

ENDOGENOUS WAGE DETERMINANTS AND RETURNS TO EDUCATION IN SPAIN

Antonio Caparrós Ruiz

Departamento de Estadística y Econometría
Universidad de Málaga
e-mail: antonio@uma.es

M^a Lucía Navarro Gómez

Departamento de Estadística y Econometría
Universidad de Málaga
e-mail: l_navarro@uma.es

Mario F. Rueda Narváez

Departamento de Estadística y Econometría
Universidad de Málaga
e-mail: mfrueda@uma.es

Abstract

One of the major contributions of human capital theory and subsequent empirical work has been to prove the important role of years of schooling as a determinant of wages. However, the exact value of such effect, the private return to schooling, and how it should be estimated remain a source of both theoretical and empirical discussion. Some of the open questions refer to which variables should be included as regressors in the wage equations and by which method they can be consistently estimated. In this paper, we add some empirical evidence for the Spanish labour market. Using the instrumental variable –IV– approach proposed by Hausman and Taylor (1981) –HT– and data from the 1994-1997 Spanish section of the European Community Household Panel (ECHP), we contribute to assess the direction and amount of the bias that affects ordinary least squares (OLS) estimation. The HT procedure allows us to take into account the possible endogeneity of education as well as other determinants of wages, while making it unnecessary to use instrumental variables excluded from the earnings equations. These equations are estimated on two incomplete panels, corresponding to male and female wage earners.

Keywords: Wage level, human capital, panel data, instrumental variable.

Area temática: Economía y Empresa

1. Introduction

In the last decades, a vast body of research has focused on individual educational attainment viewed as the result of human capital investment decisions (Becker 1964; Mincer 1958; Shultz 1961). The economic implications of such theories are related to a wide range of topics: Education has been considered as an important factor in determining economic growth (Barro, 1991), income distribution (Psacharopoulos 1976; Ram 1989) and social mobility (Boudon 1974). The extent to which human capital affects these matters will depend on the behaviour of education investors/consumers, that is, students and their families. More generally, society as a whole will be affected by the optimal resource allocation in the educational system and the labour market. From this point of view, it seems clear that information and analysis on the returns to schooling and their evolution are crucial to social and individual decision makers.

Starting with the seminal work of Mincer (1974), the empirical approach used to quantify the private return to schooling investments has focused on the estimation of earnings functions, where the education of each individual is included among other variables in order to explain some measure of individual logarithmic wages¹ (overviews can be found in Ashenfelter et al. 1999 and Psacharopoulos 1984, 1994). The empirical literature has devoted much discussion to how the rate of return to schooling should be properly estimated. Some of the econometric problems that arise in the field of the earnings functions are those related to sample selection bias (Heckman 1979; Trost and Lee 1984), unobserved -and possibly unobservable- ability (Griliches 1977; Ashenfelter and Krueger 1994; Card 1999), endogenous schooling (among others, Card 2000), and measurement errors in the educational variable (Griliches 1977). In the presence of any of these problems, it has been argued that ordinary least squares (OLS) estimates of the effect of education on wages are biased and inconsistent because, except for measurement errors², they would imply a nonzero asymptotic correlation between the residual and the regressors.

¹ In Mincer's original equation, the theoretical discussion leads to the sole inclusion of education and a quadratic form of work experience as explanatory variables. Later empirical research usually adds additional information on individual and job characteristics and also different functional forms of the experience-earnings profile (see Murphy and Welch, 1990).

One of the alternatives to avoid OLS inconsistency, proposed initially in the context of cross section data, is the use of instrumental variables (IV) procedures, where education is treated as an endogenous regressor, in the econometric sense of being correlated with the residual. Suggested instruments for the years of schooling include family background variables (parents' education and income, as in Hausman and Taylor 1981), supply-side features of the educational system as geographic proximity to a college or changes in compulsory education laws (Card 1993, Harmon and Walker 1995) or information arising from natural experiments, such as the year or season of birth (Angrist and Krueger 1991).

