Measuring wage discrimination according to an expected utility approach

Rosa Giaimo, Giovanni Luca Lo Magno

Department of Agro-Environmental Systems
University of Palermo

Abstract

Following on from the seminal works by Blinder (1973) and Oaxaca (1973), many methods have been proposed to measure wage discrimination against women. Some of these methods focus on the entire distribution of the discrimination experienced by each woman, underlining a common aspect of poverty and discrimination analysis: the latter two are both based on an idea of deprivation which originates from a poverty line (in the case of poverty) and from the expected wage in the absence of discrimination (in the case of wage discrimination) (Jenkins, 1994; Del Río et al., 2011). These approaches hinge on conditional-to-individual-characteristics expected wages, lacking in any focus regarding the entire conditional wage distribution faced by each woman.

In this paper we will discuss an expected utility approach to the study of wage discrimination. Adjusted and unadjusted for discrimination conditional-to-individual-characteristic wage distributions are evaluated for each woman by means of a utility function. And, in order to evaluate the presence and the discrimination intensity, these distributions will be compared on the basis of the respective certainty equivalent wages. As the choice of the utility function affects the results of the analysis, we will also evaluate the share of women for which the adjusted for discrimination conditional wage distribution second-order stochastically dominates the un-adjusted distribution. Finally, an empirical analysis will be performed for the Italian labour market.

1. Introduction

The classic approach to the measurement of wage discrimination in the labour market is that of the Blinder-Oaxaca decomposition (B-O) (Blinder, 1973; Oaxaca, 1973), where the gender wage differential is decomposed into a part explained by gender differences in endowments (human capital characteristics and other control variables) and a residual part, which is usually interpreted as discrimination. This decomposition employs estimates from log-wage regression models, which are separately estimated for the two genders, decomposing the gender difference in the log geometric mean wage as follows:

$$\log(\overline{W}_M) - \log(\overline{W}_F) = (\overline{\boldsymbol{Z}}_M - \overline{\boldsymbol{Z}}_F)^{'}\widehat{\boldsymbol{\beta}}_M + \boldsymbol{Z}_F^{'}(\widehat{\boldsymbol{\beta}}_M - \widehat{\boldsymbol{\beta}}_F) \tag{1}$$

where $\widehat{\boldsymbol{\beta}}_M$ and $\widehat{\boldsymbol{\beta}}_F$ are estimated coefficients for the male and female log wage regression models respectively, and $\overline{\boldsymbol{Z}}_M$ and $\overline{\boldsymbol{Z}}_F$ are vectors of mean individual characteristics for the male and the female group respectively. The explanatory variables used in the log wage regression models are the same of those of the $\overline{\boldsymbol{Z}}_M$ and $\overline{\boldsymbol{Z}}_F$ vectors. The first part of the decomposition is the so-called explained part and the second refers to the unexplained part, usually attributed to discrimination.

As the B-O decomposition measures discrimination at mean values of individual characteristics, the unexplained part of the B-O decomposition allows for the numerical compensation of individual discrimination between discriminated and non-discriminated women, thus providing the same evaluation for very different distributions of discrimination experienced. To overcome this issue

Jenkins (1994) and Del Río et al. (2011) have suggested a distributional approach. According to this approach, the entire distribution of discrimination experienced by each woman is evaluated on the basis of discrimination indices which satisfy properties borrowed from poverty analysis. While poverty hinges on the concept of income deprivation, discrimination can be conceived as the deprivation of wage from the non-discriminative wage.

This approach raises two issues: 1) how to evaluate the individual discrimination experienced by each woman; and 2) how to aggregate individual discrimination in a single index. In the distributional approach, individual discrimination is evaluated by comparing the unadjusted expected wage with the expected wage in the absence of discrimination, that is, the expected wage a woman would receive if she were paid like a man (the expected wage in the absence of discrimination has the role of a counterfactual wage in the analysis). According to Del Río et al. (2011), individual discrimination can be summarized in an index which is based on the Foster-Greer-Thorbecke (1984) class of poverty indices; this index depends on an aversion-to-discrimination parameter. When this parameter is zero, the index refers to the share of discriminated women, that is, the share of women for which the expected wage in the absence of discrimination is greater than the unadjusted expected wage. When the parameter is not zero, the index measures the discrimination intensity. Other approaches also focus on the entire distribution of individual discrimination experienced but they differ in the method used to derive the counterfactual distribution of wage (Machado and Mata, 2005; Fortin and Lemieux, 1998; DiNardo et al., 1996; Favaro and Magrini, 2008).

The approach we propose in this paper is based on expected utility theory. Using a constant relative risk aversion utility function, we can compare the unadjusted conditional-to-individual-characteristics wage distribution of each woman to the counterfactual distribution, which is obtained by assuming that she is paid as a man. Unlike the distributional approaches by Jenkins (1994) and Del Río et al. (2011), our focus does not regard the conditional mean wage but the entire conditional distribution. A similar approach has been employed by Van Kerm (2010), where the conditional-to-individual-characteristics wage distribution is assumed to follow a Singh-Maddala distribution (Singh and Maddala, 1976; Kleiber and Kotz, 2003; Biewen and Jenkins, 2005). In contrast to Van Kerm (2010), we assume that conditional wage distributions are log-normally distributed. The expected utility approach is also used in poverty analysis to provide risk-adjusted poverty measures (Makdissi and Wodon, 2003; Cruces and Wodon, 2007). The log-wage model employed in our analysis contains a heteroscedastic error term and we will be proposing a solution to take this aspect into account.

Stochastic dominance is another concept employed in our analysis. When distribution F second-order stochastically dominates distribution G, then the expected utility related to distribution F is greater than the expected utility of distribution G for every concave utility function. We enrich our statistical analysis by providing the share of women for which the conditional wage distribution in the absence of discrimination second-order stochastically dominates the unadjusted conditional distribution. This share describes how many women are discriminated against, independently of the chosen concave utility function. An assessment of discrimination testing for stochastic dominance has been used by Millimet and Wang (2006).

This Paper is organized as follows: section 2 discusses the proposed approach and derives the discrimination indices. Section 3 outlines the theoretical underpinning to the relationship between segregation and wage differential. Section 4 describes an empirical analysis, where our approach is applied (also describing our adopted estimation strategy), and an evaluation of the effect of segregation on the gender wage differential. Section 5 makes concluding comments.

2. The certainty equivalent wage and stochastic dominance for measuring wage discrimination

The wage equation model we consider is the following:

$$\log(W_{Si}) = \boldsymbol{Z}'_{Si}\boldsymbol{\beta}_S + \varepsilon_{Si} \qquad \varepsilon_{Si} \sim N(0;\ \sigma^2_{\varepsilon Si}) \qquad (S = M, F) \tag{2}$$

where W_{Si} is the hourly wage of individual i of gender S ($S = M$ for male and $S = F$ for female), $\boldsymbol{Z}_{Si}$ is his/her vector of individual characteristics, $\boldsymbol{\beta}_S$ is the parameter-vector of coefficients and ε_{Si} is the random component.

As the random variable $\log(W_{Si})$, conditional on individual characteristics, is normally distributed, the W_{Si} random variable has a log-normal distribution. We allow for heteroscedasticity in model (2), assuming that the variance of the erratic component ε_{Si} can differ among observations, according to individuals' characteristics. More generally, we can assume that every conditional-to-individual-characteristics quantile $Q_\theta(W_{Si}|\boldsymbol{Z_{Si}})$, of the order θ, can be expressed as a function of the explanatory variables:

$$Q_\theta(W_{Si}|\boldsymbol{Z_{Si}}) = \boldsymbol{Z}'_{Si}\boldsymbol{\beta}_{S\theta} \tag{3}$$

We adopt a common utility function for wage, defined as:

$$u(w) = \begin{cases} \dfrac{w^{1-r}}{1-r}, & if\ r \neq 1 \\ \log(w), & if\ r = 1 \end{cases} \tag{4}$$

where w is the received wage and r is a risk aversion parameter. This is a constant relative risk aversion utility function: when $r < 0$ it represents a risk-loving utility function while when $r > 0$ it refers to a risk-averse type. As we want to furnish an aggregate measure of discrimination using this utility function, and this measure aims at providing a social assessment of discrimination in the labour market, we prefer to use $r > 0$, thereby assuming that most people are risk-averse.

The expected utility function for individual i of gender S is:

$$U_{rSi}(\Lambda_{Si}) = \begin{cases} \displaystyle\int_0^{+\infty} \frac{w^{1-r}}{1-r} d\Lambda_{Si}(w)dw, & if\ r \neq 1 \\ \displaystyle\int_0^{+\infty} \log(w) d\Lambda_{Si}(w)dw, & if\ r = 1 \end{cases} \tag{5}$$

where $\Lambda_{Si}(\cdot)$ is the log-normal conditional cumulative distribution function of the wage of an individual i of gender S. The certainty equivalent wage C_{rSi}, that is, that wage that an individual would view as equally desirable to his/her risky wage, can be obtained by solving the equation $u(C_{rSi}) = U_{rSi}(\Lambda_{Si})$:

$$C_{rSi} = exp\left[\boldsymbol{Z}'_{Si}\boldsymbol{\beta}_S + (1/2)(1-r)\sigma^2_{Si}\right] \tag{6}$$

The discrimination measures we wish to define are based on the definition of the adusted-for-discrimination certainty equivalent wage. It is the certainty equivalent wage of a women facing a wage distribution with male parameters, which is conditional to her characteristics. Thus, this theoretical wage is the certainty equivalent wage a woman would obtain if she were not discriminated:

$$R_{ri} = exp\left[\boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_M + (1/2)(1-r)\sigma^2_{Mi}\right] \tag{7}$$

Thus a woman can be defined as being discriminated against if $R_{ri} > C_{rFi}$, that is, if the certainty equivalent wage is higher in the absence of discrimination than her unadjusted certainty equivalent wage.

A relative measure of discrimination experienced by a woman can be obtained as:

$$d_{ri} = \frac{R_{ri} - C_{rFi}}{R_{ri}} \tag{8}$$

This measure can be used to obtain the following discrimination index:

$$D_r = (1/N_F)\sum_{i=1}^{N_F} d_{ri} \tag{9}$$

where N_F is the number of women in the analysis.

