale di Previdenza Sociale,* INPS) for the two North-Eastern Italian provinces of Treviso and Vicenza. The INPS database is composed of different archives containing information on the universe of workers who had at least one employment spell eligible for the social security insurance scheme in any year between 1976 and 1997[9] and on the universe of firms which paid any social security contribution during the same period.

The database makes it possible to match each worker to each different firm he or she used to work from the beginning of their working experience; this means we can merge personal data with information on gross earnings[10], type of job, sector and firm-size for each spell of employment. Since the dataset contains information on the exact number of weeks spent working every year with any particular employer, we are able to determine precisely both the effective general experience and the firm specific experience accumulated respectively in the labour market and in the firm, net of periods of absence from the labour market. This allows us to estimate separately the effect of the different components of human capital on the level of wages. We therefore define two different variables. The variable "tenure" summarises the accumulation of specific human capital with the current employer and is defined as the total number of weeks spent in the firm employing the worker at the time we observe his or her wage; the variable "experience" captures the dimension of the accumulated generic human capital and is expressed as the total number of weeks worked in the labour market since the beginning of the worker's working experience. So it includes work experience with previous employers other than the present firm. If our sample were made of all workers, measurement errors could arise in evaluating individual experience due to the left-censored nature of the dataset. However, since we are focusing on young workers aged 15-29 in 1990, the data available to us make it possible to calculate both the total period of working experience and the total length of experience in the firm so long as the individual was working at the time we observe the wage. The availability of this data is a particularly interesting and worth feature of our analysis. In contrast to most studies on the subject that rely on household surveys, we can

[9] For detailed information about the INPS archives (for the provinces of Treviso and Vicenza) see Occari et al. (1996).

[10] Gross earnings include social security contributions. We use the ISTAT Consumer Price Index (base=1995) to deflate earnings.

avoid making use of potential experience as a proxy for actual experience and to exactly determine tenure.

One of the drawbacks of this dataset is the lack of information on worker educational levels. To overcome this problem, we construct a proxy for the highest completed level of education using the age at which an individual first started work, identified by the first non-seasonal spell of work. What we mean by first non-seasonal spell is the first inscription in the INPS archive either for a period not shorter than 17 weeks or for a period between 4 and 17 weeks but not between May and September of any year (as it could otherwise be interpreted as a working spell compatible with school attendance). Having identified the age at which an individual first entered the labour market on a career basis we derive the educational level along the following criteria: we assign an upper-secondary educational level to individuals who entered the INPS archive between 19 and 24 years old. We assign a tertiary educational level – university education – if workers entered the dataset when older than 24 years[11]. We assign a level of education lower than upper-secondary if workers entered the INPS archives when younger than 19 years. The variable education we employ in the estimates is a categorical variable assuming a value of 1 when the educational level is lower than upper-secondary (or *compulsory diploma*[12]), a value of 2 if the estimated education level is equal to upper-secondary and a value of 3 for the university level[13].

We generate two samples of young workers for the years 1990 and 1997. We focus the analysis on individuals working the whole year in order to have a homogeneous sample and to exclude seasonal working spells; moreover we exclude managers because of female under-representation in those types of occupations. The analysis is carried out for workers in the manufacturing, commerce and credit-insurance sectors[14]. Foreigners are excluded from the sample because we are not able to instrument for education. Finally, following the debate on the risks involved using INPS data for the analysis of wage gaps, we include in our sample only workers with a number of days worked equal or higher than 280[15]. The final sample is then made of 71057 workers in 1990 and 64760 workers in 1997[16]. The INPS archives provide information on yearly gross earnings (gross of social contributions and income taxes) and the number of days worker per worker-spell. We therefore derive our dependent variable – daily earnings – dividing the yearly earnings by worked days in each year (thousands of lira at constant prices 1995).

Some interesting facts emerge from looking at the sample composition (Table A1). With respect to the educational level, most workers, independently of sex, have an educational level lower than upper-secondary, in both years**.** The difference between sexes is rather small, with 69% of women and 71% of men in 1990 (52% of women and 59% of men in 1997) having an educational level lower than upper-secondary education. However, the gender difference increased during the first years of the 1990s and in 1997 there were

[11] During the 1990s most of the Italian university curricula were lasting for 4 or 5 years. Therefore we summed a 5 year period to the age of the upper-secondary diploma.

[12] For the period to which the data relate the Italian leaving school age was set at 14.

[13] Because of the way it is constructed, the educational variable could present problems of measurement error. Therefore, we instrumented the variable, as discussed in the following paragraph.

[14] The economic sectors considered in the analysis are those included in the 3, 4, 6 and 8 one-digit ATECO 1981 ISTAT classification.

[15] INPS data on wages seem to be overestimated for people working only a few days in a year. For that reason. For the debate, see Ginzburg et al. (1998, 1999), Gavosto and Rossi (1999).

[16] The percentage of women is 47.6% in 1990 and 47% in 1997.

proportionately more men with less than an upper-secondary education. The proportion of women with at least an upper-secondary educational level sharply increased, in particularly with respect to the secondary level. It increased from 29% to 45%. The positive trend was less evident for men - increasing from a 27% to a 36%. The gender educational gap at the highest level also fell, although the difference in 1997 still amounted to 1.7 percentage points (2.7% for women against 4.4% for men). The variable "education" summarises the three different levels discussed above. In average terms, the educational level of women appears to be higher than the educational level of men, in both years, and to have risen during the first half of the 1990s.

Looking at other human capital characteristics rather small gender differences appear. Gender is particularly relevant in determining the type of experience that workers accumulate; women tend to accumulate more specific human capital than men with men tending to invest more in general experience. The average tenure accumulated is slightly higher for female workers than for men and the difference increased during the period. In contrast, men have on average a longer period of experience but the gap was reduced over the period.

Summarising the information about the human capital characteristics of our samples, we see that there are only small differences between the sexes. The most significant differences are in terms of occupational and sector concentration. Women are mainly concentrated in white-collar occupations, especially in 1997, in the textile sector and in firms of small to medium size. Men are more numerous in the metal and precision tools sector and in larger firms. Both groups have similar proportions of full-time workers, although the proportion of females who worked full-time fell slightly over the period. The proportion of *Contratti di Formazione Lavoro* (CFL; Working and training contracts)[17] and *Apprenticeship* contracts decreased for both groups; the former were dismissed late in the 1990s while the latter were substituted by the introduction of other types of contracts.

