May 2002

An Investigation of the Relationship between Job Characteristics and the Gender Wage Gap

Jaume García[1]
Pedro Jesús Hernández[2]
Angel López[1]

[1] Departament d'Economia i Empresa, Universitat Pompeu Fabra
[2] Departamento de Fundamentos del Análisis Económico, Universidad de Murcia

We gratefully acknowledge financial support from DGES project PB98-1058-C03-03 and from the *Instituto de la Mujer*, Project VS/2001/0497.

Running head: Job Characteristics and the Gender Wage Gap

Abstract

This paper re-examines gender wage differences, taking into account not only worker characteristics but also job characteristics. Consideration of a wide set of "job quality" indicators can explain a fraction of the wage gap that would otherwise be attributed to pure wage discrimination. In any case, the fraction of the wage gap that remains associated to differential rewards for identical factors across sexes is still substantial. Our results suggest that in order to avoid overestimation of the fraction of the wage gap attributable to discrimination, it is necessary to control for job characteristics.

Keywords: wage differentials

JEL classification: J7, C4

1. Introduction

In most gender wage gap evaluations, the analysis has focused on the worker characteristics associated with different wages for men and women, independently of the type of job done. This paper proposes a complementary approach, asking if and by how much the same job characteristic is rewarded differently for men and women. If wages are related to productivity, and productivity influences promotion up the job ladder, then some job characteristics will have explanatory power in addition to individual proxies for productivity, such as education. Indeed, many critiques in the literature on wage discrimination focus on the fact that in wage regressions, sex dummy variables (or alternatively, the differences in coefficients in separate regressions for men and women) are just picking up the effect of unobserved productivity differentials between genders. The analysis in this paper intends to disentangle these effects by means of enriching the specification of the wage equation to account for job characteristics.

Our approach is related to that adopted in studies addressing the issue of occupational segregation in that we intend to control for the fact that assignment of women to jobs is not the same as that of men. Indeed, the majority of studies on segregation find that part of the wage gap may be explained by the fact that women tend to concentrate in low wage occupations. However, even after controlling for segregation, Miller (1987b) and Hernández (1996) find that there remains a substantial proportion of unexplained wage differences. One hypothesis arising from their results is that there is further (unobserved) segregation within each of the considered occupations. The difficulty of testing this hypothesis is rooted in the fact that information on job characteristics is usually scarce in labour surveys. This study uses information from a database, the "*Encuesta sobre Estructura, Conciencia y Biografía de Clase*" (1991), comprising responses to a series of questions designed to capture precisely the job characteristics related to the commonly "unobserved" job segregation mentioned above.

The idea motivating inclusion of these variables in a wage gap decomposition exercise is clear: if, once job characteristics have been fully taken into account, there is no unexplained wage gap then the discrimination hypothesis will appear less likely. Whether access to these job characteristics is itself a result of differential treatment of men and women in promotion and job assignment is, of course, another issue. If, on the contrary, one finds differential rewards for the same job characteristic, then our inference will be quite the opposite.[1]

We specify a model accounting for the fact that the wage equation can have sample selection problems due to participation and that individuals sort themselves into different occupations. Wages are determined by several job characteristics and individual variables such as age and education. From the model estimates, we shall implement the wage decomposition procedure proposed by Neumark (1988). Given the presence of two selection processes in our model, we shall pay special attention to how decompositions of the wage gap need to be carried out in the presence of non random assignment to different groups in the labour market. In particular we shall follow the procedures proposed by Neuman and Oaxaca (1998) in carrying out the decompositions.

Our results suggest that job characteristics are important factors in explaining wages even when controlling for individual characteristics. Moreover, when we account for job characteristics, the fraction of the gender wage gap attributable to differential rewards for men and women is reduced, reflecting the fact that men tend to be assigned to the "best" jobs. However, there remains a substantial and significant "discriminatory" component, in that the reward for job and individual characteristics is higher for men.

In section 2 we present the econometric model. Section 3 comprises discussion of the data set. The empirical results are presented in section 4, while, finally, section 5 concludes.

[1] This study is also related to studies such as Bayard *et al* (1999) and De la Rica and Felgueroso (2001) which control for the proportion of females within each occupational class, each industry and each firm. Also, work by Johnson and Solon (1986), Sorensen (1990) and Macpherson and Hirsch (1995) show that wages are negatively associated with the proportion of women within the worker's occupational class.

2. The econometric model

The objective is to estimate wage equations for both males and females in order to decompose the observed wage gap into its discrimination component and the part attributable to the different observed male and female characteristics. As usual, to obtain consistent estimates for the wage equation we have to control for the potential selectivity problem generated by the fact that wages are only observed for those who are participants.[2] Additionally, we also wish to control for the fact that having a particular occupation can also be endogenous, in the sense that the unobserved factors affecting the choice of occupation may be related to the unobserved productivity in the wage equation. Consequently, the model we are considering comprises three equations. The endogenous variables of these equations are *w, L,* and *I,* where *w* is the logarithm of the wage, *L* is a binary indicator for participation in the labour market and *I* is a categorical variable signalling occupation group. *I* and *w* are only observed when *L* is equal to one, i.e., when the individual is a worker. The model can be written as:

$$w_i = X_i^{'}\beta + u_i \qquad (1)$$

$$L_i = 1(Y_i^{'}\alpha + \eta_i > 0) \qquad (2)$$

$$I_i^{*j} = Z_i^{'}\gamma^j + v_i^j, \quad I_i = j \Leftrightarrow I_i^{*j} = Max(I_i^{*1},..., I_i^{*M}), \quad j = 1,..., M \qquad (3)$$

where *X, Y* and *Z* are vectors of explanatory variables for each equation and *u,* η and v^j are the corresponding error terms which are assumed to have zero mean and the following covariance structure:

$$\Omega^j = \begin{pmatrix} \sigma_u^2 & \sigma_{\eta u} & \sigma_{uv^j} \\ \sigma_{\eta u} & \sigma_v^2 & 0 \\ \sigma_{uv^j} & 0 & 1 \end{pmatrix} \quad j = 1,.., M$$

---

2 Méndez and Hernández (2001) analyse the sample selection bias that could arise in the estimation of an earnings equation in economies where unemployment is a relevant feature of the labour market, as is the case in Spain. The differences between non-participant and unemployed people point to a double sample selection mechanism.

