Dealing With a Potential Bias in Estimating the Share of Discriminated Women

Giamo R.[1], Lo Magno G. L.[23]

Abstract The Blinder-Oaxaca [1, 6] decomposition neglects any distributional issues of discrimination. Instead, Jenkins [5] has argued the importance of a distributional approach in evaluating wage discrimination, focusing on the entire distribution of discrimination experienced by each woman. In their distributional approach, Del Río et al. [3] have adapted the Foster, Greer and Thorbecke (FGT) [4] poverty indices in studying wage discrimination. These discrimination indices depend on a parameter which can be interpreted as a measure of aversion to discrimination. When the aversion parameter is zero, the index measures the share of discriminated women. In this paper we will demonstrate that the *naïve* approach to the estimation of the share of discriminated women – similar to that used by Del Río et al. [3] – could be considerably biased. We propose testing the significance of the discrimination experienced by each woman, using appropriate statistical tests.

1 Introduction

Jenkins [5] has proposed a distributional approach for measuring wage discrimination in which the entire distribution of individual discrimination experienced by each woman is considered. This differs from the Blinder-Oaxaca [1, 6] approach where the analysis is limited to evaluating discrimination for the mean values of individual characteristics. Individual discrimination is the difference between the wage a non-discriminated woman would receive and the unadjusted expected wage for the same woman. Del Río et al. [3] have argued that poverty analysis and wage discrimination analysis are both based on the idea of deprivation. Thus, their proposal is to adapt the class of poverty indices by Foster, Greeer and Thorbecke (FGT) [4] in analyzing discrimination, using two variants (an absolute and a relative index). The first variant provides an absolute measure of discrimination and it is given by

$$D_\alpha = \frac{1}{N_F} \sum_{i \in P} \left(R_i - Q_i \right)^\alpha, \alpha = 0, 1, \ldots \tag{}$$

[1] Rosa Giaimo, Department of Economics, Business and Finance; lomagno.gl@virgilio.it

[2] Giovanni Luca Lo Magno, Department of Economics, Business and Finance; lomagno.gl@virgilio.it

[3]

where R_i is the expected wage in the absence of discrimination for the i -th woman, Q_i is the unadjusted-for-discrimination expected wage for the same woman, N_F is the number of women, α is an aversion parameter analogous to that of the FGT indices, and P is a set identifying discriminated women, that is, women for whom $R_i - Q_i > 0$. When $\alpha = 0,$ the index provides us with the share of discriminated women, i.e. the head-count discrimination ratio.

Inferential aspects of the indices by Del Río et al. [3] have not yet been discussed in the literature. In this paper we only deal with estimation issues when $\alpha = 0$. In the next section we show that, in the case when $\alpha = 0$, estimates could have a serious bias. In order to overcome this issue, we suggest testing the discrimination experienced by each women and report the share of women who have been significantly discriminated. In Section 3 we illustrate these methodological issues by means of an empirical analysis.

2 Estimating the head-count discrimination ratio

The starting point of our wage discrimination analysis is the following wage equation:

$$log W_{Si} = Z'_{Si} \beta_M + \varepsilon_{Si}, \varepsilon_{Si} \ N\left(0 ; \sigma^2_{\varepsilon S}\right), S = M, F, \qquad ()$$

where W_{si} is the hourly wage for sex $S = M$ (male) or $S = F$ (female), Z_{si} is a vector with elements given by values of individual characteristics affecting wage, and ε_{si} is the random normal component of the model.

The expected wage in the absence of discrimination (R_i) is estimated by Del Río et al. [3] using the estimator $\widehat{R}_i = \exp\left(Z'_{Fi} \widehat{\beta}_M + \widehat{\sigma}^2_{\varepsilon F} / 2\right)$, where $\widehat{\beta}_M$ and $\widehat{\sigma}^2_{\varepsilon F}$ are estimators for β_M and $\sigma^2_{\varepsilon F}$ respectively. However, we prefer the $\widetilde{R}_i = \exp\left(Z'_{Fi} \widehat{\beta}_M + \widehat{\sigma}^2_{\varepsilon M} / 2\right)$ estimator, because it aims to estimate the conditional male distribution, which should be used as reference in a discrimination analysis. The empirical results presented by Del Río et al. [3] are based only on $\widehat{R}_i$, but the authors explain in a note that they have also calculated (but not published) estimates also using $\widetilde{R}_i$, thereby obtaining similar results for their discrimination indices. The unadjusted expected wage is estimated as $\widehat{Q}_i = \exp\left(Z'_{Fi} \widehat{\beta}_F + \widehat{\sigma}^2_{\varepsilon F} / 2\right)$.

