1. Gender-wage gap distribution with endogenous human capital: the Spanish case

Mª Lucía Navarro Gómez, *Universidad de Málaga*
Mario F. Rueda Narváez, *Universidad de Málaga*

Gender-wage gap distribution with endogenous human capital: the Spanish case[*]

Mª Lucía Navarro Gómez, *Universidad de Málaga*
Mario F. Rueda Narváez, *Universidad de Málaga*

Abstract
This paper aims to verify the existence of gender-based wage discrimination in Spain and, if so, to quantify it. To achieve this, we first estimate earnings equations for men and women using the instrumental variable method proposed by Hausman and Taylor (1981). This aims to avoid biases resulting from endogeneity of regressors with respect to the random perturbation in the model. Building on these results, we then follow the proposal of Jenkins (1994) and estimate a bivariate wage distribution for women, containing individual expected earnings when discrimination exists (applying the female wage structure) and when it doesn't (applying the male wage structure). This allows for a full distributional analysis of gender-wage gaps. Our results show that discrimination is distributed unevenly across female workers and that the degree to which women are discriminated against grows as they move upward in the wage distribution.

Keywords: Human Capital, Wage Differentials, Discrimination
JEL Codes: J24, J31, J71

1. Introduction

The main theoretical challenge posed by research on wage discrimination is to explain the sizeable observed differences between male and female average earnings (the gender gap), whether caused by gender-based differences in characteristics and/or preferences or by the existence of discrimination, which has to be modelled somehow. The latter explanation involves answering the question quoted by Cain (1986): 'Under what conditions is it possible for essentially equal goods to have different prices when exchanged in competitive markets?'—that is, why and how can men and women receive different wages for equally productive work? Theoretically, such a situation should not even arise. If any woman produces as much as a man, but receives lower wages, any employer would take advantage of the situation in order to have a cheaper workforce, thus obtaining extra benefits. One would expect many firms to do the same thing, so that in the long run female wages would rise until they caught up with male wages. However, empirical research fails to explain the entire observed wage gap in terms of differences in characteristics. This leaves the remainder of the gap consistent with the presence of discrimination in the labour market. Much theoretical work has been devoted to reconciling these two seemingly contradicting positions (for an overview, see Altonji and Blank, 1999).

From the point of view of empirical research, the focus is usually on verifying and quantifying wage discrimination. This leads methodologically to application of the human capital theory (Becker, 1964) and estimating some kind of Mincerian wage equation (Mincer, 1974) separately for men and for women. While the female parameters are thought to represent what women receive when being discriminated against, the male parameters are used to estimate the counterfactual wage a female worker would get if she were a man, but otherwise equal. From these results, the mean difference in (log) wages is usually decomposed into one part being explained

* The authors acknowledge financial assistance from the Andalusian Government under grant P09-SEJ-4859.

by differences in human capital and other characteristics (differences in productivity) and one part arising because the labour market values workers' characteristics differently depending on whether the worker is a man or a woman. This latter part (residual or unexplained) is then taken as an estimate of the mean amount of discrimination in the labour market (Oaxaca, 1973; Blinder, 1973).

In Spain, empirical research on discrimination has been relatively frequent in recent years due to the availability of new databases with suitable information on wages and other attributes of workers. Interest has been devoted to two lines of work. On the one hand, articles such as De la Rica and Ugidos (1995) use cross-sectional data to estimate the average amount of discrimination, applying to the female wage equation a variation of Heckman's (1979) model for data with sample selection problems. On the other hand, García et al. (2001) is the first in a series of papers to use quantile regression to assess the degree of discrimination at different points of the wage distribution. Evidence on discrimination in the Spanish market tends to conclude that differences in characteristics have a limited to no role in explaining observed (raw) wage differentials. This means that most of the raw differential is due to differences in the coefficients of earnings equations, resulting in residual, net, or discriminatory wage differentials[287].

