

# **Gender wage gap and segregation in Italy: an analysis on matched employer-employee data**

G. S. F. Bruno (Università Bocconi, Milano)

O. Dessy (Università Cattolica, Milano)

Novembre 2009

Presentato al Convegno 'Genere e precarietà' - Università di Trento

13-14 novembre 2009

Publicato da CSG (Centro Studi Interdisciplinari di Genere), Trento

<http://web.unitn.it/csg>

## Abstract

Per stimare correttamente l'esistenza e l'entità di discriminazione di genere è necessario comprendere la quota di differenziale salariale uomo-donna che rimane non spiegata dopo avere effettuato la maggior parte di controlli possibili: settore, occupazione, tipologia contrattuale, etc.. Comprendere a quale di questi livelli si determina una eventuale segregazione delle donne è cruciale per disegnare politiche rivolte alla riduzione del gap salariale uomo-donna. Dall'iniziale lavoro di Groshen (1991) si è evidenziato come l'aggiunta del livello di impresa sia importante nel determinare le caratteristiche della segregazione femminile negli US in lavori che pagano meno rispetto a quelli scelti dagli uomini, sebbene parte del gap rimanga non spiegato e quindi attribuibile a discriminazione.

La recente disponibilità e diffusione di dati del tipo 'matched employer-employee', in cui (grazie alla disponibilità di un identificativo di impresa) è possibile aggregare i lavoratori per impresa di appartenenza, ha permesso di perfezionare questo tipo di studi. In letteratura troviamo recente evidenza empirica per gli US (Bayard-Hellerstein-Neumark-Troske, 2003), Portogallo (Cabral-Cadoso-Portela, 2005), Danimarca (Gupta N. D. and D. S. Rothstein, 2005) e Spagna (Amuedo Dorantes-De la Rica, 2006) in cui si evidenzia come, nei vari paesi, il ruolo del tipo di occupazione, o del settore o dell'impresa siano diversi nello spiegare il differenziale salariale uomo-donna.

Il presente contributo è un primo tentativo di stima dell'impatto della segregazione femminile sui differenziali salariali uomo-donna in Italia per il periodo 1998-2002 utilizzando un dataset del tipo 'matched employer-employee', costruito utilizzando gli archivi INPS delle imprese (disponibili per gli anni 1998-2002) e dei lavoratori (disponibili per gli anni 1985-2004). La nostra fonte informativa è quindi costituita da una banca dati di tipo amministrativo, che copre il settore privato dell'economia. Il campione di riferimento, di dimensione 1:90, è costituito da tutti gli individui nati in 4 date dell'anno pre-definite, che vengono seguiti nel tempo all'interno dell'archivio lavoratori dipendenti. A tali osservazioni vengono poi agganciate informazioni su tutte le imprese da questi incontrate lungo il percorso lavorativo. La possibilità di seguire gli stessi individui nel tempo e tra le imprese consente di effettuare confronti sui percorsi occupazionali di uomini e donne, identificando i tipi di occupazione, i settori, le tipologie contrattuali e di impresa in cui si concentrano le attività femminili. La disponibilità poi di una misura del salario giornaliero permette di studiare l'impatto di tale segregazione sui differenziali salariali uomo-donna, e quindi la presenza di discriminazione in Italia e il livello a cui si verifica. Le informazioni disponibili su lavoratori ed imprese permettono di controllare per età, genere, paese (e provincia per gli italiani) di nascita, categoria occupazionale (impiegati, operai, apprendisti, etc. per 8 categorie), caratteristiche contrattuali (part-time/full-time, contratto a tempo determinato, indeterminato, stagionale, etc.), settore di attività (3 digit ateco81), dimensione d'impresa

Il risultato principale consiste nel riscontrare una misura di discriminazione nei confronti delle donne minore rispetto a quella stimata per altri paesi, ma che, a differenza di quanto accade per gli altri paesi, non riesce ad essere spiegata con variabili di segregazione.