IV estimation, however, has its own difficulties and shortcomings. Instruments must be correlated with the endogenous variables but not with the random error (rank condition), legitimately excluded from the wage equation (exclusion restriction) and be more numerous than the endogenous regressors (order condition). The IV estimator proposed by Hausman and Taylor (1981) –hereafter HT–, allows to circumvent some of these issues if panel data are available. In particular, the HT estimator exploits the panel nature of the data so that no additional information is needed apart from that included in the wage equation in order to obtain a consistent and efficient estimate of the return to schooling. Instead, sufficient assumptions have to be made about which variables are correlated with the individual-specific residual (the fixed/random effect) and which are not, so that the exogenous regressors provide at least as many instruments as endogenous variables in the model. Besides, this technique could also overcome the problem of sample selection bias if it only operates through the individual effect (Kunze 2000).

In this paper we use the HT procedure to consistently estimate the private return to education for male and female wage earners in the Spanish labour market. Our data, suitable for this kind of analysis, has been drawn from the first four waves (1994-1997) of the Spanish section of the European Community Household Panel (PHOGUE). Not until recently, panel data on this subject was practically nonexistent, and thus previous research on the Spanish case has been limited to the use of cross section data. This has made it difficult to address the problems discussed above in a general way, constraining empirical research to partial solutions. Consistent results in the presence of sample

² This kind of problem would cause a downward bias in the estimate instead.

selection bias can be found in Alba-Ramírez and San Segundo (1995), Lassibille (1998) and San Segundo and Valiente (2003)³. On the other hand, both Pons and Gonzalo (2002) and Barceinas et al. (2001)⁴ present IV estimates with cross-section data.

With our main objective in mind, the remainder of the paper is organized as follows: next section presents the econometrical framework, third and fourth sections present the data and empirical results, respectively, and the main conclusions are summarized in the last section. Besides, an annex is included with results from alternative specifications referred to in the fourth part of the paper.

2. Econometric Model

In order to estimate wage equations in their classical form when panel data are available and each cross section unit is observed in several moments of time, some considerations must be made first. These are particular problems that cast doubt on the relevance and accuracy of standard methodologies for static linear (or, as in our case, semi-logarithmic) panel data models. More precisely, we are interested in estimating the following for men and women separately:

$$y_{it} = x'_{it}\beta + z'_i\gamma + \alpha_i + \varepsilon_{it} \quad [1]$$

where y_{it} is the logarithm of monthly wages for individual i in year t , x_{it} and z_i are vectors containing k time-varying and g time-invariant variables that affect wages as determined by the coefficients in β and γ and α_i and ε_{it} are two error terms, independent of each other, with distributions $N(0, \sigma_\alpha^2)$ and $N(0, \sigma_\varepsilon^2)$.

The notation in [1] presents a standard panel data model which explicitly recognizes that the regressors in z_i are constant over time for any given individual, though they can vary from one to another. On the other hand, variables in x_{it} vary both

³ The data used in these studies were extracted from the Encuesta Piloto de Ganancias (INE, 1990), Encuesta de Presupuestos Familiares (INE, 1991) and Encuesta de Estructura, Conciencia y Biografía de Clase (INE, 1991).

⁴ Statistical sources for the former are the Encuesta de Estructura, Conciencia y Biografía de Clase (INE, 1991) and PHOGUE's first wave (INE, 1994), which is used in the latter along with the Encuesta de Presupuestos Familiares (INE, 1991).

within and between workers. The same distinction applies to the error terms, with α_i being the same every time any individual is observed and ε_{it} being time-variant. In matrix notation, if data are ordered by individual, the model is

$$y = X\beta + Z\gamma + \alpha + \varepsilon \quad [2]$$

where X and Z are $(NT \times k)$ and $(NT \times g)$ matrices and α and ε are $(NT \times 1)$ vectors. As the PHOGUE provides an unbalanced panel, each of the N individuals in the sample is observed T_i ⁵ times and so the columns in Z and α are constant in blocks of T_i observations. In our case, the only variable in Z is the number of years of completed schooling, whereas X includes the rest of the regressors: other human capital indicators (labour market experience in a quadratic form and two dummies that equal 1 if the worker has received specific training during the last year, be it self-financed or employer-financed), binary variables for personal (married or living with a partner, self-assessed health) and job characteristics (public vs. private sector, part-time vs. full time and job tenure⁶), as well as a dummy for those who were unemployed or out of the labour market the previous year and, finally, the regional unemployment rate.