Our objective in this paper is to assess the amount of wage discrimination experienced by Spanish female wage earners, using individual data from the first seven waves of the Spanish section of the European Community Household Panel (ECHP, INE, 1994–2000). To achieve this, we broaden the traditional approach outlined above in two ways. First, when estimating wages, we use the instrumental variable (IV) method for panel data developed by Hausman and Taylor (1981), HT thereafter. This is done in order to account for the possible endogeneity of several wage determinants, which is important because discrimination analysis are based on a previous estimate of what a woman would earn if she were a man but otherwise had the same characteristics. This, in turn, is compared to the amount she receives as a woman. If those estimates are biased, the discriminating and non-discriminating wages based on those equations would be biased as well. As a result, the difference between these two cannot be used to construct a reliable estimation of wage discrimination. Second, our analysis of discrimination focuses on more than the average wage differential. Following the methodology proposed by Jenkins (1994), it takes into account the entire distribution of discrimination. This is a logical step, since the original concept of discrimination arises from the comparison to the wage a woman receives with her own productivity—that is, any woman is discriminated against if she receives a wage that is lower than her productivity. This implies that the degree to which a woman is discriminated against is just another individual variable, and there is nothing that makes us expect it to be constant over the sample.

The rest of this paper is organised as follows. The next section discusses the econometric framework used to estimate wages and then compute two wage distributions for the women in the sample, one discriminating and the other non-discriminating. The data used to achieve this are presented in section 3, along with the variables used in the analysis. Section 4 presents the results, and the last section provides the main conclusions.

2. Econometric framework

In order to obtain the expected wage distributions for women when they are subject to discrimination and when they are not, we first estimate the parameters that determine such wages. We then use the male set of parameters to compute non-discriminating wages, thus

287 For an overview of empirical evidence on discrimination for Spain, see Rueda (2010).

assuming that, in the absence of discrimination, women would be treated as men. The female wage structure generates the discriminating wage distribution. Obtaining these estimates, in econometric terms, requires estimating the following model for men and for women:

$$y_{it} = \mathbf{x}'_{it}\beta + \mathbf{z}'_i\gamma + \eta_i + \varepsilon_{it} \qquad (i = 1,...N;\ t = 1...T_i), \qquad [1]$$

where y_{it} is the log wage of worker i in year t, x_{it}, and zi are vectors containing k and g (g=1) variables, respectively. While variables in xit are time-varying, zi only includes the number of years of completed schooling and thus is constant within each individual. Finally, η_i and ε_{it} are both independent random disturbances distributed with zero mean and variances O_r^2 and σ_ε^2. Finally, the fact that each individual is observed Ti times implies an unbalanced panel data set in which each worker is not present in every year of the survey, as is the case in the ECHP[288].

Concerning this model, the endogeneity of education and maybe other wage determinants is modelled as a correlation between those variables and the effect of unobservable characteristics, included in ηi. Literature on wage determination has devoted most attention to the possible endogeneity of education (see, for example, Griliches, 1977 and Card, 2000). However, the reasons for considering education as endogenous can usually be applied to other variables. It is possible, for example, that past unemployment, one of the other variables, is correlated with the unobserved factors in ηi, because these probably include the ability (or lack of ability) to find a good job, something that ought to be separated from the depreciation in the individual stock of human capital that the variable is intended to measure.

In the absence of such correlations for all covariates, efficient parameters estimates can be obtained using the random-effects (RE) procedure. If endogeneity is to be taken into account, however, the usual approach is to transform the data into deviations from individual means in order to obtain fixed-effects (FE) estimates. Yet, since our model includes a time-invariant regressor (education), this would yield no estimate for the effect of education on waves. In this kind of situation, the HT instrumental variable procedure takes an intermediate approach and allows estimating the full set of parameters (including those corresponding to time-constant variables) by assuming that only some of the variables are endogenous, while also providing a gain in efficiency, since some of the variables are treated as exogenous.