## **1. Introduction**

Since the seminar paper of Groshen (1991), the study of the female/male wage differential is based on the two components of discrimination and segregation.

Groshen (1991)'s approach sets up a wage-regression on individual data, including the usual controls for experience, tenure, human capital, and firm characteristics, where the parameter for the sex-dummy usually identifies the impact of discrimination on wage. For capturing the impact of segregation, variables such as the percentage of females in the specific sectors and occupations, who identify their degree of feminization, are included in the regression. If in this second regression the parameter for sex is lower than in the regression without segregation variables, it is concluded that segregation explains gender discrimination, and therefore lower pays for women can be partly explained by their working into sectors or occupations that on average pay less than others. The originality of Groshen (1991)'s approach is due to the availability in her data of feminization rates at the firm/establishment level, so that the control for segregation could be considered also at this level. The availability of this information is usually due to the possibility of carrying out the analysis on matched employer-employee data, where the information on employees can be linked to detailed firm characteristics.

Although Groshen (1991) focuses only on a limited number of industries, she finds that in the US segregation explains the gender wage gap with a reduction from 40% to 16%, although results vary across industries and occupations. In a more recent

analysis on all industries, Bayard et al (2003) find instead that the discrimination parameter in the US decreases only from -0.375 to -0.244 when segregation variables are included, and that the firm effect is stronger than the occupation effect.

Due to a recent widespread availability of matched employer-employee data, similar analyses have been carried out also for some European countries. Gupta and Rothstein (2005) find that in Denmark the discrimination parameter decreases from -0.341 to -0.168 thanks to the inclusion of segregation variables, with a strong impact of occupation.

Dorantes and De la Rica (2006), analysing a Spanish panel that covers the 1995-2002 period, find that, with the consideration of segregation variables, the discrimination parameter decreases from -0.24 to -0.14. Similar results are found in Portugal by Cabral Vieira et al. (2005) on a very detailed matched employer-employee data set, where also strong firm-level effects can be estimated.

In this paper we carry out a similar analysis on Italian matched employer-employee data. In this preliminary analysis that covers the manufacturing sector in the 2002-2002 period we find quite surprisingly that the discrimination parameter in Italy is lower than in the other countries, but segregation variables do not help explaining the female/male wage gap.

In Section 2 we describe our data, the results are presented in Section 3, Section 4 concludes.

## 2. The data

We use a representative 1:90 sample of administrative data on employees' wages collected by the National Institute of Social Security (Istituto Nazionale di Previdenza Sociale, INPS). The sample covers the population of workers registered at INPS born on one of the four randomly chosen days of the year. We merge three separate INPS archives: the demographic archive, the employees' archive and the firms' archive. We therefore end up working with a proper 'matched employer-employee data-set'.

The demographic archive contains information at individual level on sex, day and place of birth of individuals. Through an individual code (given by INPS, where the original scal code has been anonymised) it is possible to link the demographic archive to the employees' archive, where information on workers' characteristics (such as occupation), work histories and wages is available.

Since 1998 data come from employers' fiscal declarations, and misreporting is prosecuted. In the employees' archive data are not given at the individual level, but for each single contract registered at INPS. Therefore it might happen that the same person for a given year has more than one record within the same firm, or also across different firms. We treat such cases of multiple records for the same workers by selecting only one record at random. A firm identifying code (anonymized by INPS) is available for each record, so that it is possible to link the employees' archive to the firms' archive, supplementing the data-set with information on the size of the firm, its

date of birth/death, localization and sector of activity.

The INPS data do not cover self-employment and public employment, but only the private sector excluding agriculture. For our preliminary analysis we restrict to the manufacturing sector. The original information on sectors in the INPS data is given through an internal code (Codice Statistico-Contributivo, CSC), for which INPS gives an official conversion to ATECO81 codes.

Although the employees' archive covers the 1985 - 2004 period, we can base our empirical analysis only upon the time span 2000 - 2002 since: a) firms' data are available only over the period 1998 through 2002 and b) years 1998 and 1999 are affected by severe reporting errors for the wage variables.

