

Interpreting the Decomposition of the Gender Earnings Gap

Giaimo R., Bono F., Lo Magno G. L.

Abstract In this paper we discuss a decomposition of the gender monthly earnings differential into three components: the first is due to differences in individual characteristics, the second is due to discrimination and the last to the differential in hours worked. The Oaxaca-Blinder hourly wage decomposition is perfectly integrated into the decomposition of the monthly earnings gap. In order to evaluate the effects of this components on the monthly earnings gap, we calculate counterfactual earnings gaps, estimating what would happen if various aspects relating to the two genders were equal (for example, educational level), and then we compare these with the empirical earnings gap. The aim of the counterfactual approach is to furnish measurements that are easy to interpret and understand, even for non-specialists. This decomposition is applied using Eu-Silc 2006 data set for Italy.

Key words: labor statistics, gender pay gap, Oaxaca-Blinder decomposition, counterfactual measures

1 Introduction

Gender gaps in earnings have been on the European Union's (EU) policy agenda for several decades. With the aim of promoting equality, the EU has since the 1980s been implementing specific action programmes based on positive actions and specific projects concerning the work/life balance, women's role in decision-making process, and the participation of women in economic activity and employment. Italy has the

Giaimo R., Department of National Accounts and Social Processes Analysis, University of Palermo, Italy; e-mail: giaimo@unipa.it

Bono F., Department of National Accounts and Social Processes Analysis, University of Palermo, Italy; e-mail: f.bono@unipa.it

Lo Magno G. L., Department of National Accounts and Social Processes Analysis, University of Palermo, Italy; e-mail: lomagno.gl@virgilio.it

lowest gender pay gap (GPG) in Europe, according to data from the *Structure of Earnings Survey*.

Inequality in hourly wages between men and women can be explained by differences in the individual characteristics of employees (human capital and other aspects related to occupation) and discrimination. The methodology proposed by Oaxaca [13] and Blinder [4] quantifies the dimensions of these two aspects by decomposing the mean difference in log hourly wages for the two genders.

Some issues arise from the estimation of male and female wage equations, which are required for the decomposition. In the presence of non-random selected samples, OLS estimates are not consistent (sample selection bias), thus necessitating the application of other methods. The most common methodology for the treatment of the sample selection bias is the Heckman correction [7]; another recent method is that proposed by Lewbel [9]. A bias can also arise when the error term is correlated with one or more explanatory variables (endogeneity bias). This problem can be dealt with by applying the *instrumental variables method*, which allows us to obtain consistent estimates. Beblo et al. [1] have provided a review of the most important challenges regarding estimation, which is related to analysing gender differences in wages.

Another issue arises from the lack of consideration of the complete distribution of discrimination experienced by individuals. This is because the Blinder-Oaxaca decomposition evaluates the effect of discrimination only at the mean of explanatory variables. Jenkins [8] have proposed a class of distributionally-sensitive aggregate discrimination indices in order to summarize the distribution of discrimination. Following Jenkins, Del Río et al. [6] have applied techniques from the literature relating to poverty and deprivation to the construction of discrimination indices.

This paper has two aims: the first is to discuss a decomposition which permits us to decompose the gender differential of monthly earnings into the components already highlighted in the Oaxaca-Blinder decomposition, plus a component representing the part of the differential explained by the difference in hours worked. The second aim is to evaluate the weight on the monthly earnings differential given by the components of the decomposition and by the gender differences in the single variables used in the analysis through measures which are easy to interpret.

In Section 2 we examine the common methods used to decompose the monthly earnings differential and we discuss our approach. In Section 3 we show how counterfactual *gender earnings gaps* (GEGs) can be constructed in order to measure the contribution of the components to the earnings differential, in a way which is easy for non-specialists to understand. In Section 4 we apply the earnings decomposition and the counterfactual approach to an analysis of the monthly earnings differential in the north, centre and south of Italy for the year 2006.

2 Decomposing the Gender Earnings Gap

The standard Blinder-Oaxaca methodology decomposes the mean hourly wage differential between men and women. In order to decompose the mean monthly earnings differential and obtain information relating to discrimination, an appropriate treatment of worked hours is required. Only a few papers have neglected to use information about hours worked, and only when this information is unavailable and differences in worked hours are expected to be low [12].

