

Employer-Employee Data to Analyse the Male-Female Wage Gap¹

Archivi Employer-Employee per l'analisi dei differenziali salariali di genere

Saveria Capellari, Laura Chies, Susanna Zaccarin
Dipartimento di Scienze Economiche e Statistiche, Università di Trieste
saveria.capellari@econ.units.it, laura.chies@econ.units.it,
susanna.zaccarin@econ.units.it

Riassunto: La recente disponibilità di archivi *Employer-Employee* consente di considerare nell'analisi della segregazione e discriminazione di genere, oltre alle caratteristiche del lavoratore, anche quelle dell'impresa. Nel lavoro è presentata una analisi *multilevel* dei differenziali salariali tra maschi e femmine utilizzando i dati del Panel INPS per i lavoratori di età 20-25 anni, occupati nel 1996. I risultati ottenuti enfatizzano l'importanza della segregazione a livello di impresa come possibile spiegazione del differenziale salariale.

Keywords: Linked Employer-Employee data, Wage differential, Random effect model

1. Introduction

In recent years an international line of research on male-female segregation and discrimination at the workplace based on extensive use of employer-employee datasets has been developed. It is a matter of understanding whether women's disadvantages depend, to use Groshen's terms, on "who you are, what you do or where you work" (Groshen, 1991). Although Italian literature on wage differentials is rich in empirical studies, the field of the interaction between gender discrimination and the presence of segregation or selection at a company or local labour market level has yet to be fully explored. This study presents the results of an empirical analysis of gender wage differentials based on the INPS Panel for people aged 20 to 25 years employed in the private sector in 1996. The aim is to analyse the factors determining young male-female wage differentials, taking account of the characteristics of workers and firms. The wage equation was estimated separately for males and females, using a two-level random effects model to allow for firm heterogeneity (Snijders and Bosker, 1999). This specification is particularly suitable for the analysis of hierarchical data such as those in the employer-employee dataset, which contains information on people employed in the same firm. The analysis takes separate account of the effects of firms' sector and size and of the proportion of female employees in the firm. It thus enables us to assess whether and to what extent differences between companies count in the determination of wages and what effect a high proportion of female employees in a firm has on men's and women's wages.

¹ The present paper is financially supported by MIUR funds PRIN01, prot. 2001132941). We are grateful for the use of the INPS Panel to LABORatorio Revelli of Moncalieri (Turin, Italy)

2. The INPS Panel

From a methodological point of view, the linkage between the employer and the employee data can be achieved in different ways. In the spectrum of possible approaches, one has at one extremity the design of an *ad hoc* statistical survey of employers and their employees (Royce *et al.*, 2001), and at the other extremity, the exclusive use of administrative files and registers (the Scandinavian approach). The LEHD program of the U.S. Census Bureau (Prevost *et al.*, 2001) lies somewhere in between by combining both types (administrative and statistical) of data.

The INPS Panel data-set made available by LABORatorio Revelli of Turin (Contini, 2002) represents an example of the second type employer-employee archive. The archive covers the period from 1985 to 1996, and is based on the administrative data collected for institutional purposes on employers and employees by the Italian Social Security Institute (INPS). The Panel is derived from a longitudinal sample of employees born on the 10th of March, June, September and December of every year and registered in the INPS files in the period from 1985 to 1996. The company's longitudinal records are linked to each worker in the sample. Each yearly sample includes approximately 100,000 employees and their firms.

Information on individuals collected in the Panel comprises usual employee characteristics (gender, year of birth, occupation, hierarchical level, earnings, ...) and some aspects of his/her working history (dates of entry, number of paid month, changes in occupation, employer and sector; non-working periods). The archive does not provide information on education level and family background (especially number and age of children). As regards education a proxy may be defined for youngest people as to the possession of a high-school diploma by reference to first-job entry after the age of 18. Information on companies is confined to the average size of the firm, sector of activity, location and the age of the firm. Since the Panel do not provide a gender breakdown of employees by firm, for an analysis of the effect of proportions of female employees recourse was made to the figures of 1997 Intermediate Industry Census (Lucifora and Reilly, 1990). The mean values obtained for firms categorized according to size, detailed information on sector and province of location, were imputed to the firms in the selected dataset. Our data cover all workers born between 1971 and 1976 who began working between 1986 and 1991 and were employed in the private sector on December 1st 1996. The sample of employees comprised 14,596 people, 6,164 women and 8,432 men, working in 13,136 companies. One of the most important differences between males and females regards first-job entry and it may be correlated with the possession of a high-school diploma: 78% of females entry in the job after the age of 18, compared with 66% of males. This difference was reflected in occupational rank, with a higher concentration of women in white-collar professions and a greater proportion of men in blue-collar occupations. Overall, 14% of workers were employed on trainee contracts; the proportion of men was slightly higher. The other working history elements in the sample showed basic similarities between the genders: specifically, the values for experience and mobility were very close.

Gender gap was calculated on annual earnings which are proportional to the number of days worked. Gender earnings ratio is about 90%. Though rather low, this differential is not in contradiction with other estimates made for Italy, since it is known that differentials increase with age as a result of women's discontinuous presence in the labour market and possible vertical segregation.