Once the wage equations have been estimated, the usual way to quantify the amount of gender-based discrimination is to decompose the mean difference in log wages as:

$$\bar{y}^m - \bar{y}^f = (\bar{x}^m - \bar{x}^f)\hat{\beta}^m + (\bar{z}^m - \bar{z}^f)\hat{\gamma}^m + \bar{x}^f(\hat{\beta}^m - \hat{\beta}^f) + \bar{z}^f(\hat{\gamma}^m - \hat{\gamma}^f)\,, \qquad [2]$$

where $\bar{y}$ is the mean of the dependent variable, $\bar{x}$ and $\bar{z}$ are vectors containing the means of the regressors, $\hat{\beta}$ and $\hat{\gamma}$ are the coefficients previously obtained, and the superscripts m and f represent the male and female samples, respectively. In this decomposition, originally proposed by Oaxaca (1973) and Blinder (1973), the first two terms on the right-hand side of the equation amount to the part of the difference in observed log wages attributable to differences in mean characteristics. The last two terms, in turn, represent the wage loss by women because their characteristics are paid according to the female parameters instead of the male ones. This latter part is viewed as an estimate of the mean amount of discrimination in the labour market.
This methodology is attractive in its simplicity because it summarizes the information available in a single figure. Discrimination, however, can be considered to affect women differently across the labour force, and thus it seems clear that the same average amount can arise from very

288 For the sake of simplicity, $NT = \sum_i T_i$ will denote the total number of person-year observations in the panel.

different realities. Jenkins (1994) suggests using individual information on estimated wages in order to conduct a distributional analysis of discrimination. To do this, we compute the following two variables for each female observation in the sample:

$$\hat{y}_{it} = \exp\left(x'_{it}\hat{\beta}^{f} + z'_{i}\hat{\gamma}^{f}\right), \text{ and}$$
$$\hat{r}_{it} = \exp\left(x'_{it}\hat{\beta}^{m} + z'_{i}\hat{\gamma}^{m}\right). \qquad [3]$$

These are, respectively, the predicted (median) wage for each woman ($\hat{y}_{it}$) and the predicted amount she would earn if there were no discrimination ($\hat{r}_{it}$): the reference wage. These predictions are in monetary units (1992 pesetas) and not in logarithms[289].

Jenkins proposes several ways to summarize the information contained in the joint distribution of $\hat{y}_{it}$ and $\hat{r}_{it}$. These two wages can be compared in order to obtain an absolute ($\hat{s}_{it} = \hat{r}_{it} - \hat{y}_{it}$) or a relative ($\hat{d}_{it} = 100\left[\hat{s}_{it}/\hat{y}_{it}\right]$) estimate of the amount of discrimination faced by each female wage earner. While the former is measured in monetary units, the relative wage gap is the wage bonus a woman should receive if she were to be paid according to the male wage structure (and not the proportional wage loss arising from being paid as a woman[290]).

3. Data

We use information extracted from the first seven waves of the Spanish section of the ECHP (INE, 1994–2000). This survey provides information on individual wages, completed education, and the other variables usually included in the specification of earnings equations. While preparing the data, we had to eliminate several observations due to lack of response or abnormal values on any of the variables used and restricted the analysis to wage earners working at least 15 hours per week[291]. Added to the loss of individuals because of panel attrition, this yielded complete information on 5 369 men and 3 264 women who appear at least once in the panel. The total number of person-year observations is 19 291 and 10 349, respectively. This means that each man is observed for an average of 3.6 years and each woman for 3.2 (over a maximum of 7). Several factors affect how many times any given person appears in our data source, but a lower figure for women could be expected, based on higher labour market intermittency for women.

The explained variable is the natural logarithm of the monthly net wage (measured in constant 1992 pesetas). As for wage determinants, these can be grouped into human capital information, job and personal characteristics, and a variable with the regional unemployment rate. Human capital variables include, as stated, the number of years of completed education, labour market experience (measured as years passed since the first job was started)[292], and two dummies for workers who received training the previous year, whether employer financed or not. With regard to job information, we include dummies for public sector and part-time employment.

289 Since wages are assumed to be log-normal, sometimes (see Hansen and Wahlber, 2005), the residual variance is included in the prediction formula in order to obtain expected (mean) wages. If such an adjustment is not made, medians are estimated instead, which we consider preferable as a way to study discrimination, given the asymmetry in wages.

290 This differential can be computed as $\hat{d}_{it}/\left(1+\hat{d}_{it}\right)$ and can also be viewed as the relative wage loss a man would experience if paid as a woman.

291 Otherwise, the ECHP does not report information on other variables such as public-private employment or job seniority.

292 Computing experience this way is problematic, since the variable also includes time spent outside the market. The survey does not provide information on such interruptions. As a correction, we include a binary indicator for workers who were inactive or unemployed most of the previous year.