In order to estimate the model above we assume a probit structure for the participation equation (2) and a multinomial logit structure for the occupation equation (3). We estimate the wage equation (1) separately for females and males by OLS. In order to account for the selectivity and endogeneity issues mentioned above we include correction terms following the procedures proposed by Heckman (1979)[3] and Lee (1983), but, except for the constant term, we assume the same model for all occupations; i.e.,

$$w_i = X'_i \beta + \sigma \lambda_i + \sum_{j=1}^{M} d_i^j * \varphi^j \frac{\phi\left[J\left(Z'_i \gamma^j\right)\right]}{F\left(Z'_i \gamma^j\right)} + u_i^{'}$$

where

$$J(Z'_i \gamma^j) = \Phi^{-1}\left(\frac{\exp(Z'_i \gamma^j)}{1 + \sum_{s=2}^{M} \exp(Z_i^{'} \gamma^s)} \right)$$

$$F\left(Z'_i \gamma^j\right) = \frac{\exp(Z'_i \gamma^j)}{1 + \sum_{s=2}^{M} \exp(Z_i^{'} \gamma^s)}$$

$$\lambda_i = \frac{\phi\left(Y'_i \alpha / \sigma_\eta\right)}{\Phi\left(Y'_i \alpha / \sigma_\eta\right)}$$

$$\sigma = \sigma_u \rho_{\eta u}; \quad \varphi^j = \sigma_u \rho_{\eta v^j}; \quad d_j^i = 1(i \in j)$$

where $\Phi(\cdot)$ y $\phi(\cdot)$ are the p.d.f. and c.d.f. of the standard normal distribution, respectively.

[3] The majority of studies that analyse the relationship between wages and occupations do not correct for potential selectivity biases. See Trost and Lee (1984), Miller (1987a and 1987b), Reilly (1991) and Hernández (1996), among others, and Dolton *et al* (1989) as an exception.

3. Data and variables

Our estimating sample is extracted from the *Encuesta sobre Estructura, Conciencia y Biografía de Clase* (1991).[4] This survey contains abundant information about socio-economic variables and it is unique, at least as far as the Spanish labour market is concerned, in the sense that it contains a great deal of information on job characteristics.

The endogenous variable in the wage equation (1) is the (log) hourly wage and that of the participation equation (2) is a discrete variable taking the value 1 if the observation corresponds to a worker and 0 otherwise. On the other hand, when defining the occupational groups corresponding to the endogenous variable in equation (3), we have followed the official classification of occupations in Spain (*Clasificacíón Nacional de Ocupaciones,* 1979) with some slight changes due to small samples for some of the original cells. Namely, the classification we use is the following:

Occupation 1: Technicians, Directing staff in the public and private sector
Occupation 2: Administrative staff
Occupation 3: Sales representatives and retail staff
Occupation 4: Restaurant and hotel staff
Occupation 5: Agriculture, fisheries and mining, and manual industrial workers.

The explanatory variables used in the different equations are defined as follows:

Age: Number of years
Seniority: Number of years in the present job
Indefinite contract: Dummy variable equal to one when the individual has an indefinite contract.
Education: Dummy variables for the following educational levels: illiterate, primary school, secondary school, vocational training, university degree (intermediate) and university degree (higher).
Years of education
Married: Dummy variable equal to one if the individual is married.

[4] The survey was carried out by the *Instituto Nacional de Estadística,* the *Comunidad Autónoma de Madrid* and the *Instituto de la Mujer*. See Carabaña *et al.* (1992) for more detailed information.

Head of the household: Dummy variable equal to one if the individual is the head of the household.

Number of children

Number of dependent adults

Number of income earners

Industrial sector: Dummy variables corresponding to classification in 12 sectors we have defined.

Region: Dummy variables for each of the 17 Spanish regions (*comunidades autónomas*)

Gross wage control: Dummy variable equal to one when reported wages are gross wages.

Regarding job characteristics, we have distinguished five blocks of variables: i) those related to the degree of worker autonomy in setting the working pace within the workplace, ii) those related to the degree of control by others over the worker's output, iii) those related to the degree of supervisory/directive powers over other employees, iv) those related to the power to decide on issues related to other employees and v) those related to the education mismatch. The definition of the specific variables included in each of these groups is presented in Table 1. Unless otherwise specified, these are binary indicators (Yes=1; No=0).