The values of d_{ri}, and accordingly the resulting D_r, depend on the chosen risk aversion parameter r. It would be of interest to know if a woman can be considered as being discriminated against for each value of $r > 0$, so that her discrimination status does not depend on the analyst's subjective choice of r. The requisite concept is that of second order stochastic dominance (Quirk and Saposnik, 1962; Hadar and Russell, 1969; Hanoch and Levy, 1969; Rothschild and Stiglitz, 1970, 1971). Formally, given two probability distributions F and G, distribution F stochastically dominates G in the second order sense (F SSD G) if and only if the expected utility $U(F)$ of F is greater or equal to the expected utility $U(G)$ of G for all utility functions $u(\cdot) \in \mathfrak{U}$, employed in the calculation of the expected utilities (with strict inequality for some $u(\cdot)$), where $\mathfrak{U}$ is the set of utility functions with $du(x)/dx > 0$ and $d^2u(x)/dx^2 < 0$. According to a theorem by Levy (1973), which is valid in the context of log-normal distributions, the log-normal distribution with parameters $(\boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_M; \sigma^2_{Mi})$ second-order stochastically dominates that of parameters $(\boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_F; \sigma^2_{Fi})$ if and only if all the following conditions hold:

$$\begin{cases} \boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_M \geq \boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_F \\ \sigma^2_{Mi} \leq \sigma^2_{Fi} \\ \boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_M - \boldsymbol{Z}'_{Fi}\boldsymbol{\beta}_F \geq (1/2)(\sigma^2_{Fi} - \sigma^2_{Mi}) \end{cases} \tag{10}$$

In order to enrich the information provided by the discrimination measure D_r we define the following index:

$$S = \frac{s(i)}{N_F} \tag{11}$$

where:

$$s(i) = \begin{cases} 1, & \text{if } \Lambda_{Mi} \text{ SSD } \Lambda_{Fi} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Thus, the index S describes the share of women which can be considered as being discriminated against according to all utility functions of type (4) with $r > 0$, for which $du(x)/dx > 0$ and $d^2u(x)/dx^2 < 0$.

3. Relationship between segregation and wage differential

Occupational segregation is generally understood as the segmentation of occupations (or sectors of economic activity) on the basis of workers' gender (Anker, 1997; James and Taeuber, 1985). The classic measure to overall segregation is the segregation index by Duncan and Duncan (1955), which measures the dissimilarity between the distribution of the two genders among occupations. Many other measures have been proposed in the literature in attempting to solve specific methodological issues (Moir and Selby-Smith, 1979; Karmel and MacLachlan, 1988; Hutchens, 2004). Desirable properties of segregation indices have also been described by Hutchens (2001).

The phenomenon of segregation can be due to: employers' practices (Becker, 1957), gender differences in human capital endowments, labour market forces (Blau and Jusenius, 1976), personal constraints (for example, household responsibilities as regards childcare and care for the elderly) and preferences. Other definitions of segregation can be offered which hinge on the source of segregation itself. For example, one could be interested in measuring that part of segregation which cannot be explained by human capital characteristics, that is, the segregation due to occupational discrimination (Åslund and Skans, 2005, 2007; Kalter, 2000). Another common distinction is that between horizontal and vertical segregation, where the latter most properly refers to the different distribution of men and women among occupations with various degrees of skill, responsibility or payment (Hakim, 1981, Watts, 2005). Vertical segregation can be explained by the presence of *glass ceiling*, a subtle but pervasive barrier to the advancement of women in the career ladder.

Clearly, there is a relationship between segregation and wage differentials, as the more women are concentrated in poorly paid occupations, the lower their mean wage is. A theory explaining aspects of segregation and wages is provided by the *overcrowding hypothesis* by Bergmann (1971, 1974). If the labour market consists of sector A and sector B, and employers discriminate women in sector A (in the sense that they prefer to employ men), then women will move to sector B and thus 'overcrowd' it. This supply pressure will deflate wages in sector B, while inflating wages in sector A, thereby reducing the mean wage of women (who are more concentrated in the low-paid sector B). As men and women are equally paid when they are in the same sector, it should be noticed that there is no wage discrimination in either sector A or sector B and discrimination in employment is the source of the wage differential.

4. Estimation strategy and empirical analysis

4.1 Data and variables

The data used in the empirical analysis has been taken from the European Statistics on Income and Living Conditions (Eu-Silc) 2006 data set for Italy[1]. The sample employed in the analysis includes 8,559 male and 6,684 female employees. Explanatory variables used in the estimation of the log hourly wage equations are: *years of education*, *years of work experience*, the *square of years of work experience*, *weekly worked hours*, *occupation* (dummy variables set with 'elementary occupations' as reference category, using the one digit ISCO-88 COM classification of occupations), *economic activity* (dummy variables set with 'agriculture, hunting and forestry' as reference category, using the NACE Rev. 1.1 classification of economic activities) and *region of residence* (dummy variables north and centre with the south as reference category).

4.2 Testing for the log-normality of conditional wage distributions

The first step in the analysis is to assess the log-normality of the conditional-to-individual-characteristics hourly wage; this is equivalent to assessing the normality of the conditional-to-individual-characteristics log-hourly wage. In order to perform this step, we can take advantage of the well-known probability integral transformation theorem, in the way it was also employed by Machado and Mata (2005). According to this theorem, the random variable $F^{-1}(U)$ has distribution F, given a cumulative distribution F and a uniformly distributed random variable U, defined as [0; 1]. For each gender $S = M, F$, we generate a random sample $\{\theta_j\}_{j=1}^{m}$ of size $m = 300$ from a uniformly distributed population with parameters [0; 1]. Thereafter, we estimate $m = 300$ quantile values of order θ_j by using quantile regression analysis (Koenker and Bassett , 1978); the quantile values are conditional to several combinations of regressors. This obtains $\left\{\hat{Q}_{\theta_j}(W_{Si}|\boldsymbol{Z_{Si}})\right\}_{j=1}^{m}$, which represents a simple random sample (of size $m = 300$) from the estimated conditional distribution $\hat{Q}_{\theta_j}(W_{Si}|\boldsymbol{Z_{Si}})$, that is, the estimated conditional quantile of order θ_j. Finally, we use these samples to graphically test distribution normality by means of standardized normal probability plots.

Clearly, it is not possible to test the normality of the log-wage variable for every combination of regressors and we have, therefore, chosen only 24 combinations we consider significant and reported them in Table 2. In defining the combinations reported in Table 2, we have considered

[1] Official code of the data set: IT-SILC XUDB 2006 –April 2008

various levels of explanatory variables. Two levels for education and experience have been considered: the first level L is given by the mean value of the variable minus one standard deviation; the second level H is the mean value of the same variable plus one standard deviation. Mean values and standard deviations are separately calculated by gender. We have defined three job categories: job category A is composed of *plant and machine operators, and assemblers employed in the industry sector*; category B contains *public sector clerks*; and category C comprises *industry managers*. We have considered two regions, the north and the south of Italy, omitting the central regions. The value for weekly worked hours has been set to the mean value of the whole sample, that is, the sample including men and women.

Figure 1 shows the standardized normal probability plots of the log-wage distributions, conditional on the aforementioned 24 regressor combinations, calculated for the male sample; Figure 2 describes the results for the female sample. From observing these plots, we can generally positively conclude the assumption of the normality of the conditional log-wage and, as a consequence, the log-normality of the conditional wage, even if some distributions do not appear to be normally distributed. These results support the application of methods and formulae from Section 2 with sufficient approximation.

4.3 Estimating conditional standard deviation

The second step in the analysis is the estimation of the conditional to individual characteristic standard deviation of the erratic component of the log-wage models. If $\varepsilon_{Si} \sim N(0;\, \sigma^2_{\varepsilon Si})$ then $\log(W_{Si}|\boldsymbol{Z}_{Si}) \sim N(\boldsymbol{Z}'_{Si}\boldsymbol{\beta}_S;\, \sigma^2_{\varepsilon Si})$, while $\sigma^2_{\varepsilon Si}$ depends on the levels of regressors, as is implicit in (3). From standard statistical theory, is well-known that the difference between the mean and the quantile corresponding to the inflection point of a normal distribution is the standard deviation of the distribution, and we can thus exploit this property by defining a simple consistent estimator for σ_{Si}:

$$\hat{\sigma}_{\varepsilon SQi} = (1/2)\left[\hat{Q}_{\Phi(+1)}(W_{Si}|\boldsymbol{Z}_{\boldsymbol{Si}}) - \hat{Q}_{\Phi(-1)}(W_{Si}|\boldsymbol{Z}_{\boldsymbol{Si}})\right] \tag{13}$$

where $\Phi(\cdot)$ is the standard normal cumulative function, $\Phi(+1) \cong 0.841$ and $\Phi(-1) \cong 0.159$. Regression coefficient estimates of the models explaining the log-wage quantile of order $\Phi(+1)$ and log-wage quantile of order $\Phi(-1)$ are reported in Table 1.

Indeed, many consistent estimators for the same parameter could be defined, as for example $\hat{Q}_{\Phi(+1)}(W_{Si}|\boldsymbol{Z}_{\boldsymbol{Si}}) - \hat{Q}_{0.5}(W_{Si}|\boldsymbol{Z}_{\boldsymbol{Si}})$, therefore the choice is quite arbitrary. A non-parametric approach could be the calculation of the standard deviation based on the simulated log-wage distribution, conditional on to the individual characteristics of each observation in the sample. This estimation strategy would be very computer-intensive because 8,559 distributions for men and 6,684 distributions for women would need to be estimated in our analysis. It is, therefore, preferable to use this approach only for the log-wage distributions which are conditional to the our 24 combinations of regressor values. We display the corresponding estimates $\hat{\sigma}_{\varepsilon SDi}$ of the *simulated* approach in Table 2 to compare them with the alternative $\hat{\sigma}_{\varepsilon SQi}$ estimates. No significant differences have emerged in the comparisons between the two estimators and we interpreted that as partial evidence for the validity of our approach.

From Table 2 we can observe a positive, albeit small, effect of education and experience on standard deviation. Industry managers display the greatest standard deviation but it is considerably reduced in the case of high educational level and experience. In most cases men demonstrate greater standard deviation than is the case with women and this occurs mainly when considering $\hat{\sigma}_{\varepsilon SDi}$ estimates. The standard deviation relating to the south as the region of residence is greater than that for other regions.