When the $\widehat{R}_i$ estimator is used, the estimator for D_0 used by Del Río et al. [3] can be written as $\widehat{D}_0 = \left(1/n_F\right) \sum_{i=1}^{n_F} \widehat{d}_{0i}$, where n_F is the size of the female sample and $\widehat{d}_{0i} = 0$ if $\widehat{R}_i - \widehat{Q}_i \leq 0$ or $\widehat{d}_{0i} = 1$ when

$\hat{R}_i - \hat{Q}_i > 0$. We can note that $\hat{d}_i > 0 \iff Z'_{Fi}\left(\hat{\beta}_M - \hat{\beta}_F\right) > 0$. Moreover, $Z'_{Fi}\left(\hat{\beta}_M - \hat{\beta}_F\right) \ N\left(\delta_i; \sigma_i^2\right)$, where $\delta_i = Z'_{Fi}\left(\beta_M - \beta_F\right)$ and $\sigma_i^2 = Z'_{Fi}\left[\sigma_{\varepsilon M}^2\left(Z'_M Z_M\right)^{-1} + \sigma_{\varepsilon F}^2\left(Z'_F Z_F\right)^{-1}\right]Z_{Fi}$. Using these results, it is straightforward to demonstrate that $E\left(\hat{d}_{0i}\right) = \Phi\left(\delta_i/\sigma_i\right)$, where $\Phi\left(\cdot\right)$ is the cumulative distribution function of the standard normal variable. Thus, the expected value of $\hat{D}_0$ is $\left(1/n_F\right)\sum_{i=1}^{n_F}\Phi\left(\delta_i/\sigma_i\right).$ If no discrimination is experienced in the population, then $\delta_i = 0 \ \forall i$ (a sufficient, but not necessary, condition for this is $\beta_M = \beta_F$). Under these conditions the estimator $\hat{D}_0$ exhibits a serious upward bias of 0.5. When $\delta_i \neq 0 \ \forall i$ (a more plausible situation in the real world), the estimator $\hat{D}_0$ is asymptotically unbiased.

To overcome these issues, we suggest accompanying estimates of $\hat{D}_0$ with the share of women who are significantly discriminated against, according to a one-tail hypothesis test, which is based on the following test statistic:

$$\hat{z}_i = \frac{Z'_{Fi}\left(\hat{\beta}_M - \hat{\beta}_F\right)}{\sqrt{Z'_{Fi}\left[\hat{\sigma}_{\varepsilon M}^2\left(Z'_M Z_M\right)^{-1} + \hat{\sigma}_{\varepsilon F}^2\left(Z'_F Z_F\right)^{-1}\right]Z_{Fi}}}, \qquad ()$$

which has a standard normal asymptotic distribution.

When the $\tilde{R}_i$ estimator is used, the estimator for D_0 we propose is

$$\tilde{D}_0 = \left(1/n_F\right)\sum_{i=1}^{n_F}\tilde{d}_{0i} , \text{ where } \tilde{d}_{0i} \text{ equals 0 if } \tilde{R}_i - \hat{Q}_i \leq 0 \text{ or}$$

$\tilde{d}_{0i} = 1$ when $\tilde{R}_i - \hat{Q}_i > 0$. In this case, the derivation of the expected value of the estimator is a tricky task. Nevertheless, we found the estimator to be considerably biased in numerical simulations, especially when discrimination is low. Also in this case, we suggest testing for the significance of the individual discrimination experienced by each woman, using a one-tail test. The statistical test we would like to suggest was originally proposed by Zhou, Gao and Hui [8] and we will use it to compare the mean of two log-normal distributions:

$$\tilde{z}_i = \frac{Z'_{Fi}\hat{\beta}_M - Z'_{Fi}\hat{\beta}_F + \left(1/2\right)\left(\hat{\sigma}_{\varepsilon M}^2 - \hat{\sigma}_{\varepsilon F}^2\right)}{\sqrt{Z'_{Fi}\left[\hat{\sigma}_{\varepsilon M}^2\left(Z'_M Z_M\right)^{-1} + \hat{\sigma}_{\varepsilon F}^2\left(Z'_F Z_F\right)^{-1}\right]Z_{Fi} + \frac{1}{2}\left(\frac{\hat{\sigma}_{\varepsilon M}^4}{m_M} + \frac{\hat{\sigma}_{\varepsilon F}^4}{m_F}\right)}} \qquad ()$$

3 Empirical analysis

In this section we describe, for purely explanatory purposes, an empirical application of the methods we have suggested for dealing with the bias issue of the head-count ratio of discrimination. We have used the EU-SILC 2006 data set (European Statistics on Income and Living Conditions) for Italy. The sample we consider in this paper comprises 16-year old employees, who were in receipt of paid work when interviewed; the sample included 8,559 men and 6,684 women. The discrimination indices of the distributional approach, when $\alpha = 0$ ($\widehat{D}_0$ and $\widetilde{D}_0$), have been separately calculated for the entire sample and for each of the professional occupations in the one-digit Isco-88 (COM) classification, excluding the armed forces. This approach would also be of general interest because various authors have based their discrimination analysis on regression models, which have been separately estimated by occupation (Brown et al. [2]; Solberg [7]). The explanatory variables we used for the models estimated for each occupation are a subset of those used for the whole sample, having been selected through significance tests for beta coefficients.