(TABLE 1)

Table A.1 in the Appendix contains descriptive statistics of all the variables used in the empirical analysis for wage earners split by gender. The average male hourly wage is 14.31% higher than that for women.[5] On average, men have also accumulated more seniority (11.01 versus 7.67 years) and are more likely to have an indefinite (non fixed-term) contract than women. There are also noteworthy differences between men and women in educational levels and sectoral concentration. Firstly, the level of education would appear to be higher for women than for men and, secondly, we find that women are more likely to work in the Public Administration, the so-called "Reproduction" sector (teaching, scientific research and health

[5] When making the wage decompositions we will approximate this differential by the difference of the (log) wages, which is 13.54%.

care) and the Social Services and Domestic Work sector. We also observe that men are more likely to be in occupation 5 whereas women mainly seem to cluster in occupations 1 and 2.[6]

As for the variables related to job characteristics, we find that men occupy the majority of directive and supervising positions. Up to 7.5% of employed men occupy these positions as opposed to only 2.2% of women. Also, while more than 15.7% of men carry out supervision tasks, only 11.5% of women do so.

[6] Note that occupation 1 includes, among others, health care technicians. This is an occupation with a high percentage of women.

4. Results

The explanatory variables included in the reduced form participation equation are age and its square, marital status, worker education and education of partner, whether the individual is the head of the household, number of children, number of dependent adults, number of income earners in the household and regional dummies. The estimation results are reported in Table A.2 in the Appendix. As expected, there are substantially different patterns for the participation equations of both males and females. Education has a more important and significant effect for females. The number of children has a negative and significant effect on female participation, whereas for males the effect is positive, although not significant. This result is similar to the effect of being married, which is negative, although not significant at a 5% significance level.

The vector of explanatory variables in the occupation equation includes individual age, years of education, years of education of father and marital status. The estimation results are reported in Table A.3 of the Appendix. The number of years of education is the most relevant explanatory factor for occupational choices. There are some differences between males and females models, although the patterns are similar.

We use the estimates of the participation and occupation equations to construct the correction terms to be included in the wage equations. We consider four different specifications for the wage equation which have a common set of variables but differ on inclusion/exclusion of some individual characteristics and/or job characteristics.

The set of common variables are: having an indefinite contract, seniority, the type of occupation[7] and gross wage control plus correction terms. The group of additional individual characteristics includes age and its square, educational level, industrial sector and region. The first two variables try to capture aspects related to productivity not considered in job evaluation, because they are individual not job characteristics, and the last two try to capture industrial and geographical differentials in wages. Finally, the set of variables related to job characteristics, which are taken into account in a job evaluation process, are as defined in Table 1 in the previous section.

[7] These dummies are included to allow for different constant terms depending on the occupation, as mentioned above.

Model 1 includes the common set of variables plus those related to individual characteristics and will correspond to a standard specification of a wage equation in the analysis of gender wage discrimination. Model 2 also includes the variables related to job characteristics. By comparing models 1 and 2 we can evaluate the importance of job characteristics in the determination of wages and compare to what extent differences in the characteristics of the jobs occupied by males and females explain part of the observed wage gap. Models 3 and 4 have the same specification as models 1 and 2, respectively, but exclude the set of variables corresponding to individual characteristics. The estimation of models 3 and 4 will allow us to assess the importance of considering job characteristics when analysing gender wage discrimination in a context where the explanatory factors are basically those considered in job evaluation and those which can have an effect on wages through specific wage complements, such as seniority or having an indefinite contract. Additionally, we will be able to assess again the impact of individual characteristics on wages, but for a different reference model.

In Table 2 we present the estimates from the four specifications of the wage equation for both males and females. Having an indefinite contract brings a higher reward for females, although the difference is almost insignificant when not including the standard wage equation variables apart from the common variables. The effect of the experience in the current firm and also the overall experience proxied by age is more important for males than for females. The effect of education on wages is more significant for males than for females. By comparing models 1 and 2, we can infer a positive association between having a higher educational level and better job characteristics. On the other hand, there seems to be more variability of wages across industries for females.

Differences in wages due to type of occupation are reduced when age, education and the sectoral and regional dummies are included, in particular for males. In fact, there are no significant differences between the type of occupation for males in the model and the most complete specification (Model 2), although significant differences are found for females in all four models considered.

Including job characteristics in the wage equation increases the explanatory power of the model significantly. In the case of females, when we test for the joint significance of the job characteristic coefficients, we obtain an F statistic of 2.53 when comparing models 1 and 2, and 3.47 when comparing models 3 and 4, above the corresponding critical value (1.75) at a 5% significance level. In the case of the male equations, these values are 8.13 and 11.29, respectively, indicating that job characteristics have a more substantial effect on wages than in the case of females.

Variables related to the degree of autonomy have a higher (and significant) effect in the case of males. Also, a component of directive power over employees and power to make decisions on employee wages, is associated with higher wages for males.[8] Finally, being overeducated in terms of the educational requirements of the job has a positive and significant reward for females but has no effect on male wages.

Notice that when comparing models 2 and 4, i.e., when testing the significance of excluding the other individual characteristics in the most general model (Model 2), we obtain an F statistic of 4.92 for females, which is higher than the 2.53 obtained when comparing models 2 and 1, i.e., the significance of excluding the job characteristics in Model 2. We find the opposite result for males, an F statistic of 6.07 when comparing models 2 and 4 and 8.13 when comparing models 2 and 1. This means that the average increase in the residual sum of squares per each extra parameter not estimated is higher for the other individual characteristics for females, whereas in the case of males it is higher for the job characteristics.[9]

(TABLE 2)

[8] Notice that the negative and almost significant coefficient for this latter variable in the case of females is due to the very few cases in the female subsample with a job of these characteristics, as can be seen from the descriptive statistics in Table A.1.

[9] If we look at the adjusted R^2 the reduction is smaller for both males and females when excluding the job characteristics. This does not contradict the F statistics analysis because the number of restrictions is different.