One of the particular features of this model is the concern that some of the regressors are correlated with the individual effect. Indeed, one of the advantages of panel data is that it allows to avoid the bias that would result when estimating of β and γ in the presence of asymptotic correlation between unobservable heterogeneity, captured in α , and some of the explanatory variables. To present an important example, it has been persistently argued that OLS provides biased estimates due to the correlation between the educational variable and the first error term. This is known as the “ability bias” because α is assumed to include the unobserved ability of each individual, which should be related with the number of years of completed schooling. In a panel data framework, this kind of problem would typically lead to the OLS estimation of [2] after transforming the data in deviations from individual means, that is, using the within-

⁵ For convenience in the notation, we will write throughout the paper the total number of observations in the panel as $NT = \sum_i T_i$

⁶ Tenure is included as a set of dummy variables, since the PHOGUE only offers censored information, being impossible to know the exact length of the relationship with the employer if it has lasted for more than 15 years.

groups estimator or fixed-effects model. The only assumption that needs to be made, in this case, is that the correlation between any variable and the compound error is through the time-invariant individual (fixed) effect. On the other hand, if it is supposed that no variable is correlated with either ε or α , the parameters can be consistently and efficiently estimated by Feasible Generalized Least Squares (FGLS), in what is also known as the random effects estimator. In this case, the procedure simply takes into account the compound nature of the error term, which causes it to be autocorrelated between the observations corresponding to the same individual.

Taking into account that, as we said, some of the regressors might be endogenous with respect to α , applying the within-groups estimator seems more appropriate. Unfortunately, this implies the elimination of time-invariant variables in Z , thus making it impossible to obtain any estimation of γ . Obviously, this is so because, for every individual, the values of z_i are equal to their means over time. In matrix notation, the use of fixed effects means that [2] is pre-multiplied by an idempotent matrix $Q_v = I_{NT} - P_v$ that transforms the data, if grouped by individual, into deviations from individual means⁷. This leads to

$$\begin{aligned} Q_v y &= Q_v X \beta + Q_v \varepsilon \\ \tilde{y} &= \tilde{X} \beta + \tilde{\varepsilon} \end{aligned} \tag{3}$$

Hausman and Taylor (1981) develop their IV procedure so that this problem is solved and γ becomes possible to estimate⁸. More precisely, they suggest that the underlying assumption in the within estimator that every explanatory variable is correlated with α may be excessive. They then propose to partition both X and Z into two submatrices, $X = [X_1 : X_2]$ and $Z = [Z_1 : Z_2]$, where the subscript 1 corresponds to those variables uncorrelated with α and 2 means that the regressors are endogenous⁹. There would be, then, k_1 variables in X_1 , k_2 in X_2 , g_1 in Z_1 and g_2 in Z_2 . In our case, the

⁷ P_v is an idempotent matrix that transforms a vector of observations grouped in T_i -sized blocks into individual means.

⁸ Additionally, they argue that the fixed-effects estimator is inefficient even if there are no time-invariant regressors because it ignores the variation of y and X across individuals, only using the variance within each cross-section unit.

⁹ This set of assumptions is thus somewhere between what the random (no variable is correlated with α) and fixed (every variable is) effects models suppose.

only time-invariant regressor is the educational variable, which we believe to be correlated with α , which means that g_2 equals 1 and Z_I does not exist. Besides, HT estimation is only possible if there are at least as many exogenous time-varying regressors than endogenous constant ones, which in our model means that the condition $k_1 \geq 1$ must be satisfied by whatever set of assumptions we make. If that is the case, then both β and γ can be consistently and efficiently estimated, because X_I provides k_I instruments. Particularly, variables resulting from $Q_v X_2$ are used as instruments for X_2 , whereas $P_v X_1$ provides instruments for Z_I .

As mentioned above, another consideration that needs to be made relates to the variance-covariance structure of the compound error term $u_{it} = \alpha_i + \varepsilon_{it}$:

$$\begin{aligned} \text{Var}(u_{it}) &= \sigma_\alpha^2 + \sigma_\varepsilon^2 \\ \text{cov}(u_{it}, u_{js}) &= \begin{cases} \sigma_\alpha^2 & \text{if } i = j, t \neq s \\ 0 & \text{if } i \neq j \end{cases} \end{aligned} \quad [4]$$