Significant differences emerged in terms of the distribution of males and females classified on the proportion of females employed in the firm, which confirmed the presence of segregation in Italian companies. The employees of companies with a high proportion of women showed longer attachment to the company and had greater experience (an average of four months more), but despite this had wage levels 16% lower than employees in male-dominated firms. Furthermore, in firms with the highest proportion of females, training and occupational ranks were lower; the firms themselves tended to be of medium size and over 70% of them operated in traditional manufacturing and about 18% in the personal services. The corresponding figures for male-dominated firms are respectively 30% and 5%. As far as firm characteristics are concerned, over 65% of the sample was made up of firms with a low proportion of female employees. Firms with higher proportions of females tended to be located in north-east or southern Italy.

3. Random effect model and main results

For an assessment of the effects of individual and company characteristics on individuals' earnings, the wage equation was set out (separately for males and females) as follows:

$$W_{ij} = X_{ij}\beta + Z_j\alpha + u_j + e_{ij} \quad i = 1, \dots, n_j, \quad j = 1, \dots, J; \quad (1)$$

$$e_{ij} = N(0, \sigma_e^2), \quad u_j = N(0, \sigma_u^2)$$

where W_{ij} is the logarithm of the annual earnings of worker i employed in company j , X_i and Z_j are the vectors of individual and firm characteristics respectively and β and α are the vectors of the associated coefficients. In equation [1] the effect of individual and company characteristics are assumed to be constant for all workers, whereas variability between firms is given by the random term u_j . The coefficient estimates were obtained using the residual maximum likelihood (REML) method (Snijders and Bosker, 1999). A model was also estimated for the whole sample and the coefficient of the dummy variable for sex was significant. The proportion of variance of the annual paid wage attributable to the firm (measured by the intra-class correlation coefficient and evaluated starting from the initial model without any explanatory variables) is equal to 46.5% for males and 42.2% for females respectively. After controlling for the workers' characteristics, the firm's variables explain 60% of firm's variance for males and 59% for females. It should be pointed out that because of the particular construction of the INPS Panel (Contini, 2002), derived essentially from a population of workers, the dataset contains a very high proportion of firms represented by a single worker. Empirical analysis conducted on the sole basis of firms with at least two 20-25 year-old workers in the dataset and estimates run on the entire dataset breaking down companies by size confirm the company effect results obtained from the totality of 20-25 year-old workers in the 1996 archive. The effect of individual variables was as expected: wages increased when first-job entry was after the age of 18 (+5% for women and +13% for men), a white-collar position (+28.6% for women and +26.6% for men) and experience (+20% for women and +22% for men), which is consistent with the results generally presented in the literature. Mobility between firms proved to be significant, penalising

employees of both genders (-5.6% for women and -3.7% for men). As emerged in other analyses of INPS data, part-time employment proved to penalise both genders, especially in blue-collar occupations, though the effect in our estimates was more marked. For blue-collar workers, male and female alike, the wage was halved, while for white-collar employees it went down by 51% compared to full-time employees. Part-time work, however did not appear to penalise women (-48.5%) more than men. The effect of being employed on a trainee contract was positive with the exception of male blue-collar part-time employees (of whom there were very few). This is explained by the fact that the sample includes workers with discontinuous work experience. Compared to the latter category, young workers on trainee contracts stay longer in their jobs and are better paid. With a few exceptions, company variables proved to be significant. Wages were lower for both genders not working in heavy and light engineering, with the exception of the energy and chemical industries. Working in services, especially transport and communications and personal services, proved to penalise women. Both for males and females, the location of a firm in southern Italy as opposed to north-west proved to have a negative effect on wages. Significant gender differences did emerge in some respects: company location in north-east Italy proved to be important for women (+5.1%) and in central Italy negatively for men (-5.8%). Firm size had a positive, and slightly more marked, effect for men (+21%, as against 18.5% for women).

The hypothesis that the proportion of women employees in a firm has an effect on wages was confirmed by the data on women (-9.3%) and men (-11.9%) alike. The proportion of women employees did not appear to interact significantly with the company size variable and with the economic sectors.

These first results configure an important role for the gender composition of a firm in determining the male-female wage differential, which is consistent with the results carried out in other countries. Further research in this direction would benefit greatly from an extension of the dataset with more information on firms (part of which is already being gathered) and firms' employment strategies, as well as more precise data on individual characteristics.

References

- Contini B. (2002), *Osservatorio sulla mobilità del lavoro in Italia*, Il Mulino, Bologna.
- Groshen E. L. (1991) The Structure of the Female/Male Wage Differential: Is It Who you are, What You do, or Where You Work?, *Journal of Human Resources*, 26, 3, 457-472.
- Lucifora C., Reilly B. (1990) Wage Discrimination and Female Occupational Intensity, *Labour*, Vol. 4, 2, 147-168.
- Prevost R., Lane, J., Abowd J. (2001) The Longitudinal Employer-Household Dynamics Program at the U.S. Census Bureau, *Bulletin of the International Statistical Institute*, Book 2, 401-404.
- Royce D., Hidirologlou, M., Lavallée, P., Patak, Z. (2001) Methodological issues in Canada's Workplace and Employee Survey, *Bulletin of the International Statistical Institute*, Book 2, 397-400.
- Snijders T. A. B., Bosker, R. J. (1999) *Introduction to Basic and Advanced Multilevel Modelling*, Sage, London.