Blau and Kahn [3] adjust earnings for work hours by assuming a standard working week. They first estimate an earning equation, separately by gender, in which weekly work hours and *dummies* for part-time and full-time work are used. They then use estimated coefficients to adjust each person's earnings for work hours by assuming a 40-hour work week. Finally, earnings corrected for hours are used in the standard Blinder-Oaxaca decomposition.

In another approach, the mean monthly earnings differential is decomposed by adding log hours in the earnings equations. This approach implicitly admits the possibility that the net effect on earnings of an additional hour is different for men and women. The coefficient relating to log hours represents the elasticity of earnings to worked hours and it is usually less than one. The estimation of this earning equation is statistically equivalent to the estimation of a hourly wage equation, which controls for the log of worked hours, in which the coefficient relating to the log of worked hours equals the elasticity of monthly earnings to worked hours minus *one*.

Adjusting for hours worked, as Blau and Kahn do, the results of the final decomposition depend on the standard working week used (Rice [15] uses a 38 hour-week, Oglobin [14] a 41 hour-week). Another approach which estimates the monthly earnings equation controlling for worked hours is equivalent to controlling hourly wages for worked hours, but this latter variable can be determined by the process of discrimination under investigation. According to Cain [5], the variables held constant in the model should not be determined by the process of discrimination being examined. In the decomposition used in this paper, we did not control for worked hours in the hourly wage equation, hypothesizing that worked hours can be a result of discrimination, and this is equivalent to estimating a monthly earnings equation in which elasticity to worked hours has a value of *one*.

Monthly earnings for the i -th individual (Y_i) are here simply defined as the product between hourly wage (W_i) and the hours worked in a month (h_i). After having applied the logarithms to both members of the equality $Y_i = W_i h_i$, and separately calculated the mean for men and women, it is possible to obtain the decomposition of the gender differential in monthly earnings, expressed as:

$$\begin{aligned}
 \ln \bar{Y}_M - \ln \bar{Y}_F &= \underbrace{\ln \bar{W}_M - \ln \bar{W}_F}_{\text{hourly wage differential}} + \underbrace{\ln \bar{h}_M - \ln \bar{h}_F}_{\text{difference in hours worked}} = \\
 &= \underbrace{(\bar{Z}_M - \bar{Z}_F)' \hat{\beta}_M}_{\text{differences in characteristics}} + \underbrace{\bar{Z}_F' (\hat{\beta}_M - \hat{\beta}_F)}_{\text{discrimination}} + \underbrace{\ln \bar{h}_M - \ln \bar{h}_F}_{\text{difference in hours worked}} \quad (1)
 \end{aligned}$$

where $\bar{Y}_M$ and $\bar{Y}_F$ are the geometric averages of the monthly earnings of men and women, $\bar{W}_M$ and $\bar{W}_F$ are the geometric averages of hourly wages, $\bar{h}_M$ and $\bar{h}_F$ are the geometric averages of hours worked, $\bar{Z}_M$ and $\bar{Z}_F$ are the arithmetic averages of individual characteristics, and $\hat{\beta}_M$ e $\hat{\beta}_F$ are the regression coefficients for the model which explains hourly wages, estimated separately for groups of men and women. The log wage equation of the classical human capital earnings model by Mincer [10] [11] is used to estimate the rates of return to individual characteristics.

The monthly earnings differential in this decomposition can be explained by the hourly wage differential and by the difference in hours worked. The hourly wage differential is in turn decomposed, according to the Oaxaca-Blinder decomposition, obtaining a component explained by differences in individual characteristics and a second component attributable to discrimination.