The model estimates can be used to implement the wage decomposition procedure proposed by Neumark (1988). This method is a generalisation of those proposed by Oaxaca (1973) and Blinder (1973), which does not assume either the male or the female wage structure as the non discriminatory ideal. Instead, the latter is obtained from an estimation using the joint sample.

In short, this procedure decomposes the difference between average (log)wages for men and women into differences in observable characteristics and differences in rewards for these characteristics. In our case, given the existence of the correction terms, this wage gap decomposition takes the following form:

$$
\begin{aligned}
\overline{w}^m - \overline{w}^f = & \left[(\overline{X}^m - \overline{X}^f)'\hat{b} + (\overline{\lambda}^m - \overline{\lambda}^f)\hat{\sigma}^b + \sum_{j=1}^{M} \hat{\varphi}_j^b \left\{ \overline{\left(\frac{\phi}{F}\right)_j^m} - \overline{\left(\frac{\phi}{F}\right)_j^f} \right\} \right] + \\
& \left[\overline{X}'^m (\hat{\beta}^m - \hat{b}) + (\hat{\sigma}^m - \hat{\sigma}^b)\overline{\lambda}^m + \sum_{j=1}^{M} (\hat{\varphi}_j^m - \hat{\varphi}_j^b) \overline{\left(\frac{\phi}{F}\right)_j^m} - \right. \\
& \left. \overline{X}'^f (\hat{\beta}^f - \hat{b}) - (\hat{\sigma}^f - \hat{\sigma}^b)\overline{\lambda}^f - \sum_{j=1}^{M} (\hat{\varphi}_j^f - \hat{\varphi}_j^b) \overline{\left(\frac{\phi}{F}\right)_j^f} \right]
\end{aligned}
$$

where $\hat{b}$ is the Neumark estimator for the vector of coefficients associated to X_i and $\hat{\sigma}^b$ and $\hat{\varphi}^b$ are the estimates for the parameters associated to λ_i and $(\phi /F)_i$. The indices *m* and *f* refer to the subsample of men and women respectively. The first element on the right hand side of the last equation represents the part of the wage gap that can be explained by different attributes of men and women whereas the second term represents the "unexplained" part of the wage gap, that is, the part usually attributed to gender discrimination, which arises from a differential reward for the same observable characteristics.[10]

We follow one of the possibilities presented in Neuman and Oaxaca (1998) and consider differences in the coefficients for the selection terms (participation and occupation) as manifestations of discrimination. The differences in the average values of these terms are considered as differences in characteristics. However, as Neuman and Oaxaca indicate, a discriminatory component could be extracted from this: the difference between the correction term for females and what it would be if the parameters that govern their selection process were the same as those for males. In this case, the part attributed to discrimination would be greater than our estimation.

In Table 3 we present the decomposition of the observed wage gap between males and females (13.54%) in the two terms mentioned above: the part corresponding to differences in characteristics and the part corresponding to discrimination.

(TABLE 3)

Notice that when using the specifications without the job characteristics (models 1 and 3), the proportion of the wage gap which can be considered as discrimination exceeds 100%.[11] There are no substantial differences associated with inclusion or exclusion of the rest of the individual characteristics (Model 1 vs Model 3, Model 2 vs Model 4). However, when including the job characteristics as explanatory factors in the wage equation the proportion of the wage gap attributable to discrimination is reduced, although there is still a large portion of the wage gap which cannot be explained by either individual differences or differences in job characteristics (84.7% in Model 2 and 80.9% in Model 4).

[10] Decomposition of the sample selection correction terms follows the procedures laid out in Neuman and Oaxaca (1998).
[11] Similar results are obtained by De la Rica and Ugidos (1995) using a data set from the same survey.

5. Summary

In this paper we have attempted to examine the gender wage gap under the assumption that in addition to individual worker characteristics, job features also contribute to explaining wage differences. Our strategy was to enrich the traditional wage equation specification with an ample set of indicators accounting for important job characteristics. As expected, the results for the sequence of models that we have estimated confirm that these characteristics explain a significant proportion of wage variation. Moreover, once these characteristics are taken into account, the portion of the wage gap attributable to discrimination is reduced. This reflects the fact that a higher proportion of men attain positions of greater responsibility, autonomy, degree of control over work processes etc. than women. However, our estimates also suggest that women are rewarded differentially when they achieve these positions too.

These findings are important in that they support the view that, in the Spanish labour market, differential rewards exist for men and women even when they do the same job, i.e., there is pure wage discrimination. This complements previous Spanish evidence showing that men and women with equal characteristics (but not necessarily doing the same job) received different wages. A number of issues merit further research and they all have to do with the fact that even with this data set it is impossible to find a perfect case-control situation. The first issue is that the set of characteristics controlled for does not include all relevant work characteristics. The standard job evaluation methods suggest many variables that could pick up relevant characteristics, but unfortunately our data set does not include them. Also, the answers in our data set are provided by the worker and, as such, there may be a certain degree of subjectivity. One way to overcome these shortcomings would be construction of databases which record all relevant worker characteristics, possibly within the firm itself.

References

Bayard, K., Hellerstein, J., Neumark, D. and Troske, K. (1999): "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data", NBER Working Paper No. W7003.

Blinder, A.S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources,* 8: 436-55.

Carabaña, J. *et al.* (1992): "Encuesta de Estructura, Conciencia y Biografía de Clase: Informe Técnico", Documento de Trabajo, Comunidad Autónoma de Madrid, Consejería de Economía.