The latter is intended to capture the influence of labour supply in wages, which is known to be lower for women[293]. Also, job tenure is included as the number of years with the employer up to 10 years and as a dummy for longer jobs. This is done because the information available in the ECHP is truncated at that point, the duration of the current job being unknown when longer than 10 years. Tenure, while a job characteristic, can also be viewed as an indicator of the stock of specific human capital accumulated by the wage earner. As for personal information, we include a variable for those living with a partner (married or not) and dummies for three categories of self-reported health (good, bad, or very bad). Finally, the regional unemployment rate is computed as the weighted average of male and female rates in the autonomous communities that make up each of the seven regions the ECHP considers (NUTS 1). This variable tries to capture the effect of local labour market conditions and regional variations in wages. Descriptive statistics on these variables are included in the annex (Table A1).

4. Results

The estimation results corresponding to male and female earnings equations (RE and HT models) appear in the annex (Table A2) and allow us to analyze 'discriminating' wage differentials—that is, the difference in wages predicted using the two different estimated wage structures. With these two sets of parameters, we predict two wages for each woman in the sample, thus obtaining a bivariate distribution with wages expected with ($\hat{y}_{it}$) and without ($\hat{r}_{it}$) discrimination. As stated in the second section, this bivariate distribution allows us to obtain an absolute ($\hat{s}_{it}$) and a relative ($\hat{d}_{it}$) estimate of the amount of discrimination faced by each female wage earner.

Table 1 shows some statistics for the distribution of absolute and relative wage differentials, computed using the results obtained via the RE and HT estimators. We compare these two model because the only change in the assumptions they make is whether none (some) of the regressors included is (are) correlated with the individual effect included in the compound error term. That is, we compare results under the usual assumptions of exogeneity against a model in which several wage determinants are correlated to individual unobserved characteristics. As a result, differences in the results can be attributed to the use of a procedure that accounts for such correlation. With regard to this, our empirical results when comparing the RE and FE estimations clearly reject the exogeneity of all covariates, something that is assumed in the RE model.

Most prominent in Table 1 is the fact that estimated discrimination seems to be much larger when comparing men and women with the HT set of parameters versus the RE set, regardless of the statistic or variable used. Thus average absolute discrimination is estimated at around PTA 25.000 with the RE model, rising up to around PTA 43.000 with the HT model. In relative terms, the discriminating differential rises from 26% to 35%. The fact that median estimates are lower than mean ones implies that the distribution of discrimination is, as that of wages, positively skewed for both models. As a reference, our data show a mean wage differential of about PTA 29.000 in absolute terms. From the viewpoint of decomposition analysis, this suggests that the RE model explains only a small fraction of the raw gap (around 14%) as a result of differences in characteristics, whereas the HT model leads to a discriminating average that is greater than the original. This suggests that the mean set of female characteristics works to the advantage of female workers, because if they had the same levels of education, work experience, and so forth as men,

293 The ECHP includes information on the weekly number of hours worked, something that is arguably useful when analyzing gender-wage differentials, given that these are affected by labour supply. However, the information includes a large proportion of anomalous values, and it is not possible to know how many of those hours are due to overtime and thus paid at a different wage rate. Thus, we do not use it.

their position with respect to their male counterparts would be even worse. Taking into account the different assumptions made in the two models, the fact that estimated discrimination is higher when accounting for endogeneity suggests that the effect of unobservable characteristics, which is reflected as a bias in the RE coefficients, tends to have a positive impact on female wages relative to male wages. As a result, once the unobservables are accounted for, the resulting (unbiased) coefficients reflect a greater amount of discrimination against women.