Table 1 presents a description of the data. The average number of individuals is 28,000 per wave, whereas the average number of firms considered is about 20,640 per wave. The percentage of men on the total sample is about 68%. 8 occupations are considered: the most represented is white collars 52%, followed by blue collars 34.5% apprentices are about 7%, managers and executives are respectively 2.9% and 2%.

**Table 1: Employees' and firms' characteristics in the matched sample**

	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>Average</b>
<i>N. Individuals</i>	28130	28571	27745	28148.67
% male	67.77	68.09	67.85	67.90
<i>Occupation (%)</i>				
Executives	2.15	2.06	2.23	2.15
Managers	2.79	2.93	3	2.91
White collars	33.89	34.03	35.59	34.50
Blue collars	52.4	52.6	51.1	52.03
Apprentices	6.78	6.45	6.22	6.48
Apprentices/White collars	0.12	0.1	0.13	0.12
Apprentices/Blue collars	0.02	0.02	0.04	0.03
Others (pilots, outworkers, middlemen, teach., journ.)	1.85	1.8	1.69	1.78
<i>N. Firms</i>	20512	20904	20494	20636.67

Table 2 shows the percentage of females over the total of people employed in the sample in each occupations, that is the ‘feminization rate’ of the occupations. We can see that highest percentage of females (about 57%) in marginal activities, 43% are among white collars, the 27% of apprentices is women and the 22% of blue-collars. Among managers, only the 14% on average are women, and among executives they are about a half: 7%.

**Table 2: Feminization rate by occupations (percentages).**

	<b>Blue-collars</b>	<b>White-collars</b>	<b>Executives</b>	<b>Managers</b>	<b>Apprentices</b>	<b>Others</b>
<b>Year</b>						
1998	23.15073	42.64602	5.946792	12.31527	27.41775	58.99844
1999	22.95437	42.76024	6.583072	13.08411	27.09404	57.00483
2000	22.8015	43.24051	6.325302	13.78911	27.55013	56.83333
2001	22.58695	43.12326	7.009346	15.45064	26.48428	54.70086
2002	21.54746	43.34095	8.071749	15.32004	26.84706	57.72059

Similar rates of segregation have been constructed at industry-level (3-digit). Unfortunately, due a low representativity of our sample at the firm-level, we cannot control at this stage for segregation at this level, and we focus only on segregation at the occupation and industry level.

### **3. The model**

We consider the following two models:

a)  $y = \alpha + \beta x + \varepsilon$

b)  $y = \alpha' + \beta' x + \gamma' fpsect + \gamma'' fpocc + \gamma''' fpsect * occ + \varepsilon$

where:

y = average daily wage

$x_i$  = explanatory variables: experience (n.years same sector), tenure(n.years same firm), firm size, time dummies.

fpsect = percentage of females in sector



fpocc = percentage of females in occupation

If  $\beta' < \beta$  segregation explains discrimination.

The OLS results are given in Tables 3 and 4. As we can see the discrimination parameter  $\beta$  estimated is -0.187 and  $\beta'$  is -0.176 although the parameters estimated for the segregation variables are all significant and the controls have all the expected signs. IV panel data methods to solve for the likely endogeneity of segregation variables are in agenda.

#### **4. Conclusions**

In this paper we have carried out a preliminary analysis of gender wage gap decomposition on Italian matched employer-employee 2000-2002 data covering the manufacturing sector.

We find that:

1. The gender wage gap in Italy is not very high if compared to other countries, being about -0.187
2. Although occupation and industry segregation of women do matter, they do not explain the gender wage gap in Italy

We conclude therefore that the low degree of gender discrimination in Italy is not a good new, because not being explained by segregation of women in low-paying occupations or sectors, is completely due to discrimination.

Further analysis will be carried out considering all industries and contract segregation,

using IV panel data methods.