De la Rica, S. and Ugidos, A. (1995): "¿Son las diferencias en capital humano determinantes de las diferencias salariales observadas entre hombres y mujeres?, *Investigaciones Económicas*, 19: 395-414.

De la Rica, S. and Felgueroso, F. (2001): "Intrafirm gender wage differences: A worker's matching approach". Mimeo.

Dolton, P.J., Makepeace, G.H. and Van Der Klaauw, W. (1989): "Occupational Choice and Earnings Determination: The Role of Sample Selection and Non-Pecuniary Factors", *Oxford Economic Papers,* 41: 573-94.

Heckman, J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica,* 47: 153-61.

Hernández, P.J. (1996): "Segregación ocupacional de la mujer y discriminación salarial", *Revista de Economía Aplicada,* 11: 57-80.

Johnson, G. and Solon, G. (1986): "Estimates of the Direct Effects of Comparable Worth Policy", *American Economic Review,* 76: 1117-25.

Lee, L.F. (1983): "Generalized Econometric Models with Selectivity", *Econometrica,* 51: 507-12.

Méndez, I. and Hernández, P.J. (2001): Labour supply, sample selection and wage discrimination, 13th EALE Annual Conference, 13-16 September 2001.

Macpherson, D.A. and Hirsch, B.T. (1995): "Wages and Gender Composition: Why Do Women's Jobs Pay Less?", *Journal of Labor Economics,* 13(3): 426-71.

Miller, P. (1987a): "Gender Differences in Observed and Offered Wages in Canada, 1980", *Canadian Journal of Economics*, 20: 225-44.

Miller, P. (1987b): "The Wage Effect of the Occupational Segregation of Women in Britain", *Economic Journal,* 97: 885-96.

Neuman, S. and Oaxaca, R.L. (1998): “Estimating Labor Market Discrimination with Selectivity Corrected Wage Equations: Methodological Considerations and An Illustration from Israel”. Mimeo.

Neumark, D. (1988): “Employers Discriminatory Behavior and the Estimation of Wage Discrimination”, *Journal of Human Resources,* 23: 279-95.

Oaxaca, R. (1973): “Male-Female Wage Differentials in Urban Labour Markets”, *International Economic Review*, 14: 693-709.

Reilly, B. (1991): “Occupational Segregation and Selectivity Bias in Occupational Wage Equations: an Empirical Analysis Using Irish Data”, *Applied Economics,* 23:1-7.

Sorensen, E. (1990): “The Crowding Hypothesis and Comparable Worth”, *Journal of Human Resources,* 25: 55-89.

Trost, R.P. and Lee, L.F. (1984): “Technical Training and Earnings: a Polychotomous Choice Model with Selectivity”, *Review of Economics and Statistics,* 66:151-56.

APPENDIX

(TABLE A.1)

(TABLE A.2)

(TABLE A.3)

Table 1: Definition of the job characteristic variables

Variables related to the worker's autonomy
Own work: Can make decisions on important aspects of own work
Working pace: Can reduce working pace in working day
Working hours: Can make decisions on working hours
Variables related to the degree of control by others over the worker's output
Quality control: Degree of control over output quality (1: very easy 2: quite easy. 3: easy. 4: not easy at all)
Quantity control: Degree of control over output quantity (1: very easy 2: quite easy. 3: easy. 4: not easy at all)
Variables related to the degree of supervisory/directive power over other employees
Director: Director position
Supervisor: Supervisor position
Variables related to the employee's power to make decisions related to other employees
Raising a worker's salary: Maximum influence in the decision to raise a worker's salary.
Penalising a worker: Maximum influence in penalising a worker.
Changing methods: To make decisions regarding changes in the procedures or basic working methods
Variables related to the education mismatch
Education more than sufficient: Education more than sufficient for the present work.
Education sufficient: Educational level sufficient for the present work.