4.4 Evaluating wage discrimination

The final step in the analysis is to estimate C_{rFi} and R_{ri} in order to estimate the discrimination measures D_r and S and thus evaluate discrimination in the Italian labour market. Consistent estimators for C_{rFi} and R_{ri} are the following respectively:

$$\hat{C}_{rFi} = exp\left[\boldsymbol{Z}'_{Fi}\hat{\boldsymbol{\beta}}_F + (1/2)(1-r)\hat{\sigma}^2_{\varepsilon FQi}\right] \quad (14)$$

and

$$\hat{R}_{ri} = exp\left[\boldsymbol{Z}'_{Fi}\hat{\boldsymbol{\beta}}_M + (1/2)(1-r)\hat{\sigma}^2_{\varepsilon MQi}\right] \quad (15)$$

where $\hat{\boldsymbol{\beta}}_M$ and $\hat{\boldsymbol{\beta}}_F$ are OLS estimates for the regression coefficents of equation (2) for the male and the female group respectively. Using $\hat{C}_{rFi}$ and $\hat{R}_{ri}$, we can construct consistent estimators $\hat{D}_r$ and $\hat{S}$, for D_r and S respectively, which are simply based on formulas (9) and (11).

Table 3 shows the values of estimated $\hat{D}_r$, when different values of the aversion parameter r are used, according to educational level, years of work experience and region. These variables were employed in the analysis to evaluate the degree of discrimination for different levels of human capital characteristics and residency regions. For example, this analysis could provide insights for evaluating if more or less educated women are prone to be discriminated against or if a region is more unfavorable to women than another.

The value for the $\hat{D}_r$ index was 0.114 when the discrimination evaluation was based on the aversion-value $r = 1$. An interpretation of this result is that the adjusted-for-discrimination certainty equivalent wage is, on average, higher than 11.4%, than the unadjusted certainty equivalent wage. In other words, the certainty equivalent wage relating to the conditional wage distribution which women would have if they were not discriminated against is, on average, 11.4% higher than the value for the conditional wage distribution they currently enjoy. It is worth noting that this evaluation is not based on the mean values of the adjusted and unadjusted for discrimination conditional wage distributions, but on the corresponding whole distributions, and thus this can be considered a *distributional approach*. Moreover, this approach takes into account the risk-dimension of wage, which is neglected by other distributional approaches.

Many values of the risk-aversion parameter could be used. We have limited our analysis by using values corresponding to risk-averse utility functions. The $\hat{D}_r$ index value only ranges marginally, from 0.147 to 0.149, when the risk-aversion parameter ranges from $r = 1$ to $r = 4$. The greatest variation, which was determined by a change in the risk-aversion parameter, occurred for the south of Italy, where the $\hat{D}_r$ index range from 0.114 to 0.137.

The discrimination intensity generally increases with increasing working experience, and this is probably a consequence of the presence of a *glass ceiling* effect on the Italian labour market (see U.S. Glass Ceiling Commission (1995) for a definition of *glass ceiling*). Thus, women’s wages decline as compared with men’s with equal human capital endowments during career progression. Education plays an inverse role on the discrimination experienced by women, as increasing educational level is associated with a reduction in discrimination. By taking into account the empirical relationship which was observed between education and discrimination, we could suppose that this phenomenon might assist women to reduce future discrimination because increasing numbers of female graduates are entering the labour market.

While the evaluation of discrimination ranged from 0.155 to 0.145 for the north and from 0.156 to 0.165 for the center, the value for the $\hat{D}_r$ index went from 0.114 to 0.137 for the south. Thus, discrimination in the south appears to be less than that for other regions. Finally, it is interesting to note that the south of Italy provides researchers with an unusual case where the gender wage differential is in favor of women (the unadjusted average difference for the hourly wage is -0.227 euros).

Table 4 shows the share of women, who are second order stochastically dominated, by educational level, experience and region. This numerical analysis is not intended to provide a measure of the intensity of discrimination, but statistical information which could be complementary to results for the $\widehat{D}_r$ index and provide easily interpretable statistical information. The overall value for Italy of 0.440 means that 44% of Italian women are discriminated against in the sense that they would prefer to be paid as they were men, and this is true for every aversion parameter $r > 0$ of the utility function. The south of Italy has the greatest share of second-order stochastically dominated women, while the $\widehat{D}_r$ index reveals that the south is characterized by the least degree of discrimination. These results suggest how multifaceted the analysis of discrimination can be and the pervasive effects of discrimination, despite its possibly low intensity.

The discrimination analysis was conducted by means of calculating the shares of women, who are second-order stochastically dominated, by educational level and experience. The results revealed a similar pattern to those found for the $\widehat{D}_r$ index. Indeed, education once again reduces wage discrimination while experience generally tends to favor it. The latter phenomenon could be interpreted as a clue to the presence of a glass ceiling effect against women in the Italian labour market.

4.5 Effects of segregation on wage differential

In this sub-section we will attempt to evaluate how segregation can have an impact on the gender wage differential in the Italian labour market.

The Treiman and Hartman (1981) decomposition (T-H) provides a rough measure of the impact of segregation on the gender wage differential. This decomposition has two variants which decompose the same absolute wage differential into an inter-occupational component, that is, the part explained by the different distribution of men and women among occupations or sectors, and the intra-occupational component, that is, the part explained by gender wage differentials in occupations or sectors. We can label the first variant as *decomposition* A, which is defined as:

$$\overline{W}_M - \overline{W}_F = \overline{\boldsymbol{W}}_M^{'}(\boldsymbol{P}_M - \boldsymbol{P}_F) + \left(\overline{\boldsymbol{W}}_M - \overline{\boldsymbol{W}}_F\right)^{'}\boldsymbol{P}_F \tag{16}$$

and the second as *decomposition* B, defined as:

$$\overline{W}_M - \overline{W}_F = \overline{\boldsymbol{W}}_F^{'}(\boldsymbol{P}_M - \boldsymbol{P}_F) + \left(\overline{\boldsymbol{W}}_M - \overline{\boldsymbol{W}}_F\right)^{'}\boldsymbol{P}_M \tag{17}$$

where $\boldsymbol{P}_M$ and $\boldsymbol{P}_F$ are the column vectors of the relative frequencies of men and women in occupations (or sectors) respectively and $\overline{\boldsymbol{W}}_M$ and $\overline{\boldsymbol{W}}_F$ are the column vector of mean wages in occupations (or sectors) respectively. The inter-occupational component is given by $\overline{\boldsymbol{W}}'_M(\boldsymbol{P}_M - \boldsymbol{P}_F)$ in decomposition A and by $\overline{\boldsymbol{W}}'_F(\boldsymbol{P}_M - \boldsymbol{P}_F)$ in decomposition B; the intra-occupational component is given by $(\overline{\boldsymbol{W}}_M - \overline{\boldsymbol{W}}_F)'\boldsymbol{P}_F$ in decomposition A and by $(\overline{\boldsymbol{W}}_M - \overline{\boldsymbol{W}}_F)'\boldsymbol{P}_M$ in decomposition B.

In order to understand the T-H decomposition, is useful to pay attention to two extreme labour market configurations. When $\boldsymbol{P}_M = \boldsymbol{P}_F$, no segregation is present in the labour market and the wage differential is entirely explained by wage differences inside occupations; conversely the wage differential is fully explained by segregation when $\overline{\boldsymbol{W}}_M = \overline{\boldsymbol{W}}_F$.

In terms of computation, we notice that the T-H decomposition is a B-O decomposition where the separately estimated by sex regression wage models do not contain the constant term and all explanatory variables are occupational dummies. Mean wages $\overline{\boldsymbol{W}}_M$ and $\overline{\boldsymbol{W}}_F$ correspond to the estimated beta coefficients of the B-O decomposition and $\boldsymbol{P}_M$ and $\boldsymbol{P}_F$ correspond to the vector of mean individual characteristics of the B-O decomposition.

We can use the T-H decomposition to provide an initial and approximate picture of the effect of occupational segregation on the gender wage differential in the Italian labour market. Table 6 reports the results of the T-H decomposition (versions A and B), using occupations and economic sectors, applying it to Italy and its macro-regions. Both occupational and sector analyses reveal that, contrary to conventional wisdom, segregation has a positive impact on the relative-to-male mean, female wage (estimates of the inter-occupational components are all negative). For example, the absolute difference between male and female mean wage in Italy is 0.641 euros; according to the T-H decomposition, the effect of the differences in wage inside occupations is 1.157 euro while -0.516 is the effect of segregation. This result can be approximately interpreted in the following way: the wage differential would be 0.516 in favour of women if there were no gender wage differences inside occupations. Indeed, it must be stressed that this methodology does not control for human capital endowments and it does not provide clear information regarding discrimination. However, these results provide an initial glance about the impact of segregation on wages.

In order to enhance our understanding about the relationship between segregation and wage differentials, we can analyze the relation between male mean wage in occupations (or sectors of economic activities) and the female representation ratio, calculated as:

$$FRR_i = (F_i/N_i)/(F/N) \tag{18}$$

where F_i is the number of female employees in occupation (sector) i, N_i is the total number of workers (men and women) in occupation (sector) i, F is the total number of female employees in the labour market and N are the total number of employees in the whole labour market. The relationship between the mean wage of male workers (assumed to be the non-discriminatory group) and the female representation ratio in occupations (sectors) are plotted in Figure 5. No clear relationship appears between the two measures and we can, therefore, consider this as indicator of the not negative impact of segregation on wage differentials. Anker (1998) classifies occupations as *male dominated* ($FRR_i \leq 0.5$), *gender integrated* ($0.5 < FRR_i < 1.5$) or *female dominated* ($FRR_i \geq 1.5$), and thus we can conclude that occupations and sectors in Italy are almost all *gender integrated* and *female dominated*. Low levels of segregation in the labour market have also been observed by the European Commission (2009a, 2009b). Integrated labour markets do not only have positive features. Bettio (2002) has discussed the negative relationship between segregation and the female employment rate in the European labour market, in which the advantages of the low level of occupational segregation in Italy is counterbalanced by the disadvantages of its low female employment rate.

In order to provide a more complete picture of the impact of segregation on wage differentials and discrimination, we repeat our wage discrimination analysis in an expected utility approach using a reduced set of explanatory variables in the wage equations. The reason for this estimation choice lay in the opportunity to hold constant only the variables that are not determined by the underlying discrimination process, as highlighted by Cain (1986); we have, therefore, omitted occupational dummies, sector dummies and weekly worked hours. This estimation strategy is the same used in the seminal paper by Oaxaca (1973), where a full-scale model and a personal characteristics model are estimated. For the effect of the inclusion of occupational dummies and the choice of the occupational aggregation level on estimated discrimination, see Kidd and Shannon (1996).