4. Empirical analysis

Before discussing the econometric analysis, we can look briefly at the male and female distributions of earnings (Table A2). Recent Italian studies [ITER (2001), Flabbi (2001)] have shown different earnings trends depending on the data used. Data from the Ministry of Finance show an increase in the wage gap from 29% in 1982 to 36% in 1996; data from the same source, but for the sub-sample of year-round[18] employees, show a smaller wage difference, however steadily increasing from 19% in 1982 to 21% in 1994. Data from the Bank of Italy for the whole distribution of earnings show an increase in the wage gap between men and women during the first years of the 90s with a much stronger worsening for manual female workers than for white-collar worker. Conversely, considering only year-round employees, data from the Bank of Italy show an increasing female relative wage. Table A2 reports a much smaller gender wage gap than those estimated using data from the Bank of Italy or from the Ministry of Finance. A comparison over time shows an increase in the wage difference along the entire distribution, except for the last quartile. Young females with high earnings experience lower wage gaps than their low wage

[17] The CFL were introduced in 1985 in order to improve the youth chance (aged less than 30) to get a job. The employer is provided with a rebate on the labour cost and a full exemption from firing costs.

[18] That means that they worked in the same job all year.

colleagues. However, once more, there exist sharp differences conditional on the type of occupation: the gender wage gap decreases along the whole distribution of white-collar workers, while it increased at all wage levels in the blue-collar distribution.

Our econometric analysis presents some methodological problems that have been widely discussed in the literature. One of the issues on which most of this discussion has focused is related to the reliability of the estimates of the female earnings function. Labour force participation and unemployment rates differ substantially between men and women and are generally lower for women than for men. As a result the coefficient estimates in the female wage equations could be biased because of a non-random selection of women who work. In particular, the returns to education could be overestimated. This empirical problem can be consistently solved by a two stage procedure where in the first stage a participation equation is estimated while in the second stage the earnings function is corrected by the inverse of the Mill's ratio derived from the first stage participation equation (Heckman, 1979). In order to identify the participation equation, however, specific variables have to be used; usually, the variables of interest are those related to marriage and maternity choices as well as to the household background; household characteristics as well as educational and occupational characteristics of each individual's parents are usually recognised as good variables in order to identify the participation choice.

Unfortunately, our dataset does not contain any information on the individual's background so we are not able to correct our estimates for selection into participation. However, the logic underlying the correction for participation is that female activity and employment rates substantially differ from those of males . This, however, is not the case in our analysis. Actually, stylised facts about the region our provinces belong to –Veneto - show rather similar activity rates between men and women, especially for people aged between 18 and 29. The activity rate of this group in Veneto in 2000 was 67% against 56% in Italy as a whole; the female activity rate was 65% against a male activity rate of 70% (respectively 50% and 62% in Italy)[19]. This suggests a substantive homogeneity of labour market participation with respect to the two sexes, at least in terms of this younger age group.

It is also relevant to bear in mind the extent to which the age motherhood has risen in recent decades. In the Veneto the average age of women at marriage and at the birth of their first child is higher than it is nationally; in 2000 the average age at first childbirth in the Veneto region was 30 years and 6 months. These facts suggest that the choice between maternity and participation does not fundamentally affect most women in the age group we are studying. Both considerations let us conclude that selection bias between the sexes in participationdoes not represent a real problem that could significantly influence our results.

A second econometric problem when applying the OLS methodology to the estimation of earnings functions arises because of the likely inconsistency of the estimate of education (Card, 2001)[20]. OLS estimates of returns to education can be inconsistent principally because of endogeneity problems; the educational variable can be correlated with some unobservable characteristics – such as ability or motivation – that are included in the error

[19] These data are taken from the Regione Veneto (2002), Primo Rapporto sulla Condizione Giovanile nel Veneto.

[20] For a selected review of Italian studies that estimate the returns to education see Brunello and Miniaci (1999). A more detailed but older survey is in Lucifora and Sestito (1994). Recent estimates of returns to education in Italy are in Brunello, Comi and Lucifora (2000).

term. It follows that OLS estimates are inconsistent because of the correlation of the schooling variable with the stochastic term of the earnings function. Inconsistency of OLS estimates can arise also because of measurement errors. In our case, this problem could be particularly relevant because of the way we have constructed our variable; our proxy for education could systematically overestimate the educational level of lower ability workers who may enter the labour market later and underestimate the educational level of more able workers. The use of instrumental variables is the standard solution to a problem such as this. The most recent studies on the relation between education and earnings follow the basic idea that institutional features of the education system can be used as credible instrumental variables for individual schooling outcomes. Supply-side sources of variation in schooling are used to help resolve identification problems on the demand side of the education market (Card, 2001). Sources of variation in the supply of schooling are usually represented by the minimum school-leaving age, by tuition costs or by the geographic proximity of schools.

In our study, however, we could not use a measure of schooling supply such as the minimum school-leaving age and tuition costs. The minimum school-leaving age is not a good instrument to capture the variability of the educational level because we are analysing workers subject to the same institutional constraints; tuition costs are not a good instrument because the Italian educational system, especially in the years covered by our data, was mainly dominated by a public sector which precluded differences in tuition fees. The relevant variable we chose to instrument the educational level is the geographical availability of schools. For each *Comune* (municipality) in the 'place of birth' field of the dataset, we calculated the number of upper-secondary institutes within the *Comune* itself and within the *Comuni* within a radius of 15 kilometres from the *Comune*'s centroid[21].

Regression results are reported in Table A3[22]. Our dependent variable is the natural log of daily earnings. In addition to the human capital variables we described above, we included other controls capturing the impact of the type of occupation (white-collar/blue-collar), the type of contract (full-time/part-time, apprenticeship and training contracts - CFL), economic sector[23] and firm-size[24]. We also included dummies for each local labour market system.

Estimates confirm the positive effect of any human capital characteristic on the level of daily wages, independently of sex. However, the returns to human capital significantly decreased over the period, especially for female workers. The estimates of the returns to education confirm the results of previous empirical analyses on Italian data; women receive higher returns to education than men (Brunello, Comi and Lucifora, 2000). This in both years, although in 1997 the returns are in general much lower than in 1990. Educational levels higher than lower secondary education were estimated to produce a 14 to 23 percent higher wage in 1990 and between 9% and 12% higher in 1997. In annual

---

[21] The source of the data on secondary institutes is the Ministry of Education, University and Research. Geographical availability has been worked out using ArcView GIS.