The above implies that the residuals are autocorrelated if whenever they correspond to the same individual. This makes it necessary to use the following GLS transformation of [1]:

$$y_{it} + (\sqrt{\mathcal{G}_i} - 1)y_{i.} = \left[x_{it} + (\sqrt{\mathcal{G}_i} - 1)x_{i.} \right] \beta + \sqrt{\mathcal{G}_i} z_i \gamma + \sqrt{\mathcal{G}_i} \alpha_i + \varepsilon_{it} + (\sqrt{\mathcal{G}_i} - 1)\varepsilon_{i.} \quad [5]$$

where $y_{i.}$, $x_{i.}$ and $\varepsilon_{i.}$ are individual means¹⁰ and $\mathcal{G}_i = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\alpha^2}$. In [5] the errors are homoscedastic and not autocorrelated. Thus, HT is viewed as a three-steps least squares procedure. First, fixed-effects are applied to obtain a consistent estimator for σ_ε^2 . Secondly, a between-groups model is adjusted (using the N individual means) to estimate σ_α^2 . Finally, the data are transformed as in [5] to obtain the FGLS estimator. The only difference with respect to the standard random-effects model is that, in this

¹⁰ e.g., $y_{i\cdot} = \sum y_{it} / T_i$

case, IV estimation is applied in the last two steps on the transformed data instead of OLS.

By construction, $Q_v X_2$ is orthogonal to α and thus those instruments satisfy the condition of not being correlated with the residuals. On the other hand, the absence of asymptotic correlation between $P_v X_1$ and α is only a particular assumption of each model, which will be correct depending on which variables are included in X_1 and the underlying nature of the data. For example, it may seem safe to assume that potential labour market experience or the regional unemployment rate are independent of α , but making similar hypothesis about variables for public/private sector or specific training appears more risky. Fortunately, under some circumstances there is no need to exclusively rely on *a priori* assumptions. If the number of variables in X_1 exceeds that of time-invariant endogenous regressors ($k_1 > g_2$), then a Hausman (1978) test can be computed to compare the HT and fixed effects estimators. This test, under the null hypothesis of no correlation between $[X_1, Z_1]$ (in our case only X_1 , since Z_1 includes no variable) and α , is distributed as a chi-square with $k_1 - g_2$ degrees of freedom¹¹. Intuitively, this test compares the consistent but inefficient within estimate of β with the HT results, which are consistent and efficient under H_0 . If the null is true, then, there should not be a systematic difference between both methods. Being possible to statistically verify the *a priori* assumptions, the decision on which variables to include in X_1 can be then made on the basis of a trial and error process at the end of which the test does not reject the hypothesis that the instruments are orthogonal to α . This process will be discussed in the results section.

3. Data

Our data were drawn from the first four waves of the Spanish section of the European Union Household Panel Survey (PHOGUE) corresponding to the period 1994-1997. During these years 7,825 households were interviewed, collecting information on 20517 persons aged fifteen or older. This survey is a nationally

¹¹ This test can also be applied to the random effects estimator and, in that case, is equivalent to the usual Hausman test for fixed vs. random effects (since the latter is also, the HT estimator when no variable is included in X_2).

representative random sample of private households in Spain conducted by the *Instituto Nacional de Estadística* (INE, 1997). Its general objective is to follow the evolution of Spanish households to study changes in their quality of life, labour conditions and socioeconomic status. Since this panel includes individual information about wages, the highest level of education completed and other variables usually included as regressors in standard wage equations, it is highly suitable for our kind of study,

Our analysis is based on the sample of wage earners aged between 16 and 65 years who work 15 or more hours per week¹². Once missing and anomalous values are excluded and individuals lost because of attrition are considered, the resulting sample is an unbalanced panel of 4492 men and 2487 women, with 11582 and 5971 person-year observations for each, respectively¹³.

The statistical description of the variables used in the estimations appears in table 1. First we observe that salary dispersion is higher for women than for men, while, as expected, the greater average earnings correspond to men. In particular, men's average earnings are 27% higher than women's. With respect to human capital variables, education is measured as the minimum years of schooling required to complete the highest level of education actually achieved. The PHOGUE also provides information about the ending age of education. Then, it is possible to create a variable with the years of schooling that each worker actually attended, subtracting six (the age at the beginning of the studies) to that age. However, using this variable as a regressor has two main problems: (1) people leaving the education system and going back after some time would have anomalous values (for example, 5% of workers reports to have finished their education being 30 or more years old), (2) Moreover it seems clear, intuitively, that this measure of education is subject to more measurement errors than the first one because it depends more on individuals' memory. This is why we choose the election of the minimum years of schooling required to complete the highest level of education as an explanatory variable¹⁴.