Table 2: Estimation results for the wage equation

	Model 1				Model 2				Model 3				Model 4			
	Females		Males		Females		Males		Females		Males		Females		Males	
Variables	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*	*Coef.*	*t-stat.*
Common variables																
Constant	4.911	(21.16)	5.300	(9.21)	4.910	(20.28)	5.536	(9.92)	5.518	(41.69)	4.982	(9.12)	5.290	(36.01)	5.246	(10.05)
Indefinite contract	0.111	(2.86)	0.064	(1.76)	0.092	(2.39)	0.057	(1.62)	0.123	(3.13)	0.112	(3.07)	0.102	(2.61)	0.088	(2.54)
Seniority	0.005	(1.67)	0.008	(4.58)	0.005	(1.64)	0.008	(5.04)	0.005	(1.89)	0.006	(4.26)	0.006	(2.25)	0.007	(5.05)
Gross wage control	0.305	(7.95)	0.137	(4.45)	0.307	(7.96)	0.121	(4.05)	0.302	(7.53)	0.182	(5.74)	0.286	(7.18)	0.125	(4.08)
Occupations:																
Occupation 1	0.663	(4.03)	0.431	(0.79)	0.653	(3.88)	0.224	(0.42)	1.239	(9.07)	1.983	(3.62)	1.159	(8.33)	1.510	(2.88)
Occupation 2	0.433	(2.27)	0.260	(0.46)	0.446	(2.34)	0.134	(0.25)	0.583	(3.48)	1.729	(3.07)	0.530	(3.20)	1.309	(2.44)
Occupation 3	0.491	(1.70)	-0.605	(0.75)	0.473	(1.65)	-0.913	(1.17)	0.412	(1.46)	-0.055	(0.06)	0.411	(1.48)	-0.583	(0.72)
Occupation 5	0.467	(2.10)	0.063	(0.12)	0.479	(2.16)	-0.018	(0.04)	0.291	(1.34)	0.993	(1.82)	0.328	(1.52)	0.729	(1.40)
Other individual variables																
Age	0.017	(1.77)	0.034	(3.61)	0.015	(1.50)	0.021	(2.30)								
Age squared (/100)	-0.021	(1.75)	-0.041	(3.59)	-0.018	(1.48)	-0.026	(2.32)								
Education:																
Primary school	-0.108	(1.44)	0.048	(1.02)	-0.114	(1.53)	0.032	(0.72)								
Secondary school	-0.027	(0.26)	0.204	(2.88)	-0.067	(0.65)	0.137	(1.98)								
Vocational training	-0.085	(0.83)	0.116	(1.57)	-0.089	(0.88)	0.070	(0.97)								
University degree (interm.)	0.140	(1.27)	0.342	(4.06)	0.118	(1.07)	0.284	(3.45)								
University degree (higher)	0.303	(2.55)	0.577	(6.23)	0.224	(1.87)	0.473	(5.21)								
Industrial sector:																
Agriculture	0.080	(0.29)	0.070	(0.37)	0.074	(0.27)	0.009	(0.05)								
Basic industry	0.523	(5.00)	0.348	(1.94)	0.509	(4.84)	0.248	(1.44)								
Heavy industry	0.464	(3.94)	0.364	(2.04)	0501	(4.26)	0.319	(1.86)								
Light industry	0.196	(1.86)	0.178	(1.01)	0.209	(2.00)	0.107	(0.63)								
Building	0.464	(3.38)	0.316	(1.76)	0.476	(3.49)	0.246	(1.43)								
Retailing and Hotels	0.347	(3.82)	0.091	(0.52)	0.350	(3.84)	0.030	(0.18)								
Transport	0.388	(3.12)	0.380	(2.11)	0.357	(2.82)	0.321	(1.86)								
Economics services	0.320	(3.01)	0.425	(2.40)	0.332	(3.14)	0.335	(1.97)								
Public Administration	0.538	(5.98)	0.324	(1.86)	0.560	(6.23)	0.283	(1.70)								
Reproduction	0.397	(4.11)	0.210	(1.16)	0.412	(4.29)	0.260	(1.50)								
Social services	0.400	(3.34)	0.015	(0.07)	0.394	(3.31)	-0.018	(0.10)								

Region:								
Andalucia	0.305 (2.47)	-0.338 (3.16)	0.265 (2.09)	-0.186 (1.73)				
Aragón	0.188 (1.41)	-0.294 (2.53)	0.125 (0.93)	-0.124 (1.07)				
Asturias	0.081 (0.57)	-0.243 (1.99)	-0.017 (0.12)	-0.082 (0.67)				
Baleares	0.260 (1.84)	-0.446 (3.19)	0.248 (1.75)	-0.188 (1.33)				
Canarias	-0.108 (0.73)	-0.581 (4.55)	-0.095 (0.63)	-0.399 (3.12)				
Cantabria	0.270 (1.54)	-0.383 (2.78)	0.227 (1.27)	-0.183 (1.33)				
Castilla-La Mancha	0.053 (0.34)	-0.359 (2.72)	0.005 (0.03)	-0.161 (1.23)				
Castilla-León	0.038 (0.26)	-0.344 (2.91)	0.006 (0.04)	-0.222 (1.85)				
Cataluña	0.263 (2.24)	-0.275 (2.61)	0.192 (1.57)	0.122 (1.16)				
Comunidad Valenciana	0.162 (1.36)	-0.289 (2.73)	0.107 (0.88)	-0.107 (0.99)				
Extremadura	0.014 (0.09)	-0.636 (5.01)	-0.035 (0.08)	-0.444 (3.48)				
Galicia	-0.009 (0.07)	-0.390 (3.56)	-0.084 (0.65)	-0.253 (2.30)				
Madrid	0.228 (1.96)	-0.201 (1.92)	0.181 (1.54)	-0.043 (0.41)				
Murcia	0.363 (2.46)	-0.342 (2.22)	0.342 (2.31)	-0.073 (0.47)				
Navarra	0.220 (1.12)	-0.316 (2.17)	0.187 (0.96)	-0.169 (1.18)				
País Vasco	0.319 (2.54)	-0.213 (1.92)	0.283 (2.20)	-0.103 (0.93)				
Job characteristics:								
Autonomy:								
Own work			0.043 (1.18)	0.058 (2.05)			0.024 (0.62)	0.080 (2.70)
Working pace			0.011 (0.31)	0.069 (2.39)			0.034 (0.98)	0.083 (2.83)
Working hours			0.054 (1.10)	0.044 (1.17)			0.062 (1.19)	0.039 (0.99)
Degree of control:								
Quality control			0.029 (1.24)	-0.006 (0.33)			0.070 (2.98)	0.014 (0.73)
Quantity control			-0.045 (2.08)	-0.011 (0.58)			-0.038 (1.84)	-0.030 (1.53)
Directive power over employees								
Director			0.241 (2.31)	0.251 (4.36)			0.150 (1.35)	0.274 (4.62)
Supervisor			0.101 (2.13)	0.139 (3.83)			0.120 (2.34)	0.180 (4.79)
Power to make decisions:								
Raising a worker’s salary			-1.221 (1.64)	0.253 (1.84)			-1.566 (1.94)	0.300 (2.04)
Penalising a worker			0.134 (0.71)	0.062 (0.63)			0.059 (0.30)	0.029 (0.28)
Changing methods			0.026 (0.20)	0.067 (0.95)			0.076 (0.55)	0.047 (0.67)
Education mismatch:								
Education more than sufficient			0.106 (1.69)	-0.001 (0.01)			0.208 (3.37)	-0.038 (0.85)
Education sufficient			0.115 (2.01)	-0.013 (0.34)			0.172 (2.95)	-0.052 (1.29)