[22] We do not show the local labour market dummies (LLM); however we can provide the complete tables on request. The significant LLM dummies are different between the two genders and the estimated coefficient is often negative.

[23] We use the ISTAT definition (ATECO 1981).

[24] We also estimated a model specification with fixed effects at the firm level. The coefficient estimates in this case did not differ from the coefficients we present in this section, confirming the robustness of our specification. We preferred to drop fixed effects at the firm level because it allows us to separate firm-size effects from sector and local labour market effects.

terms, each year of school at upper-secondary level yielded a 3 to 5.5 percent higher wage in 1990 and between 2 and 3 percent higher in 1997. Any year of university study generated a 2 to 6 percent higher wage depending on the year and on the gender. Our estimates are in general lower than other estimates for Italy. This can be explained by the average age of the sample and because we are able to control for the real working experience of our workers.

General and specific experience affects the wage profiles of the two groups differently. While women get a higher premium than men the longer they stay in a firm, men receive a greater return compared to females if they accumulate general experience. However, the higher return to female specific human capital fell over the period.

White-collar workers receive a premium with respect to blue-collars which is different in the two years and for the two genders. The white-collar occupational premium for women was lower than the male in 1990 but higher in 1997. Particularly interesting are the estimates of the dummies for workers with apprenticeship contracts and CFL; wages for these workers are sharply lower than for the others, but women are less disadvantaged by this feature of the wage function and experience a lower gap.

Firm-size dummies show gains for both sexes when working in medium and large firms. Men lost a maximum of 6.5 percent in 1990 and 7.6 percent in 1997 when working in firms with less than 20 employers and gained up to about 6 percent in both years if employed in large firms. Women employed in large firms gain 8.7 percent in 1990 and 10 percent in 1997, while they lost up to a maximum of about 3.4 percent in 1990 and 5 percent in 1997 when employed in a small firm.

5. The unexplained wage gap: the distributional analysis

The first step in our distributional approach to the analysis of wage discrimination is represented by the estimate of the density function of the counterfactual earnings conditional on the estimated values; i.e. the density function $\hat{f}(\hat{r}|\hat{y})$ of equation (8). Indeed, a contour plot of this density function allows us not only to understand whether there is discrimination but also to evaluate the relationship between the extent of discrimination and the level of the estimated wage. In particular, the 45-degree diagonal in the contour plot highlights the absence of wage discrimination; if most of the density mass was concentrated along this diagonal, then estimated and reference wages for female workers would coincide. In contrast, a concentration of the probability mass away from the 45-degree diagonal and closer to the reference wage axis would indicate that substantial wage discrimination against female workers exists.

Figure 1 shows the contour plots of the Gaussian kernel density functions of the counterfactual earnings conditional on the estimated values. We represent the estimated level of wages on the horizontal axes and the counterfactual values on the vertical one; if the density function is concentrated above the diagonal, that indicates discrimination against women; otherwise we have discrimination against men.

Three main features seem to emerge. First, the projection of the density function presents a clear upwards shift at a level of about 50 thousands lire in both years. Through a more detailed analysis of the data, it was easy to identify the workers with a level of earnings lower or equal to that threshold; they are the part-time workers. Part-time workers are characterised by both positive and negative discrimination while full-time workers mainly experience negative discrimination. The second observation concerns the level of the unexplained wage gap; looking at individuals with a wage higher than 50 (focusing therefore on full-time workers), we find clear evidence that being discriminated against for women became almost a certain state of the world over the period; in 1997 the conditional density function for full time workers was concentrated almost entirely above the 45-degree diagonal. However, there does not appear to be individual discrimination; the level of the unexplained wage gap is evenly distributed across the levels of wages. The range of the unexplained wage gap does not change along the distribution of the estimated earnings.

[Figure 1 around here]

In the theoretical section we pointed out the interesting issue of distinguishing between group and individual discrimination. Our methodology makes it possible to investigate this directly by looking at the distribution of the conditional density function along the sample values of the characteristics. However, while some of the characteristics, such as the experience accumulated within the firm or the general experience accumulated during the individual's working history, are measured continuously, others are represented as categorical or dummy variables (like the level of education or the type of contract). When the characteristics are continuous, we first calculate the level of discrimination for each female, as the difference between reference and estimated earnings, and then condition the distribution of that density function on the marginal distribution of the considered characteristic (see Paragraph 2). Then we represent the conditional density function with respect to the values of the characteristic. In this case the absence of wage discrimination is represented by a concentration of the probability mass along the parallel to the horizontal axis. In the case of categorical variables, we can not calculate the conditional density function of the wage gap relative to the distribution of the characteristic. Then, we focus on the density function of the counterfactual wage conditional on the estimated wage and we represent in separate graphs these conditional density functions for any level of the characteristic of interest.

In Figure 2 we represent the conditional density functions of discrimination with respect to the years of experience accumulated in the labour market prior to the present occupation. Part a) of the graph is relative to 1990 and part b) is for 1997. A first general feature common to both years is that most of the probability mass is concentrated on discriminatory values against female workers. However, some interesting differences between the two years are evident. In 1990, discrimination even had negative values, in particular with respect to very low or very high levels of experience. In 1997, a very small part of the distribution is concentrated on negative values and in areas indicating low levels of discrimination; moreover, the level of negative discrimination noticeably reduced between the two years. The two distributions present a significant difference; while the mass of discrimination is essentially parallel to the horizontal axis in 1990, it presents an increasing trend in 1997, meaning that in 1997 we observe individual discrimination against workers with higher levels of experience. The higher is the accumulation of human capital, the higher is the dimension of discrimination. Essentially, our results show that throughout the period considered discrimination increased for individuals with more experience.

[Figure 2 around here]

Rather similar results emerge from the analysis of the relation between discrimination and experience accumulated within the firm - what we have called tenure. As was the case with experience, the position and the shape of the probability mass suggest the presence of significant wage discrimination against female workers in both years with an increase in the discrimination over the period. In particular, between 1990 and 1997 a woman's probability of being favoured (or 'negatively discriminated' against) and the level of such discrimination were noticeably reduced. In addition in 1997 female workers with higher levels of tenure are characterised by a higher probability of being discriminated against; the highest contour plots in figure 3b clearly present an upwards trend. This analysis also confirms the results with respect to general experience; throughout the first seven years of the 1990s the discriminatory phenomenon was characterised by an increasing diversification between individuals. The degree of discrimination was increasingly uneven in its distribution between individuals an presented a positive correlation with levels of experience and tenure.