¹² Employees working less than 15 hours per week are not included because there is no available information on their tenure and type of sector in the PHOGUE.

¹³ A Person-year observation corresponds to an individual in one year.

¹⁴ Given that the instrumental variables estimator is still consistent in the presence of such measurement errors, we present the results using actual years of education in appendix I. Surprisingly, these are very similar to those obtained using the required number of years.

Table 1
Descriptive statistics of the variables used in the estimations

Variable	Men		Women	
	Mean	St. Dev.	Mean	St. Dev.
Log. monthly earnings^a	11.71	0.48	11.47	0.53
Human capital				
Years of education	9.02	4.05	10.50	4.23
Experience	20.75	12.73	15.93	11.36
Experience squared	592.80	610.52	382.88	472.63
<i>Specific training</i>				
Self-financed	0.08	0.27	0.10	0.30
Financed by employer	0.03	0.16	0.05	0.21
Personal characteristics				
<i>Marital status</i>				
Stable relationship	0.71	0.45	0.57	0.50
No partner	0.29	0.45	0.43	0.50
<i>Self-assessed health</i>				
Very good	0.23	0.42	0.25	0.43
Good	0.59	0.49	0.57	0.49
Acceptable	0.18	0.38	0.18	0.38
Job characteristics				
<i>Sector</i>				
Public	0.24	0.43	0.33	0.47
Private	0.76	0.43	0.67	0.47
<i>Working time</i>				
Full-time contract	0.98	0.15	0.87	0.33
Part-time contract	0.02	0.15	0.13	0.33
<i>Tenure</i>				
Less than 6 months	0.15	0.34	0.15	0.35
6 to 12 months	0.06	0.24	0.07	0.26
1 to 2 years	0.10	0.30	0.12	0.32
2 to 5 years	0.14	0.35	0.19	0.39
5 to 10 years	0.14	0.35	0.15	0.36
More than 10 years	0.41	0.49	0.32	0.47
Not working the previous year	0.11	0.31	0.16	0.36
Unemployment rate	22.13	4.64	21.83	4.45
Sample size	11582		5971	

Source: PHOGUE (INE 1997).

(a) This variable is expressed in real terms (log 1992 pesetas) after being deflated using the consumer prices index (INE 1992-1997).

Descriptive analysis reports that women have, on average, 1.5 more years of education than men. This contrasts with the traditional idea existing in Spain about an educational differential favourable to men with respect to women, especially in higher education. This result can also be the consequence of the greater participation in the labour market of more educated women in relationship with low educated ones. Moreover, sample statistics also report that the incidence of training is greater for females than males, something that can induce to think that training and formal

education are complementary in production. On the other hand, job market experience¹⁵ is five years lower for women than men, showing the recent entry of women into the labour market. It must be noted, that the values of this variable include time spent in career interruptions. During these career breaks, no new investment in human capital occurs, especially if the investment is in training, and the stock of skills depreciates. To take into account these factors, a dummy variable is included in the specification of the econometric model to indicate if the worker was not working (inactive or unemployed¹⁶) in the previous year. Table 1 shows that 16% of the female workers were not working the previous years while this proportion is only 11% for men. This result is consistent with evidence from previous studies (see, for example, Kunze, 2000) and can suggest that career breaks occur more often and last longer for women than for men, placing the former at a disadvantage when they re-enter the labour market.

Concerning personal characteristics, we can first see that the percentage of people with a partner is about 15 percentage points higher for male workers. It's possible that women without partner necessarily have to work to support themselves, whereas married women can specialize in housework or taking care of other family member while, in these households, men take the responsibility of working. Secondly, descriptive statistics don't report significant differences by gender between the health variables, showing that 82% of people has good or very good health¹⁷.

Considering job characteristics, significant differences are observed between female and male workers. First, the portion of wage earners in the public sector is 33% for women and 24% for men. This result is expected and consistent with prior research (Lassibille, 1998), showing that public sector female workers get higher wages than their private sector counterparts for all educational level, whereas this does not always happen for male employees. Secondly, the percentage of employees working part-time is much higher for women than for men (13% vs. 2%). This result could be explained on the basis that full time job opportunities are less available to women or because other unpaid familiar commitments do not allow them to work full time. Whatever the

¹⁵ Experience in the labour market has been computed as the difference between the individual's current age and that at the beginning of his or her working life, being thus a potential labour experience.