3 Evaluating Endowment, Discrimination and Hours Worked effects

In order to evaluate the weight of the decomposition components on the monthly earnings differential, it is possible to calculate the following ratios:

$$\frac{(\bar{\mathbf{Z}}_M - \bar{\mathbf{Z}}_F)' \hat{\boldsymbol{\beta}}_M}{\ln \bar{Y}_M - \ln \bar{Y}_F}, \quad \frac{\bar{\mathbf{Z}}_F' (\hat{\boldsymbol{\beta}}_M - \hat{\boldsymbol{\beta}}_F)}{\ln \bar{Y}_M - \ln \bar{Y}_F}, \quad \frac{\ln \bar{h}_M - \ln \bar{h}_F}{\ln \bar{Y}_M - \ln \bar{Y}_F} \quad (2)$$

obtained by relating the components, which are highlighted by the decomposition (1), to the raw differential. The sum of these ratios is 1 but each of these is not necessarily included from 0 to 1 and can not, therefore, be interpreted as composition ratios. The ratio values can not be calculated if the differential is zero (a highly improbable result). Moreover, when the value of the differential is very small, the ratio value becomes so large as to render it difficult to interpret. We believe that these ratios are not particularly useful, and that the subsequent numerical and interpretative problems could be interpreted as consequences of the lack of the conceptual validity of the operation of relating an algebraic component of the differential to the differential itself.

In order to measure discrimination, Becker [2] proposes the difference between observed wage ratio and the wage ratio in the absence of discrimination. Oaxaca [13] defines a discrimination coefficient obtained by relating the difference outlined by Becker to the wage ratio in the absence of discrimination. The latter can also be calculated when the mean log wage differential is equal to zero. Furthermore, this coefficient is a-dimensional and permits comparison between different labour markets. Nevertheless, its interpretation is not immediate and does not efficaciously convey the extent of discrimination.

In order to avoid these limitations, we consider preferable to construct measurements of inequality obtained in the absence of discrimination or in the absence of differences in individual characteristics. By comparing these measurements to the inequality observed, we can obtain a measure of the effect of each dimension.

Analogous to the well-known *gender pay gap* (GPG), which is calculated on the basis of hourly wages, we can define the *gender earnings gap* (GEG) as the ratio between the gender differential of mean monthly earnings and the average monthly earnings of male employees. In order to analyse the earnings differential, it would be of great interest to estimate the *discrimination zero* GEG, $G_{Y,0}$, defined as the GEG which would theoretically be determined in the absence of discrimination, thus:

$$G_{Y,0} = (\bar{Y}_{M,0} - \bar{Y}_{F,0}) / \bar{Y}_{M,0} \quad (3)$$

Regarding the hypothesis that the regression coefficient vector estimated for the male group relates to a situation of non-discrimination, $G_{Y,0}$ can be obtained by hypothesising that the female coefficients are equal to the male coefficients ($\hat{\boldsymbol{\beta}}_F := \hat{\boldsymbol{\beta}}_M$). The zero discrimination differential, calculated as the difference between the monthly earnings (in logarithms) of women and men can, therefore, be expressed as $\ln \bar{Y}_{F,0} - \ln \bar{Y}_{M,0} = (\bar{\mathbf{Z}}_F - \bar{\mathbf{Z}}_M)' \hat{\boldsymbol{\beta}}_M + \ln \bar{h}_F - \ln \bar{h}_M$. It is easy to demonstrate the $\ln \bar{Y}_{F,0} - \ln \bar{Y}_{M,0} = \ln (1 - G_{Y,0})$ equality, which can be exploited to obtain $G_{Y,0}$, as provided by the following formula:

$$G_{Y,0} = 1 - (\bar{h}_F/\bar{h}_M) \exp[(\bar{\mathbf{Z}}_F - \bar{\mathbf{Z}}_M)' \hat{\boldsymbol{\beta}}_M] = 1 + (\bar{h}_F/\bar{h}_M)(G_0 - 1) \quad (4)$$

where G_0 is the discrimination zero GPG, that is, the GPG which would be obtained when the female coefficients are equal to those for males. It can be observed that, when the mean labor hours for men and women are equal results $G_{Y,0} = G_0$ and the observed gender earnings gap (G_Y) is equal to the observed gender pay gap (G).

One intuitive measure of the effect of discrimination is given by the difference between empirical and theoretical zero discrimination GEG, that is, $G_Y - G_{Y,0}$. Since theoretical GEG is based on geometric means, it is appropriate that empirical GEG is also calculated on the basis of the geometric means of the observed values.