[Figure 3 around here]

Figures 4 and 5 show the relation between education and discrimination for 1990 and 1997 respectively. As explained in Section 2, the variable education is a categorical variable so we are not able to represent discrimination with respect to the different values of education. Instead we are forced to plot the density function of the reference earnings conditional on estimated earnings, for any level of education. We split the sample between workers with a level lower than upper-secondary education and those with either an upper-secondary education or a university degree. In contrast to the results we previously observed for other human capital characteristics, we find that higher educational levels yield lower discrimination.

[Figures 4 and 5 around here]

Actually, the situation is slightly different between the two years. In 1990 the density mass for high educated individuals is partially located under the diagonal, meaning that workers with higher educational levels can be affected by negative discrimination. On the contrary, workers with low educational levels are almost exclusively affected by positive discrimination. Then, the higher the educational level the higher the probability of being favourably discriminated at any wage level. In 1997 workers with a higher level of education were also discriminated against less, but this advantage noticeably narrowed compared to 1990. In this year the dimension of discrimination appears almost equally distributed independently of the educational level.

6. Conclusions

In this paper we have analysed and compared the extent of discrimination between 1990 and 1997 in two Italian North-Eastern provinces.

We have proposed a new method that evaluates discrimination by focussing directly on the distributions of estimated and reference earnings making use of stochastic kernels to describe their relationship; this method makes it possible to detect both group and individual components of the unexplained wage gap, and to assign a probability of occurrence to any level of discrimination.

The distributional analysis has been focused in particular on the relationship between human capital characteristics and discrimination. With respect to general experience and tenure, the density function representation suggests the presence of significant wage discrimination against female workers in both years with an increase in the phenomenon over the period. In 1997 female workers with higher levels of both experience and tenure were characterised by a higher probability of being discriminated against, compared to 1990. Moreover, a woman's probability of being negatively discriminated and the level of such discrimination were noticeably reduced in 1997.

In contrast to the results on the relationship between experience, tenure and discrimination, we found that higher educational levels are associated with lower discrimination. The situation was slightly different between the two years. In 1990 workers with higher educational levels were affected by both positive and negative discrimination, while workers with low educational levels were almost exclusively affected by positive discrimination. In 1997 workers with a higher level of education were also subject to less discrimination, but this advantage had noticeably narrowed compared to 1990. In 1997 the dimension of discrimination appears almost equally distributed independently of the educational level.

Summing up, the comparison between 1990 and 1997 shows an increasingly unequal incidence of discrimination along the characteristics of women. In particular discrimination is correlated negatively with education and positively with levels of general and specific human capital. These results confirm the value of studying the distribution of discrimination as suggested here rather than representing it in the traditional way that focuses on the average measure of the phenomenon.

References

Addis E. and R. J. Waldmann, 1996, “Differenziali salariali e retributivi e struttura salariale in Italia”, *Economia e Lavoro*, 30, 1-2, pp. 87-103.

Altonji J.G and R.M. Blank, 1999, Race and gender in the labor market, *Handbook of Labor Economics*, 3, pp. 3143-3259.

Blinder A. S., 1973, Wage discrimination: reduced form and structural estimates, *Journal of Human Resources,* 8, pp. 436-53.

Bonjour D. and L. Pacelli, 1999, Wage formation and the gender wage gap: do institutions matter? Italy and Switzerland compared, paper presented at the International Workshop on “Understanding the labour market. Models of duration, labour market histories, job search and matching”, Venice 15-16 January 1999.

Brown R.S., Moon M. and B.S.Zoloth, 1980, Incorporating occupational attainment in studies of male-female earnings differentials, *Journal of Human Resources*, 15, pp. 3-28.

Brunello G. and R. Miniaci, 1999, The economic returns to schooling for Italian men. An evaluation based on instrumental variables, *Labour Economics,* 6, pp. 509-519.

Brunello G., Comi S. and C. Lucifora, The returns to education in Italy: a new look at the evidence, IZA Discussion Paper, 130.

Cain G.G., 1986, The economic analysis of labor market discrimination: a survey, *Handbook of Labor Economics*, 1, pp.693-785.

Card D., 2001, Estimating the return to schooling: progress on some persistent econometric problems, *Econometrica*, 69, 5, 1127-1160.

Chen X., Linton O. and Robinson P.M., 2001, The estimation of conditional densities, The Suntory Centre (LSE) discussion paper, EM/01/415.

Corcoran M.E. and G.Duncan, 1979, Work history, labor force attachment and earnings differences between the races and sexes, *Journal of Human Resources*, 14, 1, pp.3-20.

Durlauf S.N. and Quah D.T., 1999, The new empirics of economic growth. In: Taylor J. and Woodford M. (Eds.), *Handbook of Macroeconomics*, Amsterdam: North-Holland, pp. 235-308.

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.

Flabbi L., 1997, Discriminazione di genere e rendimenti dell’istruzione: un’analisi su microdati individuali, *Rivista di Politica Economica*, 87, 12, pp.173-213.

Flabbi L., 2001, La discriminazione: evidenza empirica e teoria economica, in L. Brucchi, *Manuale di economia del lavoro*, Bologna: Il Mulino.

Gavosto A. and F. Rossi, 1999, Giornate retribuite e differenziali salariali nei dati INPS, *Politica Economica*, XV, 2, pp. 253-257.

Ginzburg A., Scaltriti M., Solinas G. and R. Zoboli, 1998, Un nuovo autunno caldo nel mezzogiorno? Note in margine al dibattito sui differenziali salariali territoriali, *Politica Economica*, XIV, 3, pp. 377-410.

Ginzburg A., Scaltriti M., Solinas G. and R. Zoboli, 1999, Il mistero dei salari in Italia, *Politica Economica*, XV, 2, pp. 259-266.

ITER, 2001, *I differenziali salariali per sesso in Italia*, Napoli: ITER.

Jenkins S., 1994, Earnings discrimination measurement: A distributional approach, *Journal of Econometrics,* 61, 1, pp. 81-102.