Table 1. Descriptive statistics for discriminating wage differentials

	RE	HT
Absolute differentials (1992 *pts*)		
Mean	25 249.98	42 645.49
Median	23 700.68	31 741.09
Minimum	-1 517.26	2 371.72
Maximum	87 770.05	217 523.60
Standard deviation	12 586.54	32 091.73
Relative differentials (%)		
Mean	26.10	35.01
Q1	17.68	29.19
Median	24.57	34.46
Q3	34.01	40.17
Minimum	-1.48	13.36
Maximum	63.75	71.42
Standard deviation	11.77	7.90
% with *d* under 10%	7.01	0.00
% with *d* between 10% and 30%	58.41	28.79
% with *d* over 30%	34.58	71.21

Moving the focus away from mean values, the table shows some relevant facts. First, in a few cases, the RE model leads to negative discrimination (in the sense that women would receive lower wages if they were paid as men). This only happens for eight observations (about 0.08% of the whole sample), but illustrates the possibility that, even when some average discrimination is estimated, part of the sample suffers no discrimination at all. A first look at the distribution of discrimination can be obtained from the quartiles of the relative differentials. The RE model predicts that one-quarter of the sample is discriminated against no more than 17.68% (29.19% for the HT model). For the 25% of women experiencing the most discrimination, the HT model predicts it to be 40.17% or even higher. As for standard deviations, the data show that the distribution is somewhat more homogeneous for the HT model if measured in relative terms. This suggests that higher average wage differentials in the HT model are likely associated with higher predicted wages.

In a recent paper, De la Rica, et al. (2008) analyse how discrimination affects women along the wage structure. They show that Spanish ECHP data for 1999 apparently do not support the contention that discrimination is higher among women with higher wages (the "glass ceiling" hypothesis). They find, however, that the theory is valid for women with higher education, whereas the relationship between wages and discrimination is negative for female workers with primary or secondary education. They develop a model of statistical discrimination to explain this behaviour, arguing that low employment rates for women without a university degree (especially when compared to men) induce employers to assign them a higher risk of quitting their job and, thus, a lower wage. This theory of "sticky floors" is then successfully tested against the data using a series of quantile regressions. Although there are deep methodological differences between their work and ours, especially with respect to the econometric models used to estimate a bivariate distribution of discriminating wages and their counterfactuals, it is interesting to explore the extent to which our results are similar to theirs, keeping in mind the limitations of such a comparison. Information homologous to that presented by De la Rica et al. is shown in Tables 2 and 3 (RE and HT models, respectively). These provide the mean of the estimated relative discrimination in the four segments of the wage distribution ($\hat{y}_{it}$), delimited by its quartiles for the whole sample and the subsamples of female wage earners with and without higher education[294].

Table 2. Relative differentials (means) between quartiles of the female wage distribution (university vs. no university). RE Model

Range	All	No university	University
Minimum to Q_1	29.48	30.14	17.13
Q_1 to median	29.71	29.93	17.15
Median to Q_3	25.11	30.41	18.68
Q_3 to maximum	20.09	27.35	20.16

For the RE model, the trend is similar to that described by De la Rica et al., with decreasing discrimination for female wage earners as a whole (first column in Table 2). Average discrimination peaks at 30% for the lowest-paid quarter of the sample and falls to 20% for the highest-earning quarter. The same relationship, although much weaker, is observed for workers without any university education, while university graduates face higher discrimination as their wage prospects improve (20% for those with the highest wages versus 17% for those with the lowest).

Table 3. Relative differentials (means) between quartiles of the female wage distribution (university vs. no university). HT Model

Range	All	No university	University
Minimum to Q_1	28.16	27.33	33.47
Q_1 to median	33.47	32.18	36.67
Median to Q_3	36.40	33.89	41.09
Q_3 to maximum	42.01	38.88	46.93

In sharp contrast, the HT model results show an overall similar evolution of discriminating wage differentials along the wage distribution, irrespective of the educational level of the worker. The three columns in Table 3 show how discrimination grows by about half when switching from the lowest-paid quarter of the sample to the best-paid quarter. The only noticeable difference is that discrimination is always greater for women with higher education. Two conclusions can be drawn from comparing the results of both models. First, the HT model again seems to be more in line with the idea that, as a woman moves up in the wage distribution, the more discrimination

294 The wage scale used to obtain the quartiles is that of the corresponding set of observations in each case (the whole sample or the subsample in question). Thus the means of $\hat{d}_{it}$ are not necessarily computed with the same observations for the whole sample as for the subsample of highly educated women. This means that the average for all women is not a weighted average of the average of university graduates and the rest of workers. In fact, the overall mean can be outside the range delimited by the latter two.

she faces. Second, both models predict comparable amounts of discrimination for women with primary or secondary education (a little over 30%), but the HT model predicts much worse discrimination for women who have attended university: in the HT model these women face discrimination that is about double that predicted in the RE model. It seems plausible that this change is due, at least in part, to the change in the coefficient of the schooling variable used in the wage equations (Table A2) following estimation by the HT procedure. While this coefficient is higher for women in the RE model, thus allowing for less discrimination as education (and wages) increases, the effect of education is higher for men in the HT model, thus increasing the gap between $\hat{y}_{it}$ and $\hat{r}_{it}$ for female workers with more years of completed education.