This difference has a positive value when discrimination favours men and negative when it favours women. The value of this difference can be easily understood as more or fewer GEG percentage points due to the effects of discrimination. We believe that such a measure of discrimination would be the simpler to interpret (particularly for non-specialists), compared with the discrimination coefficient of Oaxaca or according to the second ratio in (2).

It is possible to determine a theoretical GEG, called the *endowment adjusted gender earnings gap* (EAGEG), obtained by hypothesising that the mean levels of individual female characteristics (excluding the hours worked) equal those for men ($\bar{\mathbf{Z}}_F := \bar{\mathbf{Z}}_M$). The EAGEG, $G_{Y,A}$, can be expressed thus:

$$G_{Y,A} = 1 - (\bar{h}_F/\bar{h}_M) \exp[\bar{\mathbf{Z}}_M'(\hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_M)] = 1 + (\bar{h}_F/\bar{h}_M)(G_A - 1) \quad (5)$$

where G_A is the GPG which would be obtained when the female endowments are equal to those for males. In order to evaluate the weight of the differences in characteristics on GEG, we calculate $G_Y - G_{Y,A}$, that is, the difference between empirical GEG (calculated on the basis of geometrical averages) and EAGEG.

In a similar way, hypothesising that women may be paid according to the j characteristic in the same way as for men ($\hat{\beta}_{F,j} := \hat{\beta}_{M,j}$), resulting in a $\hat{\boldsymbol{\beta}}_{F,0(j)}$ vector of remuneration coefficients, we obtain:

$$G_{Y,0(j)} = 1 - (\bar{h}_F/\bar{h}_M) \exp[\bar{\mathbf{Z}}_F' \hat{\boldsymbol{\beta}}_{F,0(j)} - \bar{\mathbf{Z}}_M' \hat{\boldsymbol{\beta}}_M] = 1 + (\bar{h}_F/\bar{h}_M)(G_{0(j)} - 1) \quad (6)$$

or women may have the same variable level of that of men ($\bar{\mathbf{Z}}_{F,j} := \bar{\mathbf{Z}}_{M,j}$), resulting in a $\bar{\mathbf{Z}}_{F,A(j)}$ vector of endowments, obtaining:

$$G_{Y,A(j)} = 1 - (\bar{h}_F/\bar{h}_M) \exp[\bar{\mathbf{Z}}_{F,A(j)}' \hat{\boldsymbol{\beta}}_F - \bar{\mathbf{Z}}_M' \hat{\boldsymbol{\beta}}_M] = 1 + (\bar{h}_F/\bar{h}_M)(G_{A(j)} - 1) \quad (7)$$

The difference between (6) and the empirical GEG is a measure of the discrimination effect of the j -th variable. The difference between (7) and the empirical GEG is a measure of the endowment effect of the same variable.

In order to evaluate the effect of the difference in hours worked on the monthly earnings differential, we compare the theoretical GEG - where women works how much as men do (this GEG equals the empirical GPG) - with the empirical GEG.

4 Results and Conclusions

The decomposition of the monthly earnings differential and the hourly wage differential were obtained from Eu-Silc 2006 data set for Italy. This data set allows us to use numerous controls for individual characteristics. Only employees aged between 16 and 64 were considered. The explanatory variables of hourly wage, used in this analysis, relate to the number of years spent in education, length of work experience, presence of a partner (cohabitation or marriage), supervisor's role, professional job type, sector of economic activity and urban density. These controls are very common in the literature relating to estimates of wage equations. We avoid the use of part-time dummies and log hours because we hypothesize these are not exogenous to the process of discrimination. The Isco-88 Com classification was used for the professional job type variable, which considers 10 categories (*unqualified employees* was chosen as reference category). The agricultural category was chosen as reference category for economic sectors (12 categories). The results are shown in Table 1.