Juhn C., Murphy K.M. and B. Pierce, 1991, Accounting for the slowdown in black-white wage convergence. In: Kosters M. (ed), *Workers and their wages*, Washington DC., AEI Press.

Kunze A., 2000, The determination of wages and the gender wage gap: a survey, IZA Discussion paper, 193.

Kunze A., 2002, Gender differences in entry wages and early career wages, IZA Discussion paper, 626.

Loprest P.J., 1992, Gender differences in wage growth and job mobility, *American Economic Review*, 82, 2, pp. 526-532.

Loury G.C., 1981, Intergenerational transfers and the distribution of earnings, *Econometrica*, 49, 4, pp. 843-867.

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 labour markets, *International Economic Review*, 14, pp. 693-709.

Overman H.G. and Y.M. Ioannides, 2001, Cross sectional evolution of the US city size distribution, *Journal of Urban Economics*, 49, 3, pp. 543-566.

Polachek S.W. and J. Robst, 2001, Trend in the male-female wage gap: the 1980s compared with the 1970s, *Southern Economic Journal*, 67, 4, pp. 869-888.

Quah D.T., 1996, Convergence empirics across economies with (some) capital mobility, *Journal of Economic Growth*, 1, 1, pp. 95-124.

Quah D.T., 1997, Empirics for growth and distribution: stratification, polarization and convergence clubs, *Journal of Economic Growth*, 2, 1, pp. 27-59.

Rosenblatt M., 1971, Curve estimates, *Annals of Mathematical Statistics*, 42, pp. 1815-1842.

Silverman B.W., 1986, *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Stokey N.L. and R.E. Jr Lucas. (with Prescott E.C.), 1989, *Recursive methods in economic dynamics*, Cambridge MA: Harvard University Press.

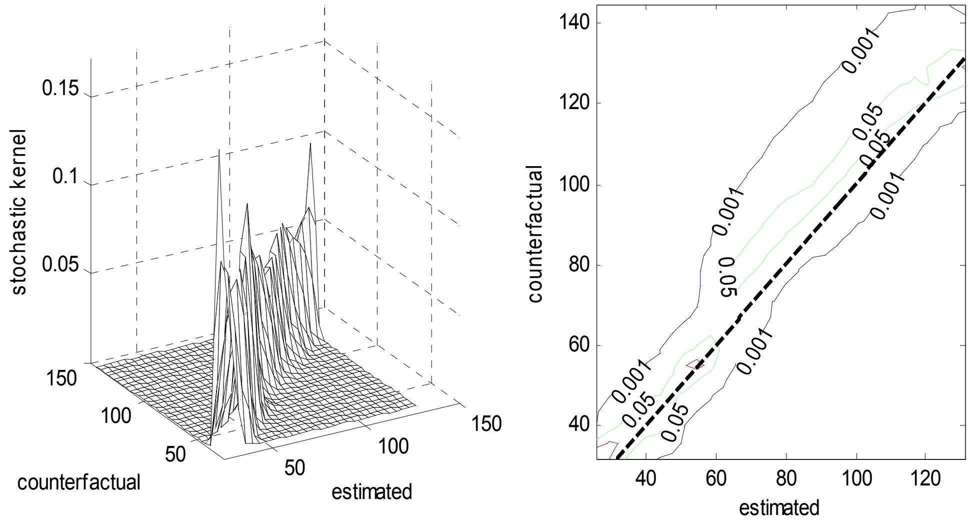
Thurow Lester C., 1975, *Generating Inequality*, New York, Basic Books.

Wood R.G., Corcoran M.E. and P.N. Courant, 1993, Pay differences among the highly paid: the male-female earnings gap in lawyers' salaries, *Journal of Labor Economics,* 11, 3, pp. 417-441.

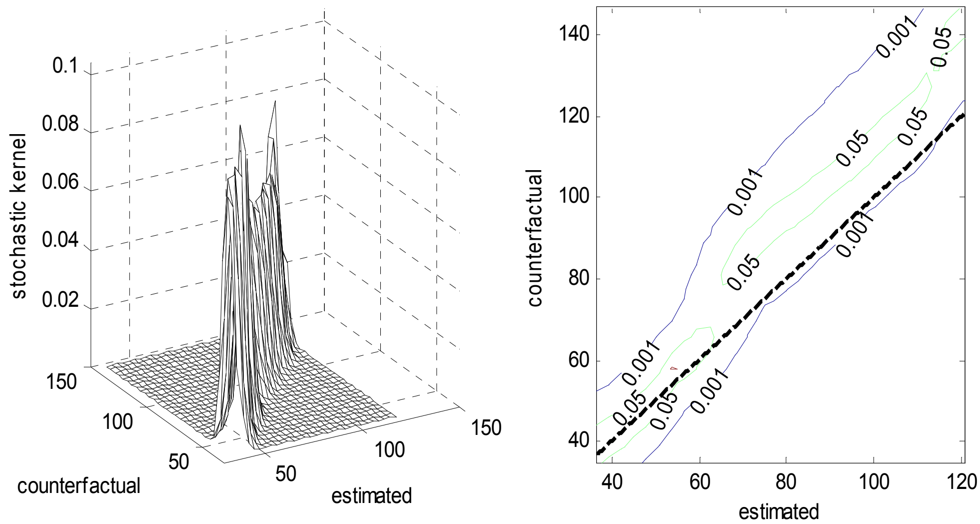
Figures

Figure 1. Probability density functions of counterfactual earnings conditional to estimated earnings