Finally, the econometric framework used in this paper also allows us to analyze more deeply the associations between wage discrimination and educational attainment. An intuitive approach is to use the information on $\hat{d}_{it}$ in order to identify those groups of female workers facing less (or more) discrimination depending on their highest level of completed education. Table 4 presents mean values of relative discrimination for those groups for both specifications (RE and HT). Additionally, the last (third) column shows the proportion of individuals in each group for the total female sample. It must be noted however, that since the discrimination estimates used to construct this table are computed using actual information on each individual for every variable, the effect of the degrees considered can be mixed with that of other correlated variables.

Table 4. Estimated discriminating wave differentials (%) for women according to educational level.

Group	RE	HT	Relative size (%)
Total	26.10	35.01	100.00
Education			
No education	42.55	32.89	1.36
Primary education	37.23	33.17	16.57
1st level secondary ed.	29.48	31.68	21.38
Vocational training 1	26.53	31.96	8.54
Vocational training 2	21.99	31.84	8.03
2nd level secondary ed.	25.04	36.44	14.07
University (short)	19.16	39.04	15.44
University (long)	17.35	40.07	14.60

Viewing the information on Table 4, it becomes clear that wage differentials behave quite differently for the two models. The RE model shows a sharp decrease in discrimination as education grows. The wage gain that women would receive if they were paid as men decreases from 37% for those with primary education to about half (17%) for those with long university degrees. The HT model shows the opposite trend, although less pronounced: the differential increases from 33% to 40% for the same two groups. In any case, both models place workers with a university degree at a separate level, either well below the average discrimination (as in the RE model) or above it (the HT model). Discrimination among wage earners with secondary education is relatively stable, irrespective of the exact level (for example, vocational or not).

5. Conclusions

Throughout this paper, we have undertaken an empirical analysis of wage discrimination against women in the Spanish labour market using data from the first seven waves of the Spanish section of the European Community Household Panel (INE, 1994–2000). We depart from the approach taken in previous research for this context in two ways, in particular concerning the estimation of wage equations and how these are then used to estimate the degree of discrimination in the market. First, we have consistently and efficiently estimated the necessary earnings equations using the procedure proposed by Hausman and Taylor (1981) for use with panel data when some of the regressors are time-constant, as education, and some are potentially endogenous. Second, and following the methodology suggested by Jenkins (1994), we have used those estimates to estimate a bivariate distribution of wages with and without discrimination, using the male and female set of parameters, respectively. This provides information on the degree of discrimination faced by each of the female wage earners in the sample and thus makes it possible to take a closer look at the subject.

The results of the analysis have been presented and compared for the RE and HT models, because both share the compound structure of the random disturbance and differ only in the assumptions made on the exogeneity/endogeneity of the wage determinants. Results are quite different according to each model. Both in absolute and relative terms, the RE model finds average discriminating differentials that are slightly narrower than the raw difference in wages between men and women, something common in the literature for the Spanish case when using comparable methodologies. This means that almost all of the observed difference in wages is due to the fact that the same characteristics (such as education and experience) are valued by the market differently for women than for men. However, the HT model's estimated mean discrimination is higher than the observed raw gender gap. This suggests that, if women and men were equally endowed with respect to wage determinants, the wage gap between both groups would be higher than that actually observed.

A closer look at the results provides some interesting additional information on the distribution of discrimination across our sample of female workers. A descriptive analysis shows that there is considerable variation in the degree of discrimination that women face, irrespective of the estimation procedure. The distribution of discriminating wage differentials across the wage scale shows that the HT model predicts higher average discrimination because those differentials are notably wide for women with higher wages, whereas for women on the left side of the wage distribution both models lead to similar results. Finally, we have analyzed the degree to which discrimination varies depending on completed education. The HT model shows that the relative position of women worsens with higher educational achievement. Thus, it seems that as wage prospects improve for women, they are also subject to higher levels of discrimination.