Column (a) shows the results based on decomposition (1). Column (b) shows the ratios between each component and the monthly earnings differential, providing a measure of the components' weight on the differential, but they do not yet generate an intuitive idea of their true importance. Column (c) shows the ratio values between each component of the Oaxaca-Blinder decomposition and the hourly wage differential. When the hourly wage differential is particularly low, as was found to be the case regarding the south of Italy (0.003), ratio values are particularly high and they seem to indicate greater effects for the south than those recorded for other areas. The difference between the discrimination effect for the south of Italy and that of the north (or the center) seems exorbitant. An exorbitant difference also appears for the endowment effect. As previously mentioned in Section 3, these ratios do not lend themselves well to comparisons between areas. Column (d) contains the evaluations, based on counterfactual GEGs, of the effects of the components on the monthly earnings differential. The counterfactual GEG used varies according to which of the following effects is to be evaluated: discrimination (4); differences in characteristics (5); single variable in terms of discrimination (6); single variable in terms of differences in characteristics (7).

A value of -10.7 is obtained for the effect of the differences in endowments for the south of Italy and this can easily be interpreted as a reduction, in terms of percentage points, which GEG would undergo if women had the same average level of individual characteristics as that for men. Values obtained in a similar way to those in column (d) are reported in column (e), although the latter refer to GPG. These have been obtained as the difference between empirical GPG and appropriate theoretical GPG. These measures can be easily interpreted and, in our opinion, are more correct than the data in column (c) for making regional comparisons.

Focusing our attention on analysing discrimination, the log monthly earnings differential pertaining to the south of Italy is equal to 0.219, of which 0.078 (35.9% of the differential) can be attributed to discrimination. In column (c) the ratio between discrimination and hourly differential is 2,968.5%. This high value is due to the very small hourly wage differential (0.003) and it is not easy to interpretate.

Table 1: Decomposition and analysis of monthly and hourly wage differentials. Northern, central and southern Italy, 2006.

	North					Centre					South				
	(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)
Monthly differential (log)	0.284	100	---	---	---	0.274	100	---	---	---	0.219	100	---	---	---
Hourly differential (log)	0.093	32.6	100	---	---	0.095	34.8	100	---	---	0.003	1.2	100	---	---
Endowments	-0.006	-2.1	-6.5	-2.9	-3.5	-0.004	-1.4	-4.1	-3.0	-3.6	-0.076	-34.7	-2868.5	-10.7	-13.2
<i>Education</i>	-0.020	-7.0	-21.5	-1.4	-1.7	-0.017	-6.0	-17.4	-1.2	-1.4	-0.070	-32.2	-2665.0	-3.9	-4.9
<i>Experience</i>	0.019	6.7	20.5	1.3	1.5	0.018	6.6	19.1	1.6	2.0	0.029	13.3	1102.5	2.4	3.0
<i>Partner</i>	-0.001	-0.5	-1.6	-0.1	-0.1	0.003	0.9	2.7	0.0	0.1	0.010	4.8	394.9	0.7	0.8
<i>Supervisor</i>	0.016	5.6	17.3	0.9	1.1	0.016	5.9	16.8	0.5	0.6	0.007	3.2	266.8	0.5	0.6
<i>Occupation</i>	-0.019	-6.6	-20.1	-2.4	-2.9	-0.006	-2.2	-6.4	-0.4	-0.5	-0.017	-7.7	-640.1	-5.8	-7.2
<i>Sector</i>	-0.001	-0.2	-0.6	-1.2	-1.4	-0.017	-6.2	-17.8	-3.4	-4.1	-0.035	-16.2	-1338.5	-4.8	-6.0
<i>Urban density</i>	-0.001	-0.2	-0.5	-0.0	-0.0	-0.001	-0.4	-1.1	-0.1	-0.1	0.000	0.1	10.8	0.0	0.0
Discrimination	0.099	34.7	106.5	7.8	9.5	0.099	36.3	104.1	7.9	9.5	0.078	35.9	2968.5	6.6	8.1
<i>Education</i>	0.022	7.8	24.0	1.7	2.1	0.018	6.6	19.0	1.4	1.7	0.116	53.1	4394.4	9.9	12.3
<i>Experience</i>	0.048	17.0	52.1	3.7	4.5	0.024	8.7	24.9	1.8	2.2	0.032	14.6	1204.0	2.6	3.2
<i>Partner</i>	0.017	5.9	18.1	1.3	1.5	0.050	18.1	52.0	3.9	4.6	0.017	8.0	661.7	1.4	1.8
<i>Supervisor</i>	0.006	2.2	6.7	0.5	0.6	0.015	5.7	16.2	1.2	1.4	0.001	0.5	44.6	0.1	0.1
<i>Occupation</i>	-0.044	-15.3	-46.9	-3.2	-3.9	-0.156	-57.1	-163.9	-11.0	-13.2	-0.048	-21.8	-1804.2	-3.7	-4.6
<i>Sector</i>	0.128	45.0	138.0	10.3	12.5	0.139	50.9	146.1	11.4	13.6	-0.099	-45.4	-3756.1	-7.6	-9.4
<i>Urban density</i>	0.005	1.8	5.6	0.4	0.5	0.005	1.8	5.1	0.4	0.4	0.015	6.7	551.8	1.2	1.5
<i>Constant</i>	-0.085	-29.8	-91.2	---	---	0.004	1.6	4.6	---	---	0.044	20.2	1672.2	---	---
Hourly component	0.192	67.4	---	15.9	---	0.178	65.2	---	14.8	---	0.216	98.8	---	19.4	---