¹⁶ A previous estimation showed that considering two separate dummies (for unemployed and inactive) did not yield significantly different coefficients.

underlying explanation, it seems that in general, careers are more precarious and intermittent for women than for men.

Finally, unemployment regional rate is around 22% for both sexes. This variable is a weighted average of the unemployment rates in the provinces that form each of the seven wider Spanish regions that appear in the PHOGUE, where the weight is the number of active people in each province.

4. Results

In this section, we now present the wage earnings models estimated for females and males (see Table 2). We think it is interesting to analyse and compare the estimates when different estimation methods are used and the initial hypothesis change. Thus, we present the results obtained after applying the HT, OLS, within-groups and random effects estimators.

First, OLS estimation suggests that an additional year of schooling increases the net monthly wage by about 6% in both sexes. This finding is similar to what recent Spanish research (see, for example, Alba-Ramirez and San Segundo, 1995; Vila and Mora, 1998; Pons and Gonzalo, 2002)¹⁸ reports, obtaining estimates of the economic return to schooling between 5% and 7%. However, for our data, the use of Breusch-Pagan test (1980) rejects the null hypothesis that the individual effects α do not exist. Moreover, the random effects estimate of ρ ¹⁹ reveals that the proportion of the total error variance accounted for by the unobservable individual-specific error term is roughly three fourths.

¹⁷ These variables are obtained from a subjective indicator about health. In this indicator, people answer if their health is very good, good, acceptable, bad or very bad. As the two last options only represent 2.5% of all people, these categories are counted as acceptable health.

¹⁸ The wage earnings equations estimated in these papers are similar to the classical Mincerian specification of the human capital model. Thus, they only include educational variables, experience and experience squared as regressors. In this paper we estimate an alternative model structure that includes more explanatory variables. As discussed by Pereira and Martins (2001), the inclusion of explanatory variables to the Mincerian specification significantly lowers the estimated return to schooling.

¹⁹ This coefficient may be thought of as representing the proportion of the observed total variance of the error term accounted for by unobserved heterogeneity:

$$\rho = \frac{\sigma_{\epsilon}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2}$$

Table 2. Earnings functions estimates by different methods

Variable	Men ^a				Women ^a			
	HT ^b	OLS	Random effects	Fixed effects	HT ^b	OLS	Random effects	Fixed effects
Constant	9.747***	10.874***	10.828***	11.199***	9.378***	10.651***	10.576***	11.112***
Human capital								
Years of education	0.150***	0.056***	0.056***	--	0.155***	0.057***	0.061***	--
Experience	0.037***	0.022***	0.027***	0.037***	0.032***	0.019***	0.022***	0.029***
Experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
<i>Specific training</i>								
Self-financed	0.007	0.001	0.002	0.007	0.023**	-0.005	0.018	0.024**
Financed by employer	0.025***	0.117***	0.043***	0.025***	0.008	0.056	0.016*	0.008
Personal characteristics								
Stable relationship	0.037***	0.128***	0.094***	0.023	0.027**	0.045***	0.035***	0.019
Self-assessed health								
Very good	0.003	0.024***	0.005	0.002	0.004	0.024***	0.008	0.004
Acceptable	-0.006	-0.028***	-0.015**	-0.008	-0.010	-0.029***	-0.016*	-0.014
Job characteristics								
Public sector	0.009	0.018**	0.030	0.008	0.053***	0.181***	0.173***	0.048**
Part-time contract	-0.300***	-0.472***	-0.359***	-0.299***	-0.259***	-0.509***	-0.362***	-0.258***
<i>Tenure</i>								
6 to 12 months	0.004	0.017	0.012	0.003	-0.011	-0.001	-0.002	-0.011
1 to 2 years	0.028***	0.026**	0.030***	0.026***	0.001	0.006	0.007	0.002
2 to 5 years	0.043***	0.088***	0.069***	0.039***	0.028***	0.100***	0.065***	0.024
5 to 10 years	0.029**	0.117***	0.099**	0.028***	0.026	0.164***	0.113***	0.018
More than 10 years	0.043***	0.233***	0.185***	0.040**	0.016	0.251***	0.176***	0.003
Not working the previous year	-0.035***	-0.133***	-0.078***	-0.034***	-0.068***	-0.095***	-0.088***	-0.067***
Unemployment rate	-0.001	-0.007***	-0.006***	-0.003	0.002	-0.004***	-0.004***	0.000
Observations:								
N	4492	4492	4492	4492	2487	2487	2487	2487
NT	11582	11582	11582	11582	5971	5971	5971	5971
R²		0.530	0.524	0.084		0.643	0.634	0.138
σ_α	0.549		0.524	0.173	0.619		0.287	
σ_ε	0.173		0.289		0.163		0.163	0.162
ρ	0.909		0.773		0.935		0.757	
Hausman test (vs. fixed effects)	5.510		359.190		4.580		366.490	
(degrees of freedom)	10		16		10		16	