1990



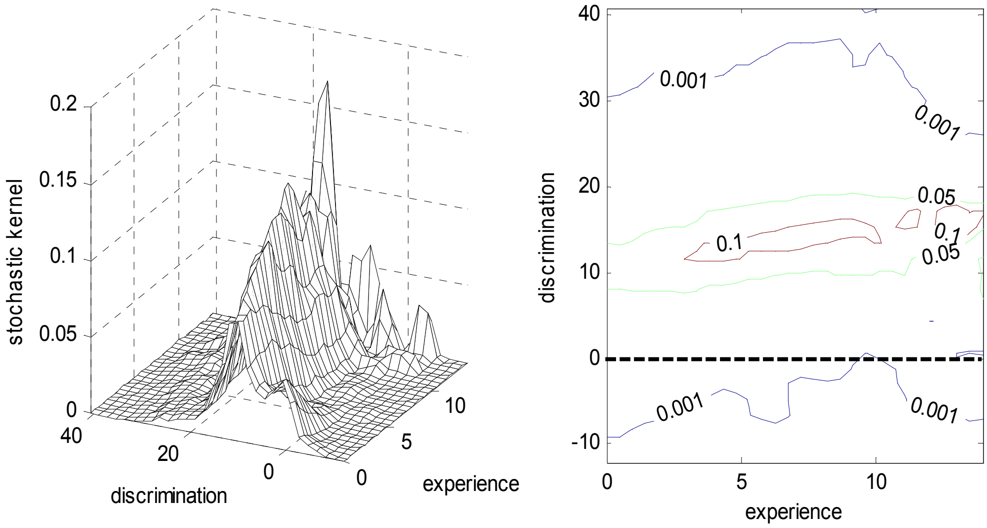
1997



Note: Earnings are expressed in thousands lire (constant prices, base 1995).

Figure 2. Probability density functions of discrimination conditional on working experience

1990



1997

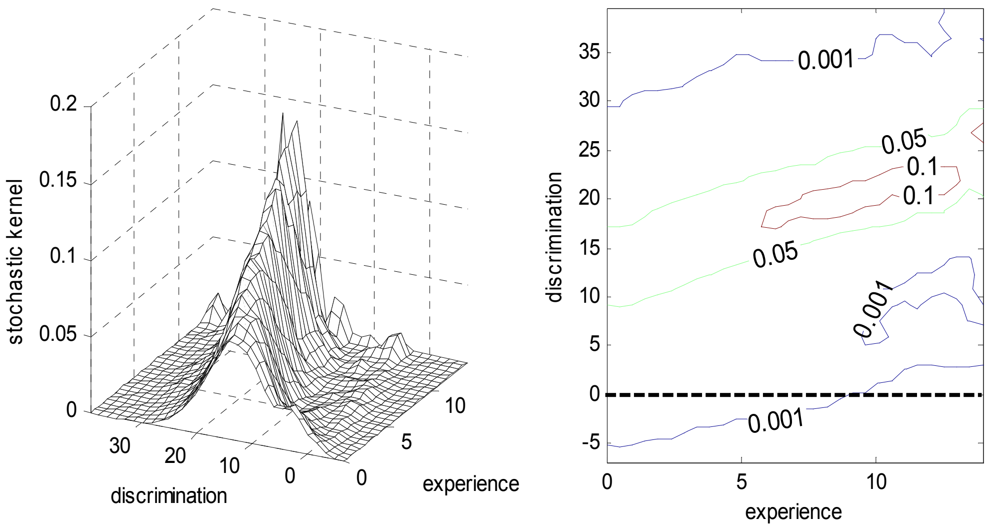
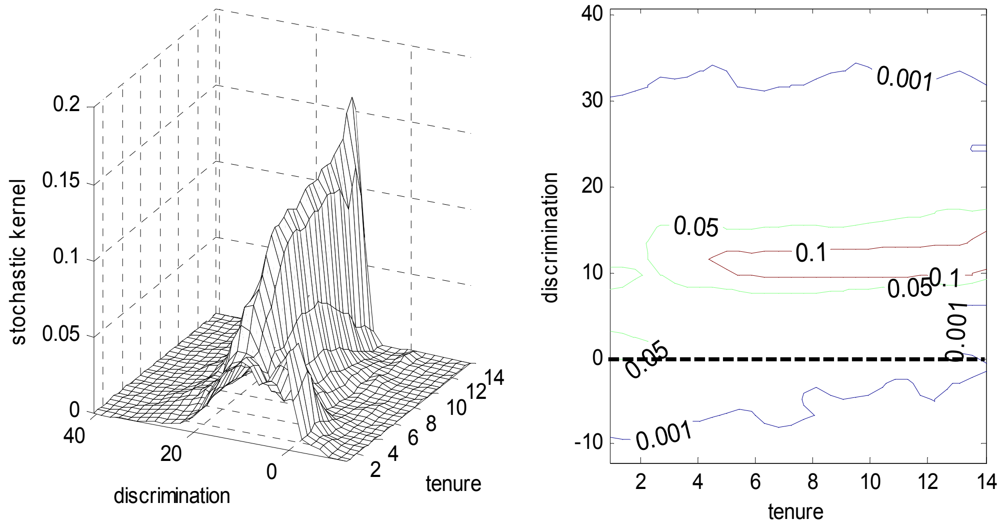


Figure 3. Probability density functions of discrimination conditional on tenure with the firm

1990



1

997

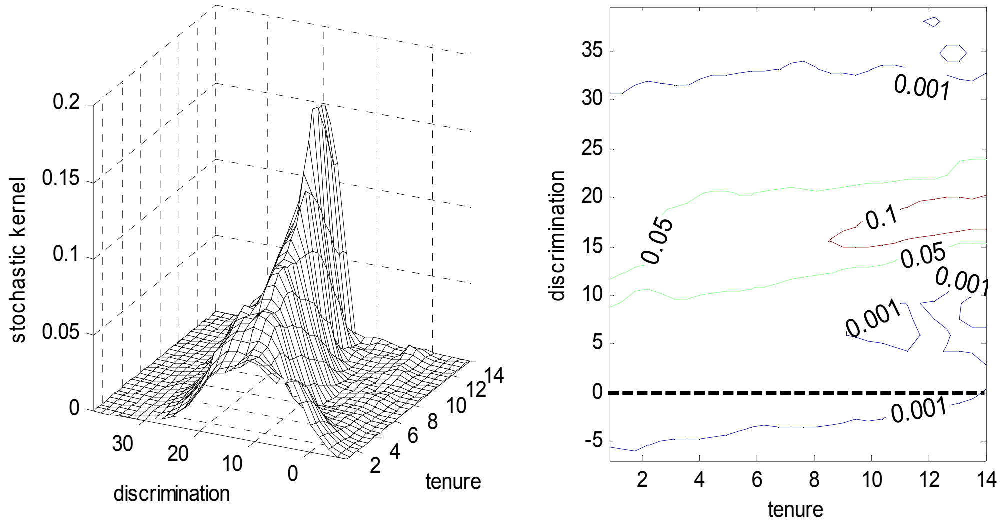
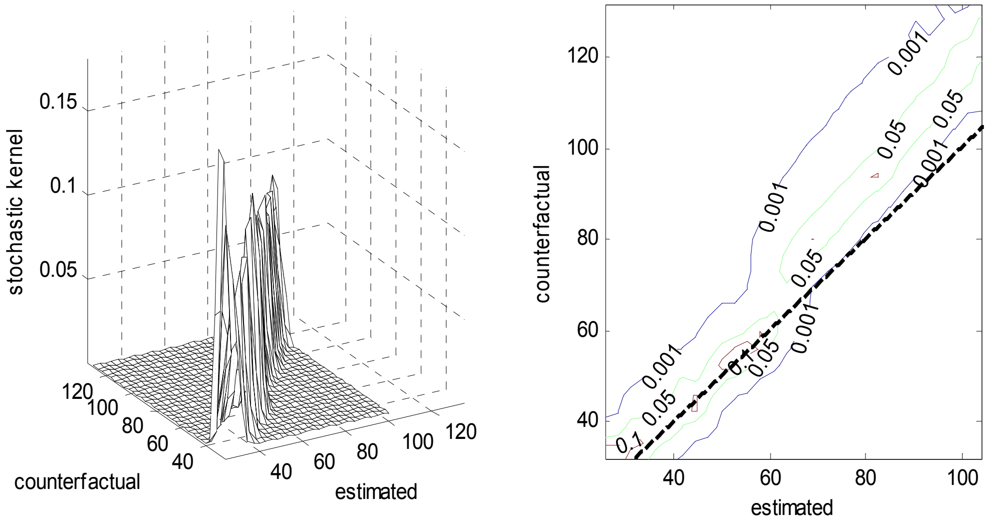