Correction terms:																
λ	-0.032	(1.09)	-0.007	(0.27)	-0.033	(1.09)	-0.003	(0.12)	-0.033	(1.40)	-0.098	(4.65)	-0.029	(1.18)	-0.063	(3.15)
$(\phi / F)_1$	-0.179	(2.33)	-0.058	(0.82)	-0.175	(2.24)	-0.031	(0.45)	-0.384	(5.94)	-0.338	(5.93)	-0.356	(5.56)	-0.288	(5.33)
$(\phi / F)_2$	-0.013	(0.12)	-0.034	(0.32)	-0.037	(0.34)	-0.010	(0.10)	0.081	(0.77)	-0.253	(2.47)	0.078	(0.74)	-0.182	(1.87)
$(\phi / F)_3$	-0.242	(1.60)	0.433	(1.39)	-0.242	(1.61)	0.531	(1.77)	-0.080	(0.52)	0.659	(2.00)	-0.110	(0.72)	0.749	(2.40)
$(\phi / F)_4$	0.213	(2.21)	0.089	(0.31)	0.207	(2.15)	0.046	(0.17)	0.302	(3.17)	0.613	(2.05)	0.265	(2.80)	0.475	(1.67)
$(\phi / F)_5$	-0.135	(1.01)	0.027	(0.40)	-0.140	(1.02)	0.049	(0.76)	0.086	(0.65)	0.245	(4.46)	-0.028	(0.21)	0.217	(4.17)
Observations	651		968		651		968		651		968		651		968	
Adjusted R-squared	0.5247		0.4830		0.5387		0.5269		0.4142		0.3681		0.4402		0.4405	

Omitted dummies are: illiterate, domestic services, La Rioja and occupation 4.

Table 3: Wage gap decomposition (%)

	Discrimination	**Differences in characteristics**
Model 1	14.15	-0.61
Model 2	11.47	2.07
Model 3	14.81	-1.27
Model 4	10.95	2.59

Table A.1: Descriptive statistics for wage earners

	Females		Males	
Variable	*Mean*	*S.D.*	*Mean*	*S.D.*
Hourly wage (log)	6.266	0.517	6.401	0.505
Hourly wage (pesetas)	609.4	409.5	696.6	501.6
Indefinite contract	0.665	0.472	0.781	0.414
Age	33.48	10.25	38.50	11.63
Seniority	7.67	7.30	11.01	10.18
Education				
Illiterate	0.042	0.201	0.075	0.263
Primary school	0.496	0.500	0.515	0.500
Secondary school	0.109	0.312	0.109	0.312
Vocational training	0.088	0.284	0.097	0.296
University degree (interm.)	0.173	0.378	0.076	0.265
University degree (higher)	0.079	0.271	0.077	0.266
Industrial sector				
Agriculture	0.003	0.055	0.034	0.182
Basic industry	0.066	0.249	0.106	0.308
Heavy industry	0.039	0.194	0.115	0.319
Light industry	0.121	0.326	0.146	0.354
Building	0.021	0.144	0.103	0.305
Retailing and Hotels	0.192	0.394	0.117	0.322
Transport	0.027	0.162	0.069	0.254
Economics services	0.060	0.238	0.096	0.294
Public Administration	0.194	0.396	0.115	0.319
Reproduction	0.204	0.403	0.079	0.270
Social services	0.031	0.174	0.014	0.117
Domestic services	0.026	0.160	0.004	0.064
Region				
Andalucía	0.092	0.289	0.130	0.337
Aragón	0.045	0.207	0.043	0.204
Asturias	0.028	0.165	0.028	0.165
Baleares	0.028	0.165	0.016	0.126
Canarias	0.026	0.159	0.025	0.158
Cantabria	0.013	0.111	0.017	0.129
Castilla-La Mancha	0.022	0.147	0.020	0.139
Castilla-León	0.030	0.171	0.037	0.188
Cataluña	0.165	0.372	0.148	0.355
Comunidad Valenciana	0.110	0.313	0.133	0.340
Extremadura	0.024	0.153	0.025	0.155
Galicia	0.057	0.232	0.079	0.269
Madrid	0.213	0.410	0.175	0.380
Murcia	0.025	0.156	0.011	0.116
Navarra	0.008	0.091	0.013	0.114
País Vasco	0.093	0.291	0.085	0.280
La Rioja	0.012	0.110	0.009	0.196

Occupation				
Occupation1	0.312	0.464	0.185	0.389
Occupation2	0.273	0.446	0.182	0.386
Occupation3	0.102	0.302	0.056	0.230
Occupation4	0.173	0.378	0.072	0.259
Occupation5	0.141	0.348	0.505	0.500
Job characteristics				
Autonomy				
Own work	0.447	0.498	0.467	0.499
Working pace	0.352	0.479	0.341	0.474
Working hours	0.123	0.329	0.176	0.381
Degree of control				
Quality control	2.071	0.947	2.261	0.929
Quantity control	2.108	0.980	2.121	0.938
Directive power over employees				
Director	0.022	0.148	0.075	0.264
Supervisor	0.115	0.319	0.157	0.364
Power to make decisions				
Raising a worker's salary	0.001	0.019	0.012	0.120
Penalising a worker	0.007	0.083	0.024	0.154
Changing methods	0.013	0.114	0.035	0.183
Education mismatch				
More than sufficient	0.332	0.471	0.274	0.446
Sufficient	0.586	0.493	0.610	0.488
Gross wage control	0.188	0.391	0.220	0.415
Observations	651		968	