(a) The reference is an individual without partner, with good health, working full-time in the private sector, not training and with less than 6 month of tenure.

(b) As the computed R² for instrumental variables models is not bound between 0 and 1, it is not suitable as a goodness of fit measure and thus not presented.

(***) Statistically significant at 1%, (**) at 5% and (*) at 10%.

Having positively tested the significance of α implies that the OLS estimator is not efficient, but it would be consistent if no regressor were correlated with the error term²⁰. In this case, the computed Hausman test, comparing the within-groups and random effects models, strongly rejects the hypothesis that α and the regressors are uncorrelated. Therefore, OLS estimates are neither efficient nor consistent. Besides, under the correlation between α and the regressors, the fixed-effects estimator is

consistent while the random-effects one is not. However, as we discussed, the fixed-effects procedure does not yield any estimate of the returns to schooling since the variable ‘years of completed education’ is a constant within each person. Moreover, this model would mean a loss of efficiency in the coefficients for X variables if α is not correlated with some of them.

On the contrary, the HT estimator is both consistent and efficient if the set of hypotheses about which explanatory variables are not correlated with α is correct. The results we present were obtained assuming as endogenous regressors the following variables: years of education, specific training (whether financed by the worker or his or her employer), public sector and part-time work. *A priori* selection of the rest of the variables as exogenous is supported by the Hausman test results. Moreover, any attempt to increase the described set of exogenous variables is rejected by such tests.

Turning now to the effect of our variables on earnings, the HT estimator shows that an additional year of schooling increases earnings by about 15% for both men and women, more than doubling the effect obtained using OLS or random effects models. In this case, the nature of the correlation between schooling and the error term in the earnings function implies a downward bias in the OLS estimates of the return to schooling. This result is common in previous empirical research (see Hausman and Taylor, 1981; Guillotin and Sevestre, 1994 and Wright (1999) for panel data examples and Card, 1993 for a cross-sectional data framework). Several arguments have been suggested to explain such a bias. Grilliches (1977) emphasizes that measurement errors in schooling are expected to lead to a downward bias in any OLS estimator of the relationship between schooling and earnings. Additionally, Card (1999) proposes an explanation based in the presence of heterogeneous returns to education, which would be lower for high educational levels, something captured in IV estimation if the instruments satisfy certain conditions. Finally, OLS estimates may be subject to a discount rate bias arising from individuals with higher discount rates choosing less education in a model of investment in human capital.

²⁰ In any case, the existence of α implies that OLS standard errors are inconsistent. Therefore, statistical inference about coefficients is invalid.

Another advantage of using HT estimator is that standard errors are somewhat lower than those obtained by means of the within-group estimator, reflecting an improvement in the efficiency of the estimates. Thus, for example, the variable that shows if the individual has a stable relationship is relevant in the HT model, but not significant if fixed-effects are applied.

Concerning other human capital variables we can first see that wages are, as expected, increasing in experience with a decreasing slope, suggesting that wages increase with experience but at a slower rate for older workers. Secondly, investments in job-specific training have a positive effect on earnings for male and female workers. However, for men this is only when training is financed by the employer, whereas for women the effect is positive if the investment is financed by the worker herself.

Other interesting results are related to job characteristic. First, working in the public sector has a positive influence on female wages. This result is similar to that obtained by De la Rica and Ugidos (1995), using information from the Survey of Social Status (INE 1991), or by Lassibille (1998) with data provided by the Family Budget Survey (INE 1990-1991).