The explanatory variables included in the reduced models are: *years spent in education, length of work experience in years*, the *square of years of work experience* and *regional dummies* (the north and center of Italy, with the south as a reference). Table 7 shows the estimated standard deviation of the erratic component of the reduced quantile model for eight regressor combinations. In order to form these combinations we used two educational levels (low and high) and two levels of work experience (low and high), as defined in Section 4.2; only the north and south levels relating to the region categorical variable were used. Standard deviations are estimated by means of the $\hat{\sigma}_{\varepsilon SQi}$ and $\hat{\sigma}_{\varepsilon SDi}$ estimators. A high degree of normality condition was observed for the male case: the null-hypothesis of normality was not refused in seven cases out of eight, according to the Shapiro-Wilk (1965) normality test at a 5% significance level. However, the results of the application of this test are merely indicative because the sample obtained through the use of the integral transformation theorem cannot be considered to be a proper a simple random sample from the true underlying distribution; indeed, the sample is obtained by using an estimated (and, therefore, not true) inverse cumulative probability distribution. Standardized normal probability plots for simulated male log-wage distributions, shown in Figure 3, confirm the accurate approximation to normal distribution. The same cannot be said for female distributions (see Figure 4), which are negatively skewed, displaying higher deviation from normality. Furthermore, the estimates $\hat{\sigma}_{\varepsilon SQi}$ and $\hat{\sigma}_{\varepsilon SDi}$ appear to be quite different from each other and we can interpret this as a consequence of the non-normality of distributions. In spite of this result, the discrimination indices D_r and S were estimated and results displayed in Table 8 and 9 respectively The discrimination index D_r appeared generally lower than that calculated using the full models and we interpreted this as a positive effect of segregation on the gender differences in wages. However, a complementary picture is provided by the second-order stochastically dominated share of women, according to reduced models (Table 9), which is often 100% in the sub-samples under analysis. It can, therefore, be stated that discrimination against women remains a pervasive phenomenon, but its intensity is lower than the discrimination measured by using the full set of variables.

5. Conclusion

The expected utility approach is based on the estimation of two conditional-to-individual-characteristics wage distributions for each woman in the sample: the unadjusted distribution and the adjusted for discrimination distribution. Discrimination emerges from the comparison between the two conditional distributions, while other methods for evaluating discrimination hinge on the conditional expected wages only. Conditional distributions are compared on the basis of certainty equivalent wages, that is, the wage which make an individual indifferent between that wage and the risky conditional distribution. This comparison is based on an expected utility function which is assumed to be the same for each individual.

Although evaluating discrimination according to the expected utility approach depends on the chosen utility function, this approach provides us with a wealth of information, for two reasons: 1) individuals are not only interested in the expected wage but in the entire distribution, and other moments of the distribution can affect their utility, thus a utility approach is useful for analyzing discrimination; and 2) many utility functions can be employed in the same analysis, making explicit the socially, evaluative choice of the researcher. The second-order stochastic dominance criterion provides further information regarding discrimination, permitting us to estimate the share of women who are discriminated against regardless of the chosen concave utility function. We suggest using this as auxiliary information in the analysis but we would like to underline that it does not measure discrimination intensity.

Our method works well when conditional to individual characteristic wage distribution are log-normally distributed. We have tested this distribution form and established that it can be often assumed with reasonable approximation. However, we are of the opinion that the choice of the most appropriate theoretical distribution is a matter of judgment. Nevertheless, the expected utility approach is quite flexible and we believe it can be applied by assuming different wage distributions or a non-parametric framework. The empirical analysis for the Italian labour market revealed some interesting findings: 1) female discrimination increases with work experience and we interpreted this as a consequence of the *glass ceiling*; 2) the impact of discrimination is inversely linked with educational level and we consider this as progress for women as their educational level is currently increasing; 3) the Italian labour market conceals considerable regional differences; 4) segregation and discrimination in employment have less impact than pure wage discrimination; and 5) discrimination intensity and the diffusion of discrimination (measured by the share of discriminated against women) are two different viewpoints, both of which should be monitored.

Tables

Table 1 – Estimations of the coefficients for the full regression models

	men			women		
	$\Phi(+1)$	$\Phi(-1)$	OLS	$\Phi(+1)$	$\Phi(-1)$	OLS
education	0.0285***	0.0196***	0.0263***	0.0271***	0.0237***	0.0244***
	(0.0020)	*(0.0019)*	*(0.0014)*	*(0.0018)*	*(0.0021)*	*(0.0015)*
experience	0.0276***	0.0333***	0.0316***	0.0157***	0.0268***	0.0210***
	(0.0018)	*(0.0018)*	*(0.0013)*	*(0.0019)*	*(0.0021)*	*(0.0017)*
experience (squared)	-0.0004***	-0.0006***	-0.0005***	-0.0001**	-0.0004***	-0.0003***
	(0.0000)	*(0.0000)*	*(0.0000)*	*(0.0001)*	*(0.0001)*	*(0.0000)*
weekly hours worked	-0.0097***	-0.0113***	-0.0102***	-0.0125***	-0.0067***	-0.0097***
	(0.0008)	*(0.0009)*	*(0.0007)*	*(0.0006)*	*(0.0008)*	*(0.0006)*
occupation 1	0.7407***	0.3246***	0.5031***	0.5923***	0.3292***	0.5233***
	(0.0400)	*(0.0429)*	*(0.0398)*	*(0.0494)*	*(0.0582)*	*(0.0607)*
occupation 2	0.5283***	0.3449***	0.3888***	0.4528***	0.3897***	0.4279***
	(0.0286)	*(0.0318)*	*(0.0226)*	*(0.0246)*	*(0.0314)*	*(0.0233)*
occupation 3	0.2430***	0.2255***	0.2159***	0.2291***	0.3143***	0.2713***
	(0.0237)	*(0.0239)*	*(0.0160)*	*(0.0197)*	*(0.0235)*	*(0.0174)*
occupation 4	0.1144***	0.1398***	0.1090***	0.1529***	0.3115***	0.2410***
	(0.0245)	*(0.0245)*	*(0.0161)*	*(0.0210)*	*(0.0247)*	*(0.0179)*
occupation 5	0.1826***	0.1680***	0.1529***	0.0624**	0.1943***	0.1384***
	(0.0282)	*(0.0287)*	*(0.0189)*	*(0.0222)*	*(0.0274)*	*(0.0196)*
occupation 6	0.0609	0.0861	0.0488	-0.0841	0.0941	0.0346
	(0.0451)	*(0.0483)*	*(0.0350)*	*(0.0585)*	*(0.0741)*	*(0.0515)*
ocupation 7	0.0470*	0.0898***	0.0517***	0.0319	0.1202***	0.0678**
	(0.0225)	*(0.0223)*	*(0.0143)*	*(0.0272)*	*(0.0324)*	*(0.0236)*
occupation 8	0.1050***	0.1292***	0.0926***	0.0534	0.1954***	0.1387***
	(0.0237)	*(0.0234)*	*(0.0152)*	*(0.0292)*	*(0.0342)*	*(0.0230)*
occupation 9	0.3317***	0.2949***	0.3039***	0.0928	0.2288	0.1868
	(0.0404)	*(0.0415)*	*(0.0260)*	*(0.1404)*	*(0.1698)*	*(0.1647)*
economic activity 1	0.1546***	0.2178***	0.2000***	0.0146	0.2090***	0.1284***
	(0.0337)	*(0.0353)*	*(0.0248)*	*(0.0358)*	*(0.0438)*	*(0.0352)*
economic activity 2	0.1613***	0.1725***	0.1762***	0.0255	0.1777**	0.0914
	(0.0355)	*(0.0376)*	*(0.0258)*	*(0.0550)*	*(0.0669)*	*(0.0544)*
economic activity 3	0.0717*	0.1175**	0.0831**	-0.0388	0.1476**	0.0573
	(0.0365)	*(0.0381)*	*(0.0268)*	*(0.0364)*	*(0.0458)*	*(0.0360)*
economic activity 4	-0.0550	0.0389	-0.0402	-0.0704	0.0472	0.0051
	(0.0499)	*(0.0505)*	*(0.0373)*	*(0.0416)*	*(0.0514)*	*(0.0411)*
economic activity 5	0.2191***	0.2667***	0.2459***	0.1390**	0.2317***	0.2079***
	(0.0369)	*(0.0391)*	*(0.0271)*	*(0.0440)*	*(0.0539)*	*(0.0429)*
economic activity 6	0.3862***	0.4196***	0.4031***	0.2565***	0.2645***	0.2778***
	(0.0450)	*(0.0457)*	*(0.0326)*	*(0.0416)*	*(0.0508)*	*(0.0420)*
economic activity 7	0.0594	0.1018*	0.0806**	-0.0610	0.1271**	0.0450
	(0.0402)	*(0.0415)*	*(0.0297)*	*(0.0377)*	*(0.0460)*	*(0.0371)*
economic activity 8	0.1499***	0.2660***	0.2153***	0.1171**	0.3418***	0.2360***
	(0.0364)	*(0.0382)*	*(0.0263)*	*(0.0372)*	*(0.0446)*	*(0.0356)*
economic activity 9	0.1594***	0.2651**	0.2239***	0.1681***	0.3900***	0.3290***
	(0.0420)	*(0.0452)*	*(0.0306)*	*(0.0347)*	*(0.0431)*	*(0.0347)*
economic activity 10	0.2519***	0.2814***	0.2688***	0.0891*	0.2556***	0.1987***
	(0.0390)	*(0.0422)*	*(0.0298)*	*(0.0359)*	*(0.0439)*	*(0.0357)*
economic activity 11	0.1693***	0.1226**	0.1380***	-0.0584	0.0718	0.0140
	(0.0372)	*(0.0388)*	*(0.0282)*	*(0.0349)*	*(0.0430)*	*(0.0357)*
north	0.1151***	0.1429***	0.1292***	0.0621***	0.1420***	0.1066***
	(0.0124)	*(0.0127)*	*(0.0085)*	*(0.0125)*	*(0.0153)*	*(0.0111)*
center	0.0612***	0.0993***	0.0844***	0.0386**	0.0810***	0.0587***
	(0.0144)	*(0.0145)*	*(0.0097)*	*(0.0141)*	*(0.0171)*	*(0.0125)*
constant	1.9442***	1.4466***	1.6546***	2.1572***	1.0379***	1.5846***
	(0.0520)	*(0.0533)*	*(0.0391)*	*(0.0439)*	*(0.0533)*	*(0.0441)*

Note: The label $\Phi(+1)$ column contains the estimated coefficients of the log-wage quantile model of the order $\Phi(+1) \cong 0.841$. The label $\Phi(-1)$ column contains the estimated coefficients of the log-wage quantile model of the order $\Phi(-1) \cong 0.159$. The OLS label contains the estimated coefficients of the log-wage model (2), which are estimated by means of OLS. Standard errors are in parentheses. Occupations: 1 legislators, senior officials and managers; 2 professionals; 3 technicians and associated professionals; 4 clerks; 5 service workers and shop and market sales workers; 6 skilled agricultural and fishery workers; 7 craft and related trades workers; 8 plant and machine operators and assemblers; 9 armed forces. The reference for the occupation dummy set is *elementary occupations*. Economic activities: 1 industry; 2 construction; 3 wholesale, retail trade and repair; 4 hotels and restaurants; 5 transport, storage and communications; 6 financial intermediation; 7 real estate, renting and business activities; 8 public administration and defence, compulsory social security; 9 education; 10 health and social work; 11 other activities. The reference for the economic activities dummy set is *agriculture, hunting and forestry*.