Figure 4. Probability density functions of counterfactual earnings conditional on estimated earnings, by educational level. Year 1990.

Education level lower than upper-secondary



Upper-secondary or university educational level

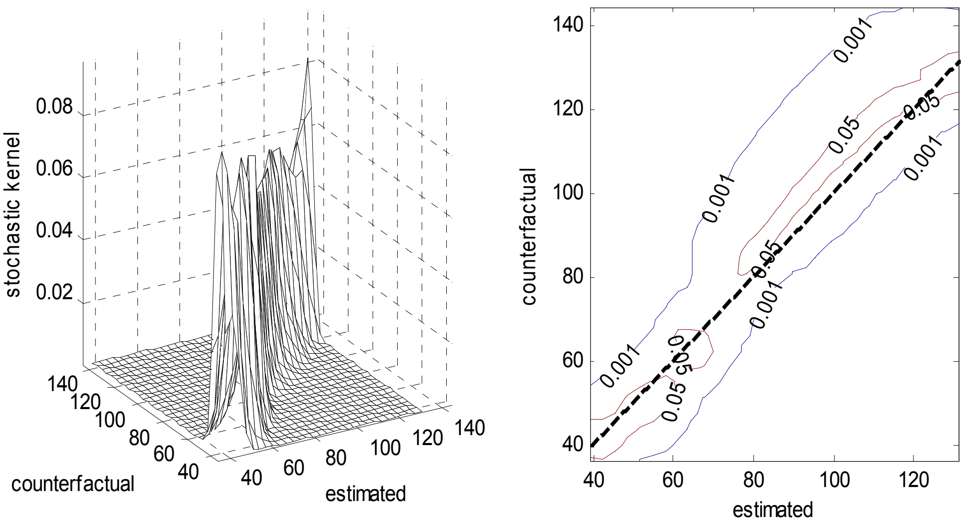
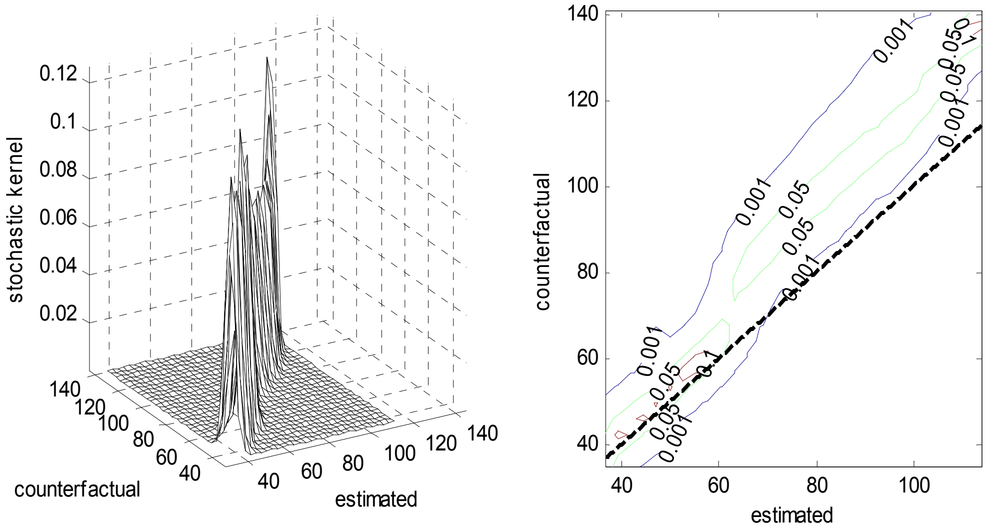
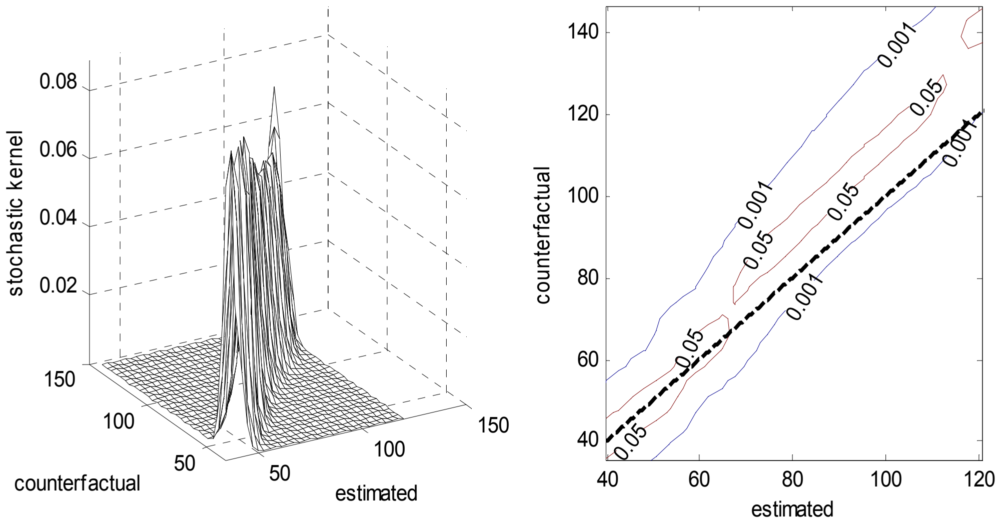


Figure 5. Probability density functions of counterfactual earnings conditional on estimated earnings, by educational level. Year 1997.