Source: Authors' calculations on Eu-Silc 2006.

Notes: (a) absolute values; (b) values of (a) divided by monthly differential (log); (c) values of (a) divided by hourly differential (log); (d) evaluation of effect on GEG, based on counterfactual GEG; (e) evaluation of effect on GPG, based on counterfactual GPG.

For this reason evaluations made on the basis of counterfactual GEGs (column d) and counterfactual GPGs (column e) are very useful. They indicate that discrimination favours men by 6.6 percentage points more for GEG and 8.1 percentage points more for GPG. On the basis of this evaluation, the south of Italy seems to be an area characterised by the least discrimination, painting a rather different picture to the situation described in column (c).

We have interpreted this low value of discrimination in the south of Italy as a result of the greater presence of southern employees in the public sector where, compared to the private sector, discriminatory behaviour is more tightly controlled.

References

1. Beblo, M., Beninger, D., Heinze, A., Laisney, F.: Methodological Issues Related to the Analysis of Gender Gaps in Employment, Earnings and Career Progression. Project carried out for the European Commission, Employment and Social Affairs DG (2003).
2. Becker, G.S.: The economics of discrimination. University of Chicago Press, Chicago (1957).
3. Blau, F.D., Kahn, L.M.: Wage Structure and Gender Earnings Differentials: An International Comparison. *Economica* **63**(250), 29-62 (1996).
4. Blinder, A.: Wage Discrimination: Reduced Form and Structural Estimates. *The J. of Hum. Resour.* **8**(4), 436-455 (1973).
5. Cain, G.G.: 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 (1986)
6. Del Río, C., Gradin, C., Cantó, O.: The Measurement of Gender Wage Discrimination: The Distributional Approach Revisited. Paper prepared for the 29th General Conference of the International Association for Research in Income and Wealth, Joensuu, Finland, August 20-26 2006 (2006).
7. Heckman, J.J.: Sample Selection Bias as a Specification Error. *Econometrica* **47**(1), 153-161 (1979).
8. Jenkins, S.P.: Earnings Discrimination measurement: A Distributional Approach. *J. of Econometrics* **61**(1), 81-102 (1994)
9. Lewbel, A.: Selection Model and Conditional Treatment Effects, Including Endogenous Regressors. Mimeo, Boston College (2001).
10. Mincer, J.: Investment in Human Capital and Personal Income Distribution. *J. of Political Economy* **66**(4), 281-302 (1958).
11. Mincer, J.: Schooling, Experience, and Earnings. NBER Press, New York (1974)
12. Ng, Y.C.: Gender Earnings Differentials and Regional Economic Development in Urban China, 1988-9. *Rev. of Income and Wealth* **53**(1), 148-166 (2007).
13. Oaxaca, R.: Male-Female Wage Differential in Urban Labor Markets. *Int. Econ. Rev.* **14**(3), 693-709 (1973)
14. Oglobin, C.G.: The Gender Earnings Differential in the Russian Transition Economy. *Industrial and Labor Relat. Rev.* **52**(4), 602-627 (1999).
15. Rice, P.: Gender Earnings Differentials: The European Experience, The World Bank, Policy Research Report on Gender and Development, Working Paper Series **8** (1999).