Table 2 – Estimated standard deviation of the erratic component of the log-wage equation for various combinations of regressor values (full models)

					$\hat{\sigma}_{\varepsilon SQi}$		$\hat{\sigma}_{\varepsilon SDi}$	
	education	**experience**	**job**	**region**	**men**	**women**	**men**	**women**
1	L	L	A	north	0.239	0.231	0.249	0.252
2	L	L	A	south	0.253	0.271	0.267	0.295
3	L	L	B	north	0.212	0.208	0.239	0.207
4	L	L	B	south	0.226	0.248	0.258	0.250
5	L	L	C	north	0.459	0.434	0.454	0.453
6	L	L	C	south	0.473	0.474	0.471	0.492
7	L	H	A	north	0.237	0.200	0.240	0.232
8	L	H	A	south	0.251	0.240	0.257	0.275
9	L	H	B	north	0.210	0.177	0.227	0.188
10	L	H	B	south	0.224	0.217	0.245	0.230
11	L	H	C	north	0.457	0.403	0.445	0.436
12	L	H	C	south	0.471	0.443	0.461	0.474
13	H	L	A	north	0.272	0.244	0.276	0.263
14	H	L	A	south	0.286	0.284	0.293	0.307
15	H	L	B	north	0.245	0.220	0.263	0.219
16	H	L	B	south	0.259	0.260	0.281	0.262
17	H	L	C	north	0.493	0.446	0.481	0.464
18	H	L	C	south	0.506	0.486	0.497	0.503
19	H	H	A	north	0.270	0.213	0.269	0.243
20	H	H	A	south	0.284	0.253	0.284	0.287
21	H	H	B	north	0.243	0.189	0.253	0.200
22	H	H	B	south	0.257	0.229	0.270	0.242
23	H	H	C	north	0.270	0.213	0.269	0.243
24	H	H	C	south	0.284	0.253	0.284	0.287

Note: L (H) stands for *low level variable* (*high level variable*), that is, the sample mean value of variable minus (plus) one sample standard deviation (mean and standard deviation are calculated separately by gender). A *job* is a combination of occupation and economic activity as follows: job A = "Occupation: plant and machine operators and assemblers; Economic activity: industry"; job B = "occupation: clerks; economic activity: public administration and defence, compulsory social security"; Job C = "occupation: legislators, senior officials and managers, economic activity: industry". The value for weekly worked hours has been set to the mean value of the whole sample, that is, the sample including male and female workers. The values $\hat{\sigma}_{\varepsilon SQi}$ and $\hat{\sigma}_{\varepsilon SDi}$ are estimates of the standard deviations of the erratic component of log-wage equations, conditional on different combinations of regressor values. The estimate $\hat{\sigma}_{\varepsilon SQi}$ is obtained as $(1/2)\left[\hat{Q}_{\Phi(+1)}(W_{Si}|\boldsymbol{Z}_{Si}) - \hat{Q}_{\Phi(-1)}(W_{Si}|\boldsymbol{Z}_{Si})\right]$. The estimate $\hat{\sigma}_{\varepsilon SDi}$ is the standard deviation based on the simulated conditional log-wage distribution.

Table 3 – Index of discrimination $\widehat{D}_r$ for different values of aversion parameter r by education, work experience and region (full models)

		Italy			north			center			south		
Education		*exp ≤5*	*exp>5*	*Tot.*	*exp≤5*	*exp>5*	*Tot.*	*exp≤5*	*exp>5*	*Tot.*	*exp ≤5*	*exp>5*	*Tot.*
primary	***r=1***	0.115	0.188	0.181	0.126	0.195	0.191	0.099	0.200	0.186	0.118	0.164	0.160
	r=2	0.130	0.198	0.192	0.136	0.200	0.196	0.109	0.209	0.196	0.146	0.186	0.182
	r=3	0.144	0.208	0.203	0.146	0.205	0.201	0.119	0.219	0.205	0.173	0.207	0.204
	r=4	0.158	0.218	0.213	0.155	0.209	0.205	0.129	0.228	0.214	0.198	0.227	0.225
secondary	***r=1***	0.114	0.157	0.150	0.114	0.165	0.157	0.123	0.167	0.159	0.101	0.122	0.118
	r=2	0.118	0.158	0.151	0.114	0.162	0.154	0.130	0.170	0.163	0.114	0.129	0.126
	r=3	0.122	0.158	0.152	0.112	0.159	0.151	0.136	0.173	0.166	0.127	0.135	0.134
	r=4	0.126	0.158	0.152	0.111	0.156	0.148	0.142	0.176	0.170	0.139	0.141	0.141
tertiary	***r=1***	0.095	0.131	0.123	0.109	0.141	0.135	0.112	0.143	0.136	0.062	0.099	0.090
	r=2	0.095	0.128	0.120	0.102	0.133	0.127	0.115	0.142	0.136	0.066	0.102	0.093
	r=3	0.094	0.124	0.118	0.096	0.125	0.119	0.117	0.141	0.135	0.071	0.104	0.096
	r=4	0.093	0.120	0.114	0.089	0.117	0.111	0.120	0.139	0.135	0.075	0.107	0.099
Total	***r=1***	0.110	0.155	0.147	0.114	0.163	0.155	0.120	0.164	0.156	0.091	0.120	0.114
	r=2	0.113	0.155	0.147	0.112	0.160	0.152	0.126	0.167	0.159	0.102	0.128	0.122
	r=3	0.116	0.155	0.148	0.110	0.156	0.149	0.131	0.169	0.162	0.112	0.134	0.130
	r=4	0.119	0.155	0.149	0.108	0.153	0.145	0.137	0.172	0.165	0.123	0.141	0.137

Note: *exp* stands for *years of work experience*. r is the aversion parameter.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Table 4 – Share of second order, stochastically dominated women (full models)

	Italy			north			center			south		
education	***exp ≤5***	***exp>5***	***Tot.***	***exp≤5***	***exp>5***	***Tot.***	***exp≤5***	***exp>5***	***Tot.***	***exp ≤5***	***exp>5***	***Tot.***
primary	0.719	0.702	0.704	0.583	0.614	0.612	0.818	0.681	0.700	0.778	0.879	0.870
secondary	0.471	0.438	0.444	0.348	0.349	0.349	0.551	0.532	0.535	0.653	0.579	0.594
tertiary	0.270	0.370	0.348	0.193	0.289	0.269	0.378	0.405	0.399	0.280	0.485	0.433
Total .	0.432	0.442	0.440	0.322	0.355	0.350	0.521	0.516	0.517	0.552	0.583	0.577

Note: *exp* stands for *years of work experience*
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Table 5 – Estimation of regression model coefficients (reduced models)

	men			women		
	$\Phi(+1)$	$\Phi(-1)$	OLS	$\Phi(+1)$	$\Phi(-1)$	OLS
education	0.0584***	0.0371***	0.0485***	0.0666***	0.0476***	0.0559***
	(0.0018)	*(0.0013)*	*(0.0012)*	*(0.0023)*	*(0.0018)*	*(0.0014)*
experience	0.0308***	0.0341***	0.0342***	0.0252***	0.0320***	0.0308***
	(0.0019)	*(0.0017)*	*(0.0014)*	*(0.0028)*	*(0.0022)*	*(0.0019)*
experience (squared)	-0.0004***	-0.0005***	-0.0005***	-0.0003***	-0.0005***	-0.0004***
	(0.0000)	*(0.0000)*	*(0.0000)*	*(0.0001)*	*(0.0001)*	*(0.0001)*
north	0.1010***	0.1719***	0.1217***	-0.0121	0.1661***	0.0779***
	(0.0128)	*(0.0115)*	*(0.0091)*	*(0.0188)*	*(0.0162)*	*(0.0128)*
center	0.0568***	0.1186***	0.0739***	-0.0295	0.0769***	0.0270
	(0.0151)	*(0.0135)*	*(0.0107)*	*(0.0214)*	*(0.0184)*	*(0.0144)*
constant	1.5347***	1.0760***	1.2853***	1.4811***	0.8785***	1.1573***
	(0.0289)	*(0.0223)*	*(0.0198)*	*(0.0410)*	*(0.0323)*	*(0.0260)*

Note: The label $\Phi(+1)$ column contains the estimated coefficients of the log-wage quantile model of the order $\Phi(+1) \cong 0.841$. The label $\Phi(-1)$ column contains the estimated coefficients of the log-wage quantile model of order $\Phi(-1) \cong 0.159$. The OLS label column contains the estimated coefficients of the log-wage model (2), which are estimated by means of OLS. Standard errors are in parentheses.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Table 6 – Treiman-Hartman decomposition of gender wage differential by occupation and sector of economic activity

	occupational analysis							
	decomposition A				***decomposition B***			
	Italy	***north***	***center***	***south***	***Italy***	***north***	***center***	***south***
inter-occupational component	-0.516	-0.321	-0.493	-0.876	-0.695	-0.439	-0.568	-1.415
intra-occupational component	1.157	1.340	1.433	0.649	1.336	1.458	1.508	1.188
gender wage differential	0.641	1.018	0.940	-0.227	0.641	1.018	0.940	-0.227

	sector analysis							
	Decomposition A				***Decomposition B***			
	Italy	***north***	***center***	***south***	***Italy***	***north***	***center***	***south***
inter-sector component	-1.164	-1.130	-1.234	-1.433	-0.663	-0.585	-0.672	-0.868
intra-sector component	1.806	2.148	2.174	1.206	1.304	1.604	1.612	0.640
gender wage differential	0.641	1.018	0.940	-0.227	0.641	1.018	0.940	-0.227