The estimates $\widehat{D}_0$ and $\widetilde{D}_0$ and the shares of statistically discriminated women for the significance levels 5%, 1% and 0.1% are reported in Table 1. According to $\widehat{D}_0$, 96.9% of women in the whole sample are discriminated against. Of these women, 90.4% are significantly discriminated at a level of 5%, 85.0% at a level of 1% and 77.8% at the more severe level of 0.1%; the test statistic used here is $\hat{z}_i$ from (3). Surprisingly high differences between the estimated shares of discriminated women and the relative share of significantly-discriminated women occur as regards Isco 1 occupation (legislators, senior officials and managers) and Isco 6 occupation (skilled agricultural and fishery workers). These results show that crude point estimates of the share of discriminated women could led to a misleading evaluation of discrimination and highlight the importance of the inferential information which we have added.

Table 1: *Head-count ratio of discriminated women and share of statistically-discriminated women at different levels of significance.*

	(1)				*(2)*			
	$\widehat{D}_0$	5%	1%	0.1%	$\widetilde{D}_0$	5%	1%	0.1%
All occupations	0.969	0.904	0.850	0.778	0.969	0.901	0.848	0.776
Isco 1	**0.571**	**0.114**	**0.043**	**0.000**	**0.571**	**0.114**	**0.043**	**0.000**
Isco 2	0.984	0.728	0.525	0.325	0.993	0.774	0.587	0.376
Isco 3	0.920	0.701	0.586	0.444	0.924	0.715	0.598	0.456
Isco 4	0.885	0.549	0.404	0.244	0.913	0.597	0.449	0.282
Isco 5	0.989	0.867	0.773	0.617	0.990	0.874	0.778	0.642
Isco 6	**0.925**	**0.377**	**0.226**	**0.000**	**0.925**	**0.396**	**0.245**	**0.000**
Isco 7	0.995	0.875	0.785	0.665	0.993	0.858	0.765	0.645

Isco 8	0.976	0.882	0.804	0.743	0.976	0.882	0.804	0.743
Isco 9	0.980	0.707	0.586	0.464	0.962	0.662	0.557	0.403

Note: The statistics in (1) refer to the model where the adjusted-for discrimination expected female

wage is $\exp\left(Z'_{Fi}\beta_M + \sigma^2_{\varepsilon F}/2\right)$, while in (2) it is $\exp\left(Z'_{Fi}\beta_M + \sigma^2_{\varepsilon M}/2\right)$.

The occupation of the armed forces has not been singly considered, but observations from armed forces have been included in the all occupations model.

Isco codes: (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) crafts and related trades workers; (8) plant and machine operators and assemblers; (9) Elementary occupations.

References

1   Blinder, A.S: Wage Discrimination: Reduced Form and Structural Estimates. J. Hum. Resour. 8(4), 436-455 (1973)
2   Brown, R.S., Moon, M., Zoloth, B.S.: Incorporating Occupational Attainment in Studies of Male/Female Earnings Differentials. J. Hum. Resour. 15(1), 3-28 (1980)
3   Del Río, C., Gradín, C., Cantó, O.: The Measurement of Gender Wage Discrimination: the Distributional Approach Revisited. J. Econ. Inequal. 9(1), 57-86 (2011)
4   Foster, J., Greer, J., Thorbecke, E.: A Class of Decomposable Poverty Measures. Econometrica 52(3), 761-766 (1984)
5   Jenkins, S. P.: Earnings Discrimination Measurement. A Distributional Approach. J. Econometrics 61(1), 81-102 (1994)
6   Oaxaca, R.L.: Male-Female Wage Differentials in Urban Labor Markets. Int. Econ. Rev. 14(3), 693-709 (1973)
7   Solberg, E.J.: The Gender Pay Gap by Occupation: A Test of the Crowding Hypothesis. Contemp. Econ. Policy 23(1), 129-148 (2005)
8   Zhou, X., Gao, S., Hui, S.L.: Methods for Comparing the Means of Two Independent Log-normal Samples. Biometrics 53, 1129-1135 (1997)