Table A.2: Probit estimates of the participation equation

	Females		Males	
Variables	*Coef.*	*z-stat.*	*Coef.*	*z-stat.*
Age	0.026	(1.08)	0.132	(5.14)
Age squared (/100)	-0.034	(1.14)	-0.166	(5.43)
Married	-0.228	(1.77)	0.004	(0.02)
Education of partner:				
Primary education	-0.108	(0.89)	-0.019	(0.14)
Secondary education	0.354	(1.92)	-0.050	(0.24)
Vocational training	0.068	(0.34)	0.029	(0.09)
University diploma	0.565	(3.22)	0.524	(2.26)
University degree	0.428	(2.63)	0.524	(2.26)
Education:				
Primary school	0.336	(2.49)	0.229	(1.55)
Secondary school	0.778	(4.52)	0.327	(1.79)
Vocational training	0.530	(3.03)	0.462	(2.50)
University degree (intermediate)	0.899	(5.94)	0.173	(0.99)
University degree (higher)	0.643	(3.88)	0.274	(1.77)
Household variables:				
Head of household	0.503	(4.55)	0.456	(2.96)
Number of children	-0.092	(2.46)	0.056	(1.27)
Number of dependent adults	0.110	(2.92)	0.176	(4.74)
Number of income earners	-0.289	(5.16)	-0.596	(11.80)
Region:				
Andalucia	-1.110	(2.21)	-0.662	(1.07)
Aragón	-0.865	(1.65)	-0.041	(0.06)
Asturias	-0.909	(1.79)	-0.175	(0.26)
Baleares	-0.248	(0.43)	-0.173	(0.25)
Canarias	-1.167	(2.18)	-0.775	(1.20)
Cantabria	-0.515	(0.80)	-0.024	(0.03)
Castilla-La Mancha	-1.205	(2.29)	-0.274	(0.42)
Castilla-León	-1.365	(2.66)	-0.309	(0.49)
Cataluña	-0.605	(1.21)	0.058	(0.09)
Comunidad Valenciana	-0.773	(1.53)	-0.354	(0.57)
Extremadura	-1.121	(2.06)	-0.501	(0.77)
Galicia	-1.020	(1.98)	-0.240	(0.38)
Madrid	-0.751	(1.51)	-0.139	(0.22)
Murcia	-0.715	(1.33)	-0.869	(1.32)
Navarra	-0.872	(1.48)	0.0222	(0.03)
País Vasco	-0.922	(1.80)	-0.533	(0.84)
Constant	0.278	(0.42)	-1.510	(1.91)
Observations	1701		1986	
Log L	-994.765		-708.037	

Omitted dummies: Not married, illiterate, illiterate partner, not being head of the household and La Rioja.

Table A.3: Multinomial logit for occupational choice (females)

Absolute value of z-statistics in parenthesis

Variables	$Ln\left[\frac{\Pr.(Oc1)}{\Pr.(Oc.5)}\right]$	$Ln\left[\frac{\Pr.(Oc2)}{\Pr.(Oc.5)}\right]$	$Ln\left[\frac{\Pr.(Oc3)}{\Pr.(Oc.5)}\right]$	$Ln\left[\frac{\Pr.(Oc4)}{\Pr.(Oc.5)}\right]$
Age	0.078	-0.010	-0.020	0.011
	(3.72)	(0.36)	(0.76)	(0.54)
Years of education	0.667	0.416	0.146	-0.028
	(6.72)	(4.64)	(1.38)	(0.33)
Father's years of education	0.123	0.107	0.076	-0.032
	(2.13)	(1.95)	(1.18)	(0.52)
Married	0.276	0.337	-0.412	-0.522
	(0.54)	(0.71)	(0.75)	(1.19)
Constant	-9.929	-3.892	-1.160	0.423
	(7.88)	(3.65)	(1.01)	(0.39)
Observations		651		
Log L		-787.688		

Table A.4: Multinomial logit for occupational choice (males)
Absolute value of z-statistics in parenthesis

Variables	$Ln\left[\frac{\Pr.(Oc1)}{\Pr.(Oc.5)}\right]$	$Ln\left[\frac{\Pr.(Oc2)}{\Pr.(Oc.5)}\right]$	$Ln\left[\frac{\Pr.(Oc3)}{\Pr.(Oc.5)}\right]$	$Ln\left[\frac{\Pr.(Oc4)}{\Pr.(Oc.5)}\right]$
Age	0.059	0.033	0.001	-0.016
	(2.99)	(2.44)	(0.03)	(0.86)
Years of education	0.593	0.338	0.163	0.050
	(11.38)	(8.04)	(3.17)	(1.14)
Father's years of education	0.145	0.096	0.084	-0.002
	(3.77)	(3.25)	(2.30)	(0.04)
Married	0.172	0.332	0.032	0.396
	(0.44)	(1.10)	(0.08)	(0.82)
Constant	-10.743	-6.166	-4.014	-2.002
	(10.61)	(8.61)	(5.23)	(2.73)
Observations		968		
Log L		-1006.882		