Note: Wages are expressed in euros. Occupations used in the analysis are: legislators, senior officials and managers; professionals; technicians and associate professionals; clerks; service workers and shop and market sales workers; skilled agricultural and fishery workers; craft and related trades workers; plant and machine operators and assemblers; elementary occupations; the armed forces. Economic activities used in the analysis are: agriculture, hunting and forestry ; industry; construction; wholesale, retail trade and repair; Hotels and restaurants; transport, storage and communication; financial intermediation; real estate, renting and business activities; public administration and defence, compulsory social security; education; health and social work.
Source: Author's calculation on the Eu-Silc 2006 Italian data set

Table 7 – Estimated standard deviation of the erratic component of the log-wage equation for various combinations of regressor values (reduced models)

				$\hat{\sigma}_{\varepsilon SQi}$		$\hat{\sigma}_{\varepsilon SDi}$	
	education	**experience**	**region**	**men**	**women**	**men**	**women**
1	L	L	north	0.264	0.277	0.273	0.334
2	L	L	south	0.299	0.366	0.316	0.430
3	L	H	north	0.276	0.267	0.282	0.332
4	L	H	south	0.311	0.356	0.324	0.428
5	H	L	north	0.343	0.347	0.344	0.375
6	H	L	south	0.378	0.436	0.386	0.471
7	H	H	north	0.355	0.337	0.354	0.373
8	H	H	south	0.390	0.426	0.396	0.469

Note: L (H) stands for *low level variable* (*high level variable*), that is, a sample mean value of variable minus (plus) one sample standard deviation (mean and standard deviation are calculated separately by gender). The values $\hat{\sigma}_{\varepsilon SQi}$ and $\hat{\sigma}_{\varepsilon SDi}$ are estimates of the standard deviations of the erratic component of the log-wage equations, conditional on various combinations of regressor values. The estimate $\hat{\sigma}_{\varepsilon SQi}$ is obtained as $(1/2)\left[\hat{Q}_{\Phi(+1)}(W_{Si}|\boldsymbol{Z_{Si}}) - \hat{Q}_{\Phi(-1)}(W_{Si}|\boldsymbol{Z_{Si}})\right]$. The estimate $\hat{\sigma}_{\varepsilon SDi}$ is the standard deviation based on the simulated conditional log-wage distribution.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Table 8 – Index of discrimination $\widehat{\boldsymbol{D}}_{\boldsymbol{r}}$ for different values of aversion parameter $\boldsymbol{r}$ by education, work experience and region (reduced models)

		Italy			**north**			**center**			**south**		
education		***exp ≤5***	***exp>5***	***Tot.***	***exp≤5***	***exp>5***	***Tot.***	***exp≤5***	***exp>5***	***Tot.***	***exp ≤5***	***exp>5***	***Tot.***
primary	***r=1***	0.127	0.133	0.133	0.138	0.143	0.142	0.137	0.146	0.144	0.099	0.106	0.106
	r=2	0.135	0.137	0.137	0.140	0.141	0.141	0.146	0.151	0.150	0.115	0.119	0.119
	r=3	0.143	0.141	0.141	0.142	0.140	0.140	0.154	0.157	0.156	0.132	0.131	0.131
	r=4	0.152	0.145	0.145	0.144	0.138	0.138	0.163	0.162	0.162	0.148	0.144	0.144
secondary	***r=1***	0.081	0.100	0.096	0.089	0.107	0.104	0.091	0.109	0.106	0.050	0.064	0.061
	r=2	0.088	0.102	0.099	0.089	0.103	0.101	0.100	0.114	0.112	0.068	0.078	0.076
	r=3	0.094	0.103	0.102	0.089	0.100	0.098	0.109	0.120	0.118	0.087	0.092	0.091
	r=4	0.100	0.105	0.104	0.088	0.096	0.095	0.118	0.125	0.124	0.105	0.105	0.105
tertiary	***r=1***	0.033	0.050	0.046	0.046	0.060	0.057	0.050	0.063	0.060	0.002	0.017	0.013
	r=2	0.041	0.051	0.049	0.043	0.053	0.051	0.058	0.067	0.065	0.022	0.031	0.029
	r=3	0.048	0.052	0.051	0.040	0.046	0.044	0.067	0.071	0.070	0.041	0.045	0.044
	r=4	0.055	0.053	0.054	0.037	0.039	0.038	0.076	0.075	0.075	0.060	0.059	0.059
total	***r=1***	0.072	0.093	0.089	0.082	0.102	0.099	0.083	0.102	0.099	0.038	0.058	0.054
	r=2	0.078	0.095	0.092	0.081	0.098	0.095	0.092	0.107	0.104	0.057	0.072	0.068
	r=3	0.085	0.097	0.095	0.080	0.094	0.092	0.101	0.112	0.110	0.075	0.085	0.083
	r=4	0.091	0.099	0.097	0.079	0.090	0.088	0.110	0.117	0.116	0.094	0.099	0.098

Note: *exp* stands for *years of work experience*. r is the aversion parameter.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set.

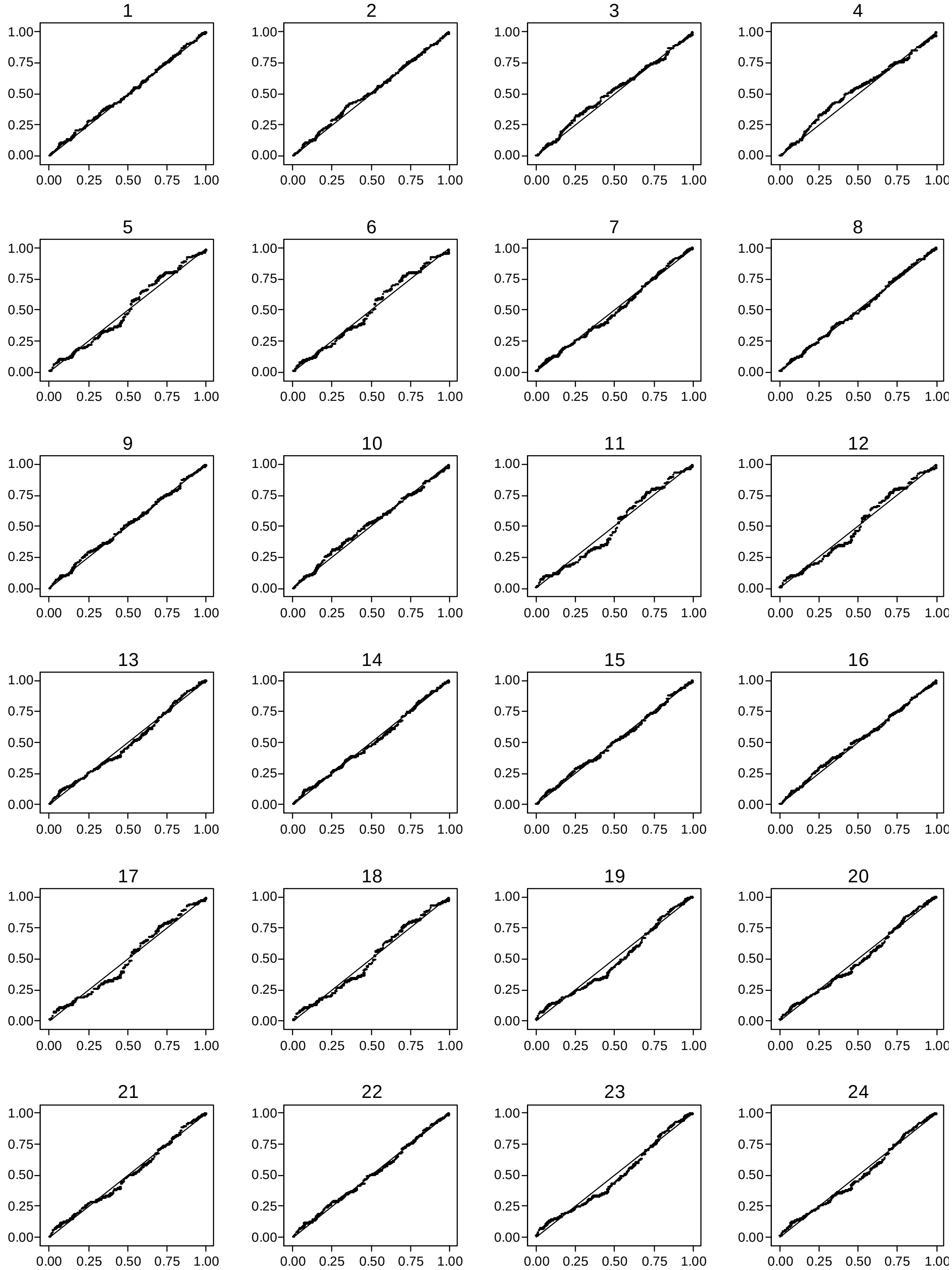
Table 9 – Second order, stochastically dominated share of women (reduced models)

	Italy			**north**			**center**			**south**		
education	***exp ≤5***	***exp>5***	***Tot.***	***exp≤5***	***exp>5***	***Tot.***	***exp≤5***	***exp>5***	***Tot.***	***exp ≤5***	***exp>5***	***Tot.***
primary	1.000	0.512	0.556	1.000	0.042	0.107	1.000	1.000	1.000	1.000	1.000	1.000
secondary	0.668	0.440	0.480	0.344	0.000	0.056	1.000	1.000	1.000	1.000	0.997	0.998
tertiary	0.274	0.439	0.403	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.665	0.495
Tot.	0.588	0.444	0.470	0.290	0.003	0.050	1.000	1.000	1.000	0.717	0.927	0.884

Note: *exp* stands for *years of work experience.*
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