Education level lower than upper-secondary



Upper-secondary or university educational level



Apperndix

Table A1
Sample average for each variable

	1990		1997	
	Women	Men	Women	Men
Primary education	.690	.709	.518	.591
Secondary education	.295	.269	.453	.363
University education	.013	.021	.027	.044
Education	1.323	1.311	1.509	1.452
Tenure	.217	.209	.231	.219
Experience	.091	.107	.109	.114
White-collar	.329	.173	.419	.194
Full-time	.979	.998	.968	.997
CFL	.164	.160	.069	.073
Apprenticeship	.135	.115	.078	.059
Sector 3 (Metals, precision tools)	.161	.439	.236	.478
Sector 4 (Food, clothing, wood)	.614	.374	.503	.347
Sector 6 (Commerce)	.150	.158	.166	.143
Sector 8 (Credit)	.074	.026	.094	.030
Firm size 1_5	.271	.236	.237	.191
Firm size 6_10	.170	.138	.129	.123
Firm size 11_20	.229	.164	.205	.161
Firm size 21_50	.127	.164	.159	.184
Firm size 51_100	.071	.101	.101	.132
Firm size > 101	.132	.197	.169	.209

Table A2
Daily earnings (1000 lire) and wage gap by quartiles and occupation*

	1990			1997		
	Women	Men	Wage gap	Women	Men	Wage gap
Mean	77.24	89.43	13.63	82.3	95.73	14.03
S.D:	16.74	23.72		19.31	24.2	
10%	58.85	66.79	11.89	65.03	73.13	11.08
25%	70.51	76.65	8.01	72.19	81.12	11.01
Median	75.66	86.25	12.28	80.21	91.7	12.53
75%	83.51	98.57	15.28	90.11	105.64	14.70
90%	94.19	115.49	18.44	102.7	124	17.18
			White-collar			
Mean	88.98	109.9	19.04	94.12	112.57	16.39
S.D.	19.08	33.28		19.84	30.99	
10%	72.04	78.09	7.75	75.78	81.68	7.22
25%	77.17	86.2	10.48	81.6	89.89	9.22
Median	84.34	101.76	17.12	89.34	104.69	14.66
75%	95.53	126.51	24.49	100.8	128.35	21.46
90%	111.63	151.86	26.49	118.26	153.69	23.05
			Blue-collar			
Mean	71.84	85.16	15.64	74.43	91.73	18.86
S.D.	12.22	18.54		14.29	20.33	
10%	54.52	62.65	12.98	56.98	71.84	20.68
25%	68.22	75.34	9.45	69.02	79.48	13.16
Median	73.11	84.44	13.42	74.74	89.54	16.53
75%	78.46	94.93	17.35	82.42	101.7	18.96
90%	84.86	106.41	20.25	89.93	115.46	22.11

* Only year-round workers. Wage gap in percentage

Table A3

a) Earnings functions. Dependent variable logW*. Year 1990

	Women		Men	
	Coefficient	S.E.	Coefficient	S.E.
Constant	3.1499	0129	3.2574	.1301
Education	.2310	.0916	.1466	.0748
Experience**	.7069	.1861	.7224	.1766
Experience^2	-.7351	.1682	-.6987	.1830
Tenure**	1.1431	.1659	.8992	.1093
Tenure^2	-1.0173	.0919	-.7503	.0466
White-collar	.0928	.0335	.1164	.0365
Full-time	.6126	.0109	.7750	.0398
CFL	-.0217	.0036	-.0751	.0068
Apprenticeship	-.0943	.0378	-.2598	.0397
Sector 3	.0088	.0056	.0529	.0019
Sector 4 (base category)				
Sector 6	.0295	.0053	.0207	.0029
Sector 8	-.0530	.0092	.1057	.0118
Firm-size 1-5	-.0346	.0030	-.0658	.0035
Firm-size 6-10	-.0114	.0028	-.0196	.0034
Firm-size 11-20 (base category)				
Firm-size 21-50	.0288	.0039	.0220	.0035
Firm-size 51-100	.0428	.0061	.0348	.0046
Firm-size 101+	.0875	.0073	.0627	.0073
Observations	33862		Observations	37195
R^2	0.4108		R^2	0.5395

Table A3

b) Earnings functions. Dependent variable logW*. Year 1997

	Women		Men	
	Coefficient	S.E.	Coefficient	S.E.
Constant	3.2745	.1203	3.3320	.0924
Education	.1237	.0641	.0930	.0454
Experience**	.4858	.1411	.6473	.1151
Experience ^{2}	-.4863	.1168	-.4653	.1076
Tenure**	.9356	.1109	.8988	.0855
Tenure^2	-.7922	.0524	-.6247	.0487
White-collar	.1582	.0269	.1324	.0217
Full-time	.6353	.0092	.7809	.0306
CFL	-.0136	.0053	-.0663	.0041
Apprenticeship	-.0951	.0248	-.2236	.0256
Sector 3	.0398	.0028	.0690	.0022
Sector 4 (base category)				
Sector 6	.0501	.0039	.0348	.0033
Sector 8	.0419	.0071	.1369	.0093
Firm-size 1-5	-.0508	.0038	-.0763	.0036
Firm-size 6-10	-.0173	.0035	-.0191	.0039
Firm-size 11-20 (base category)				
Firm-size 21-50	.0450	.0039	.0291	.0034
Firm-size 51-100	.0623	.0050	.0272	.0042
Firm-size > 101	.1053	.0073	.0591	.0058
Observations	30447		Observations	34313
R^2	0.4563		R^2	0.4249

** Estimates corrected by the White var-cov matrix.*

** ‘Tenure’ and’ experience’ are scaled by 1000