References

Altonji, Joseph G., and M. Blank Rebecca (1999): "Race and Gender in the Labor Market", in O. Ashenfelter and D. Card (eds.). *Handbook of Labor Economics,* Vol. 3C. Elsevier, Amsterdam, The Netherlands.

Becker, Gary S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. The University of Chicago Press, Chicago and London. 3rd edition, 1993.

Blinder, Alan S. (1973): "Wage Discrimination: Reduced Forms and Structural Estimates". *Journal*

of Human Resources, 8(4), pp. 436–455.
Cain, Glenn G. (1986): "The Economic Analysis of Labor Market Discrimination: A Survey", in O. Ashenfelter and R. Layard (eds.). *Handbook of Labor Economics*. Elsevier, Amsterdam, The Netherlands.

Card, David (2000): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems". Working Paper 7769, National Bureau of Economic Research, Cambridge, Mass.

De la Rica, Sara, Juan J. Dolado, and Vanesa Llorens (2008): "Ceilings or Floors?: Gender Wage Gaps by Education in Spain". *Journal of Population Economics*, forthcoming.

De la Rica, Sara, and Arantza Ugidos (1995): "¿Son las diferencias en capital humano determinantes de las diferencias salariales observadas entre hombres y mujeres?". *Investigaciones Económicas,* 19(3), pp. 395–414.

García, Jaume, Pedro J. Hernández, and Ángel López (2001), "How Wide Is the Gap? An Investigation of Gender Wage Differentials using Quantile Regression". *Empirical Economics*, 26(1), pp. 149–167.

Griliches, Zvi (1977): "Estimating the Returns to Schooling: Some Econometric Problems". *Econometrica*, 45(1), pp. 1–22.

Hansen, Jörgen y Roger Wahlberg (2005): "Endogenous schooling and the distribution of the gender wage gap". *Empirical Economics*, 30(1), pp. 1-22.

Hausman, Jerry A. (1978): "Specification Tests in Econometrics." Econometrica, 46(6), pp. 1251–1272.

Hausman, Jerry A., and William E. Taylor (1981): "Panel Data and Unobservable Individual Effects". *Econometrica*, 49(6), pp. 1377–1399.

Heckman, James J. (1979): "Sample Selection Bias as a Specification Error". *Econometrica* 47(1), pp. 153–162.

Jenkins, Stephen P. (1994): "Earnings Discrimination Measurement: A Distributional Approach". *Journal of Econometrics*, 61(1), pp. 81–102.

Mincer, Jacob (1974): *Schooling, Experience and Earnings.* National Bureau of Economic Research and Columbia University Press, New York.

Oaxaca, Ronald (1973): "Male-Female Wage Differentials in Urban Labour Markets". *International Economic Review,* 14(3), pp. 693–709.

Rueda, Mario F. (2010): *Discriminación salarial por razón de género y capital humano: Un análisis con datos de panel.* Civitas, Pamplona.

Appendix

Table A1. Descriptive statistics for variables used in the estimations

	All		Men		Women	
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Log wages[a]	11.6437	0.5125	11.7299	0.4797	11.4830	0.5327
Human capital						
Education (years)	9.6602	3.9895	9.1413	3.8756	10.6275	4.0186
Experience	18.6299	12.4882	20.2299	12.7463	15.6475	11.4089
Experience sq.	503.02	565.16	571.71	601.02	374.99	465.05
Labour market training						
Firm-financed	0.0929	0.2904	0.0876	0.2827	0.1029	0.3039
Self-financed	0.0378	0.1908	0.0301	0.1709	0.0522	0.2224
No work previous year	0.1284	0.3345	0.1093	0.3120	0.1641	0.3704
Job characteristics						
Public sector	0.2567	0.4368	0.2262	0.4184	0.3138	0.4640
Part-time	0.0600	0.2375	0.0217	0.1458	0.1314	0.3379
Tenure < 10 years						
Tenure	1.7221	2.4858	1.6334	2.4519	1.8874	2.5398
Tenure sq.	9.1448	18.7573	8.6794	18.4647	10.0121	19.2617
Tenure ≥ 10 years	0.3731	0.4836	0.4033	0.4906	0.3167	0.4652
Personal characteristics						
Good health	0.2318	0.4220	0.2291	0.4202	0.2368	0.4252
Bad health	0.1632	0.3695	0.1633	0.3697	0.1629	0.3693
Married / with partner	0.6563	0.4749	0.7009	0.4579	0.5733	0.4946
Unemployment rate	19.5742	5.6060	19.7893	5.6248	19.1732	5.5489
N	8 633		5 369		3 264	
NT	29 640		19 291		10 349	