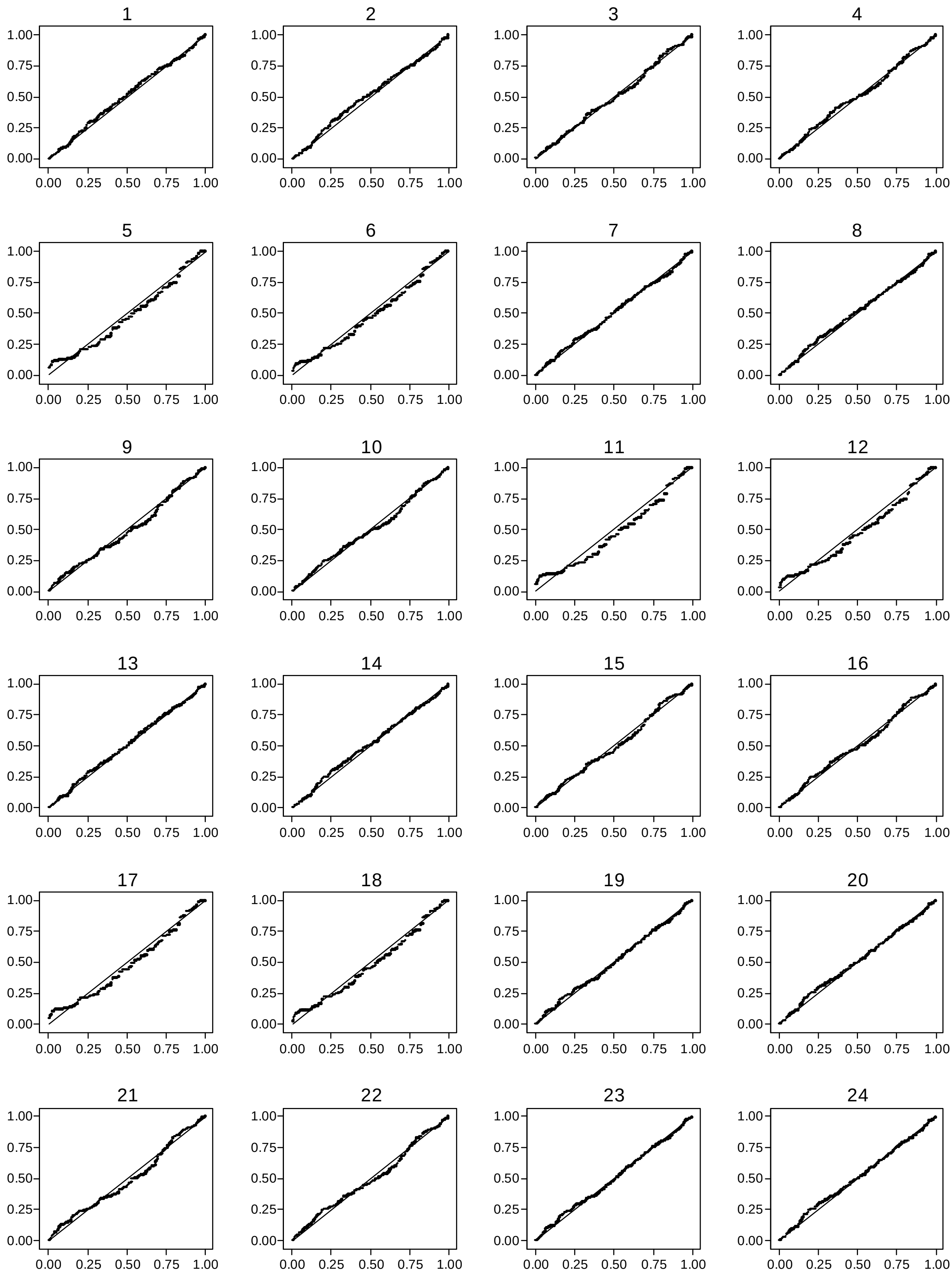
Figures

Figure 1 - Standardized normal probability plots for male log-wage distributions, conditional on various combinations of regressor values (full models)



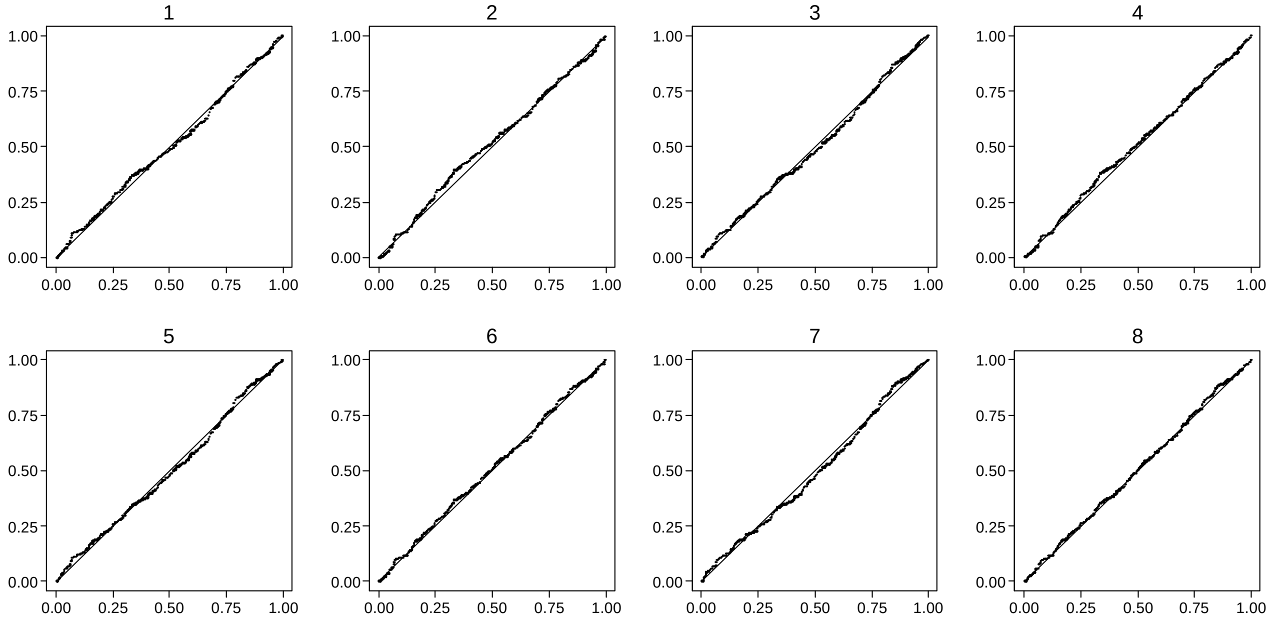
Note: For explanatory notes relating to the chose combinations of regressor values (combinations 1-24), see Table 7.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Figure 2 - Standardized normal probability plots for female log-wage distributions, conditional on various combinations of regressor values (full models)



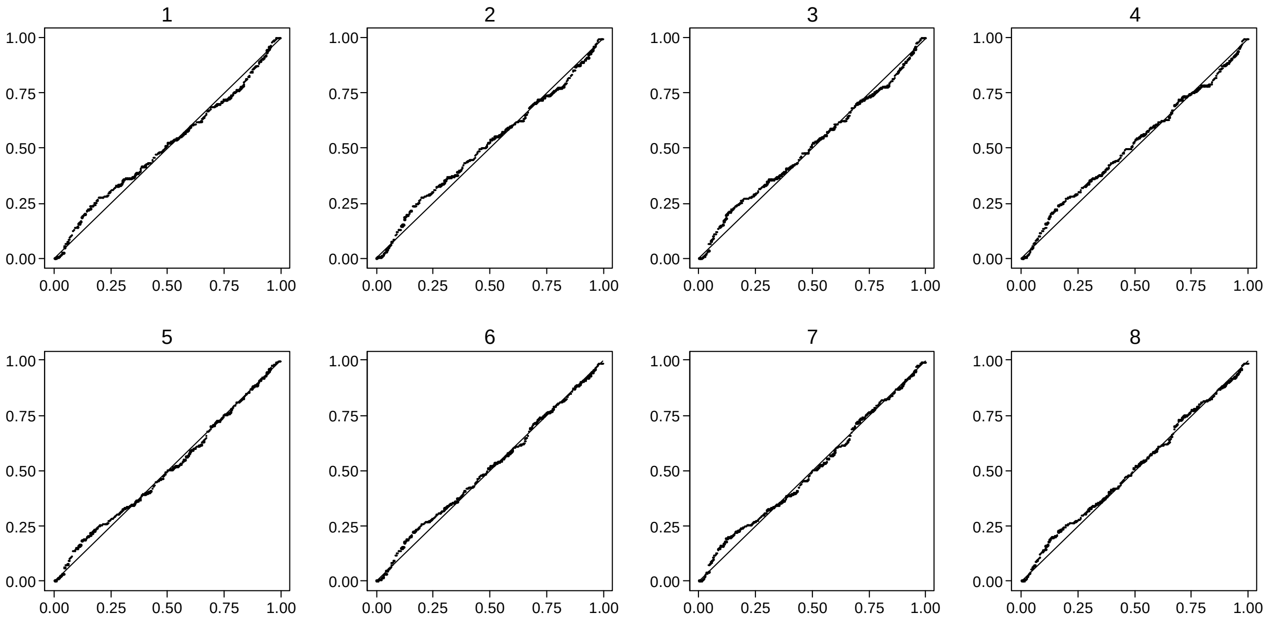
Note: For explanatory notes relating to the chose combinations of regressor values (combinations 1-24), see Table 2.
Source: Author's calculation on the Eu-Silc 2006 italian data set.