**Table 3: Estimation results for model a)**

Source	SS	df	MS			
Model	12127.7177	10	1212.77177	Number of obs = 253365		
Residual	44116.9193253354		.174131529	F( 10,253354) = 6964.69		
Total	56244.637253364		.221991431	Prob > F = 0.0000		
				R-squared = 0.2156		
				Adj R-squared = 0.2156		
				Root MSE = .41729		

  

lcomp_gior	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sesso	-.1864984	.0019165	-97.31	0.000	-.1902546	-.1827422
time1	(dropped)					
time2	.0225541	.0026729	8.44	0.000	.0173153	.0277929
time3	.0337785	.0026672	12.66	0.000	.0285508	.0390062
time4	.0555082	.0026493	20.95	0.000	.0503157	.0607008
time5	.0850149	.0026202	32.45	0.000	.0798795	.0901504
ldim	(dropped)					
ldim2	.0249221	.0002029	122.81	0.000	.0245243	.0253198
eta	.0454261	.000518	87.70	0.000	.0444109	.0464413
eta2	-.0004489	6.65e-06	-67.48	0.000	-.000462	-.0004359
tenure	.0052767	.0033404	1.58	0.114	-.0012705	.0118239
tenure2	.0014958	.0005198	2.88	0.004	.000477	.0025146
_cons	2.748663	.0101181	271.66	0.000	2.728832	2.768494

**Table 4: Estimation results for model b)**

Source	SS	df	MS			
Model	12869.9661	13	989.997393	Number of obs = 253365		
Residual	43374.6709253351		.171203867	F( 13,253351) = 5782.56		
Total	56244.637253364		.221991431	Prob > F = 0.0000		
				R-squared = 0.2288		
				Adj R-squared = 0.2288		
				Root MSE = .41377		

  

lcomp_gior	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sesso	-.1753249	.0022117	-79.27	0.000	-.1796598	-.17099
time1	(dropped)					
time2	.0232527	.0026504	8.77	0.000	.018058	.0284474
time3	.0339585	.0026448	12.84	0.000	.0287748	.0391421
time4	.056857	.0026272	21.64	0.000	.0517078	.0620062
time5	.0899812	.0025999	34.61	0.000	.0848855	.095077
ldim	(dropped)					
ldim2	.023404	.0002039	114.78	0.000	.0230044	.0238037
eta	.0441633	.000514	85.92	0.000	.0431559	.0451707
eta2	-.0004316	6.60e-06	-65.36	0.000	-.0004446	-.0004187
tenure	.0037163	.0033127	1.12	0.262	-.0027765	.0102092
tenure2	.0018	.0005155	3.49	0.000	.0007897	.0028103
fpsettore	.00199	.0001058	18.81	0.000	.0017827	.0021974
fpoccupation	.0075537	.0001245	60.66	0.000	.0073097	.0077978
fpocccsett	-.0030813	.0001014	-30.39	0.000	-.00328	-.0028825
_cons	2.60078	.0104691	248.42	0.000	2.58026	2.621299

## References

Amuedo-Dorantes C. and S. De la Rica (2006), 'The role of segregation and pay structure on the gender wage gap: evidence from marche employer-employee data for Spain', *Contributions to Economic Analysis & Policy*, 5(1), Art.10.

Bayard K., Hellerstein J., Neumark D. and K. Troske, (2003), 'New evidence on Sex segregation and sex differences in wages from marche employee-employer data', *Journal of labor Economics*, 21(4), pp.887-922.

Cabral Vieira J., Cadoso A. R. and M. Portela (2005), 'Gender segregation and the wage gap in Portugal: ana analysis at the establishment level', *Journal of Economic Inequality*, 3, pp. 145 - 168.

Groshen, E. (1991), 'The structure of the female/wage differential: is it who you are, what you do, or where you work?', *The Journal of Human Resources*, 26(3), pp.457-472.

Gupta N. D. and D. S. Rothstein (2005), 'The impact of worker and establishment-level characteristics on male-female wage differentials. Evidence from Danish marche employee-employer data', *Labour*, 19(1), pp.1-34.