a. Natural log of net monthly wages in constant 1992 pesetas.

Table A2. Estimation results from earnings equationsa

Variable	Men						Women					
	RE			HT[b]			RE			HT[b]		
	Coef.		Std. Err.	Coef.		Std. Err.	Coef.		Std. Err.	Coef.		Std. Err.
Education (years)	0.0604	***	0.0012	0.1296	***	0.0070	0.0648	***	0.0015	0.1194	***	0.0134
Experience	0.0276	***	0.0010	0.0422	***	0.0017	0.0202	***	0.0013	0.0395	***	0.0024
Experience sq.	-0.0004	***	0.0000	-0.0005	***	0.0000	-0.0003	***	0.0000	-0.0005	***	0.0000
Public sector	0.0089		0.0075	-0.0075		0.0095	0.1204	***	0.0093	0.0458	***	0.0103
Part time	-0.3861	***	0.0129	-0.3369	***	0.0142	-0.3873	***	0.0086	-0.3159	***	0.0084
No work previous year	-0.1009	***	0.0067	-0.0835	***	0.0071	-0.0960	***	0.0083	-0.0862	***	0.0077
Labour market training												
Firm-financed	0.0300	***	0.0060	0.0161	***	0.0061	0.0190	**	0.0075	0.0095		0.0066
Self-financed	-0.0022		0.0095	-0.0006		0.0097	0.0030		0.0099	0.0080		0.0087
Tenure												
Tenure	0.0192	***	0.0032	0.0145	***	0.0033	0.0266	***	0.0044	0.0206	***	0.0040
Tenure sq.	-0.0011	***	0.0004	-0.0014	***	0.0004	-0.0013	***	0.0005	-0.0021	***	0.0004
Over 10 years (dummy)	0.1261	***	0.0090	0.0262	**	0.0108	0.1750	***	0.0132	0.0145		0.0145
Health												
Very good	0.0038		0.0040	0.0037		0.0040	0.0048		0.0055	0.0004		0.0048
Normal or bad	-0.0120	**	0.0049	-0.0061		0.0050	-0.0186	***	0.0068	-0.0179	***	0.0061
Unemployment rate	-0.0096	***	0.0004	-0.0064	***	0.0005	-0.0079	***	0.0006	-0.0048	***	0.0009
Married or with a partner	0.0855	***	0.0078	0.0400	***	0.0092	0.0278	***	0.0084	0.0122		0.0094
Intercept	10.8839	***	0.0192	10.0630	***	0.0735	10.6445	***	0.0250	9.8925	***	0.1496
N	5 369			5 369			3 264			3 264		
NT	19 291			19 291			10 349			10 349		
R^2	0.5277			0.4625			0.6349			0.5505		
σ_u	0.27872			0.38638			0.2721			0.9397		
σ_e	0.18521			0.18516			0.1803			0.1802		
ρ	0.69369			0.81324			0.6949			0.9645		
$\chi 2^c$	475.09			5.57			553.48			3.27		
(degrees of freedom)	14			6			14			6		

a: The reference is a worker with no labour market training, single, in good health, and working full time in the private sector.

b: Although it is possible to compute the R2, this is not bound between 0 and 1 for IV models and is thus not shown.

c: Test statistic for the Hausman test (1978). For RE estimates, it is a test of exogeneity of all covariates. Degrees of freedom of the statistic under the null of no endogeneity appear on the row below.

***: 1% significant, **: 5%, *: 10%.