Figure 3 - Standardized normal probability plots for simulated male log-wage distributions, conditional on to various regressor combinations (reduced models)



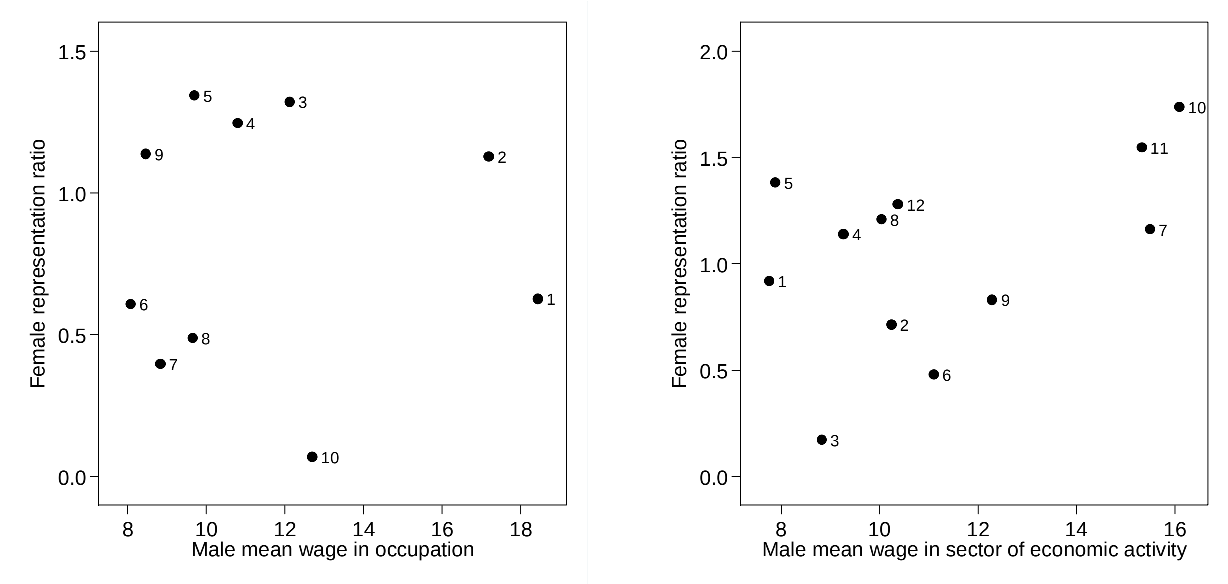
Note: For explanatory notes relating to the chose combinations of regressor values (combinations 1-24), see Table 7.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Figure 4 - Standardized normal probability plots for female log-wage distributions, conditional on various regressor combinations (reduced models)



Note: For explanatory notes relating to the chose combinations of regressor values (combinations 1-24), see Table 7.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

Figure 5 – Female representation ratio and male mean wage by occupation and sector of economic activity



Note: Occupations are: 1 legislators, senior officials and managers; 2 professionals; technicians and associate professionals; 3 clerks; 4 service workers and shop and market sales workers; 5 skilled agricultural and fishery workers; 6 craft and related trades workers; 7 plant and machine operators and assemblers; 8 elementary occupations;9 the armed forces. Economic activities include: 1 agriculture, hunting and forestry; 2 industry; construction; 3 wholesale, retail trade and repair; 4 hotels and restaurants; 5 transport, storage and communication; 6 financial intermediation; 7 real estate, renting and business activities; 8 public administration and defence, compulsory social security; 9 education; 10 health and social work.
Source: Authors' calculation on the Eu-Silc 2006 Italian data set

References

Anker R. (1997), Theories of Occupational Segregation by Sex, *International Labour Review*, Vol. 136, No. 3, pp. 315-339.

Anker R. (1998), *Gender and Jobs: Sex Segregation of Occupations in the World*, International Labour Office, Geneva.

Åslund O., Skans O. N. (2005), Measuring Conditional Segregation – Methods and Empirical Examples, *Working Paper of IFAU – Institute for Labour Market Policy Evaluation*, No. 12.

Åslund O., Skans O. N. (2007), How to Measure Segregation Conditional on the Distribution of Covariates, *Working Paper Series of the Department of Economics of the Uppsala University*, No. 27.

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

Bergmann B. R. (1971), The Effect on White Incomes of Discrimination in Employment, Journal of Political Economy, Vol. 79, No. 2, pp. 294-313.

Bergmann B. R. (1974), Occupational segregation, Wages and Profit when Employers Discriminate by Race or Sex, Eastern Economic Journal, Vol. 1, No. 2, pp. 103-110.

Bettio F. (2002), The Pros and Cons of Occupational Gender Segregation in Europe, *Canadian Public Policy – Analyse de Politiques*, Vol. 28, special issue No. 1.

Biewen M., Jenkins S. P. (2005), Accounting for Differences in Poverty Between the USA, Britain and Germany, *Empirical Economics*, Vol. 30, No. 2, pp. 331-358.

Blau F. D., Jusenius C. L. (1976), Women and the Workplace: The Implications of Occupational Segregation, *Signs*, Vol. 1, No. 3, pp. 181-199.

Blinder A. (1973), Wage Discrimination: Reduced Form and Structural Estimates, *The Journal of Human Resources*, Vol. 8, No. 4, pp. 436-455.

Cain G. G. (1986), The Economic Analysis of Labor Market Discrimination: A Survey. In Ashenfelter O., Laynard R. (eds.), *Handbook of Labor Economics Vol. 1*, Elsevier Science Publisher BV.

Cruces G., Wodon Q. (2007), Risk-Adjusted Poverty in Argentina: Measurement and Determinants, *Journal of Development Studies*, Vol. 43, No. 7, pp. 1189-1214.

Del Río, C., Gradín, C., Cantó, O. (2011), The Measurement of Gender Wage Discrimination: the Distributional Approach Revisited, *Journal of Economic Inequality*, Vol. 9, No. 1, pp. 57-86.

DiNardo J., Fortin N. M., Lemieux T. (1996), Labor Market Institutions and the Distribution of Wages, 1973-1992: a Semiparametric Approach, *Econometrica*, Vol. 64, No. 5, pp. 1001-1044.

Duncan O. D., Duncan B. (1955), A Methodological Analysis of Segregation Indices, *American Sociological Review*, Vol. 20, pp. 210-217.

European Commission (2009a), *Gender Segregation in the Labour Market. Root causes, implications and policy response in the EU*, Publication Office of the European Union, Luxembourg.

European Commission (2009b), *Indicators for Monitoring the Employment Guidelines Including Indicators for Additional Employment Analysis. 2009 Compendium*.

Favaro D., Magrini S. (2008), Group Versus Individual Discrimination Among Young Workers: A Distributional Approach, *Journal of Socio-Economics*, Vol. 37, No. 5, pp. 1856-1879.

Fortin N. M., Lemieux T. (1998), Rank regression, wage distributions, and the gender gap, *Journal of Human Resources*, Vol. 33, No. 3, pp. 610-643.

Foster J., Greer J., Thorbecke E. (1984), A Class of Decomposable Poverty Measures, *Econometrica*, Vol. 52, No. 3, pp. 761-766.

Hadar J., Russell W. R. (1969), Rules for Ordering Uncertain Prospects, *The American Economic Review*, Vol. 59, No. 1, pp. 25-34.

Hakim C. (1981), Job Segregation: Trends in the 1970s', *Employment Gazette* (December), pp. 521-529.

Hanoch G., Levy H. (1969), The Efficiency Analysis of Choices Involving Risk, *The Review of Economic Studies*, Vol. 36, No. 3, pp. 335-346.

Hutchens R. (2001), Numerical Measures of Segregation: Desiderable Properties and their Implications, *Mathematical Social Sciences*, Vol. 42, pp. 13-29.

Hutchens R. (2004), One Measure of Segregation, *International Economic Review*, Vol. 45, No. 2, pp. 555-578.

James D. R., Taeuber K. E. (1985), Measures of Segregation, Sociological Methodology, Vol. 15, pp. 1-32.

Jenkins S. P. (1994), Earnings Discrimination Measurement. A distributional approach, *Journal of Econometrics*, Vol. 61, No. 1, pp. 81-102.

Katler F. (2000), Measuring Segregation and Controlling for Independent Variables, Arbeitspapiere – Mannheimer Zentrum für Eropäische Sozialforschung (MZES), No. 19.

Karmel T., MacLachlan M. (1988), Occupational Sex Segregation. Increasing or Decreasing?, *The Economic Record*, Vol. 64, pp. 187-195.

Kidd M. P., Shannon M. (1996), Does the Level of Occupational Aggregation Affect Estimates of the Gender Wage Gap, *Industrial and Labor Relations Review*, Vol. 49, No. 2, pp. 317-329.

Kleiber C., Kotz S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, Hoboken, New Jersey.

Levy H. (1973), Stochastic Dominance among Log-Normal Prospects, *International Economic Review*, Vol. 15, No. 3, pp. 601-614.

Machado J. A. F., Mata J. (2005), Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression, *Journal of Applied Econometrics*, Vol. 20, No. 4, pp. 445-465.

Makdissi P., Wodon Q. (2003), Risk-Adjusted Measures of Wage Inequality and Safety Nets, *Economics Bulletin*, Vol. 9, No. 1, pp. 1-10.

Millimet D. L., Wang L. (2006), A Distributional Analysis of the Gender Earnings Gap in Urban China, *The B.E. Journal of Economic Analysis & Policy*, Vol. 5, No. 1, Article 5.

Moir H., Selby Smith J. (1979), Industrial Segregation in the Australian Labour Market, Journal of Industrial Relations, Vol. 21, No. 3, pp. 281-291

Oaxaca R. (1973), Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

Quirk J. P., Saposnik R. (1962), Admissibility and Measurable Utility Functions, Review of Economic Studies, Vol. 29, No. 2, pp. 140-146.

Rothschild M., Stiglitz J. E. (1971), Increasing Risk: II. Its Economic Consequences, *Journal of Economic Theory*, Vol. 3, No. 1, pp. 66-84.

Rothschild M., Stiglitz J. E., Increasing Risk (1970): I. A definition, *Journal of Economic Theory*, Vol. 2, No. 3, pp. 225-243.

Shapiro S. S., Wilk M. B. (1965), An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, Vol. 52, No. 3/4, pp. 591-611.

Singh S. K., Maddala G. S. (1976), A Function for Size Distribution of Incomes, *Econometrica*, Vol. 44, No. 5, pp. 963-970.

Treiman D. J., Hartmann H. (1981), *Women, Work, and Wages: Equal Pay for Jobs of Equal Values*, National Academy Press, Washington, D.C.

U.S. Glass Ceiling Commission (1995), *A Solid Investment: Making Full Use of the Nation's Human Capital* (Final Report of the Commission), U.S. Government Printing Office, Washington, DC.

Van Kerm P. (2010), Generalized Measures of Wage Differentials, *Paper Prepared for the 31st General Conference of The International Association for Research in Income and Wealth, St. Gallen, Switzerland, August 22-28, 2010*.

Watts M. J. (2005), On the Conceptualisation and Measurement of Horizontal and Vertical Occupational Gender Segregation, *European Sociological Review*, Vol. 21, No. 5, pp. 481-488.