With respect to the effects of tenure on wages, on one hand, the specific human capital theory (Becker, 1964; Mincer, 1974) predicts a positive influence on earnings, because the amount of time spent in the same firm is viewed as evidence of accumulation of job specific skills. On the other hand, job matching theory (Jovanovic, 1979) indicates that is the quality of matching that causes the wage growth. In any case, it has also been argued that more able workers tend to keep their job longer (unobserved heterogeneity bias), and thus if the estimation method does not take this into account, an increasing wage-tenure profile will be observed. Considering this possibility and using a method similar to HT, Altonji and Shakotoko (1987) and Ris (2001) find that the effect of tenure on wages is considerably lower. In our case, the influence of unobserved heterogeneity can be assessed comparing the results from the OLS and HT models. While in the former a positive relationship is observed between wage and tenure for both men and women, the HT results suggest a weaker effect for men and practically no influence on women's wages.

Finally, table 2 also shows that the individual labour status in the last year has a significant influence on earnings. Indeed, people not working the previous year has lower wages than the rest. This suggest that the market stigmatizes more intermittent workers, who are perceived to be less productive than the rest.

5. Conclusions

The main objective of this paper has been to estimate for Spain the causal relationship between education and earnings proposed the human capital theory. The chosen econometric method has been the HT estimator (Hausman and Taylor, 1981), so that it is possible to consider in the estimation of the earnings functions the endogeneity of human capital and the unobserved ability bias.

Our results suggest that the return of an additional year of education is about 15% for male and female workers. This result is higher than those obtained in previous Spanish research using OLS that estimate returns to schooling between 5% and 7% (see, for example, Alba and San Segundo, 1995; Vila and Mora, 1998; Pons and Gonzalo, 2001). This result shows, on one hand, the existence of a negative bias in the OLS estimates and, on the other hand, that the return to schooling is still quite high in a context where the educational level has increased (see, for example, Lassibille and Navarro, 1998). The increase of the return once the HT estimator is applied is a common result in the economic literature. For example, Wright (1999) obtains an increase of 10 percentage points using the HT estimator with respect to OLS, with data provided from the British Household Panel. Finally, the use of HT estimator has controlled for and evaluated the influence of unobservable heterogeneity on wages. In particular, the proportion of the total error variance accounted for by unobservable individual heterogeneity represents more than 90% of the variance of the random term.

References

1. Alba-Ramírez, A. and San Segundo, M. J. (1995): "The returns to education in Spain". *Economics of Education Review*, 14(2), pp. 155-66.
2. Altonji, J. and Shakotko, R. (1987): "Do wages rise with job seniority?". Review of
3. Economic Studies, 54(3), pp. 437-59.

4. Angrist, D. and Krueger, A. (1991): "Does compulsory schooling affect schooling and earnings?". *Quarterly Journal of Economics*, 106(4), pp. 979-1014.
5. Ashenfelter, O. and Krueger, A. (1994): "Estimates of the economic return to schooling from a new sample of twins". *American Economic Review* 84(5), pp. 1157-73.
7. Ashenfelter, O., Harmon, C. and Oosterbeck, H. (1999). "A review of estimates of the schooling/earnings relationship, with tests for publication bias" *Labour Economics*, 6(4), pp. 453-470
8. Barceinas, F. et al. (2001): Spain, en *Education ad Earnings in Europe: a cross country analysis of the returns to education*. Editores: C. Harmon, I. Walker y N. Westergaard-Nielsen.
9. Barro, R. (1991): "Economic Growth in a Cross Section of Countries" *Quarterly Journal of Economics*, 106(2) pp. 407-43.
10. Becker, G. (1964): *Human capital*. NBER, New York.
11. Boudon, R. (1974): *Education, opportunity, and social inequality: changing prospects in Western society*. New York, John Wiley and Sons.
12. Breusch, T. and Pagan, A. (1980): "The Lagrange multiplier test and its applications to model specification in Econometrics". *Review of Economic Studies*, 47 (1), pp. 239-53.
13. Card, D. (1993): "Using geographic variation in college proximity to estimate the returns to education". NBER Working Paper 4483.
14. Card, D. (1999): "The causal effect of education on earnings", en Ashenfelter y Card (editores), *Handbook of Labor Economics Volume 3A*. Amsterdam, Elsevier.
15. Card, D. (2000): "Estimating the returns to schooling: progress on some persistent econometric problems". NBER Working Paper 7769.
16. 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(3), pp. 395-414.
17. Griliches, Z. (1977): "Estimating the returns to schooling: some econometric problems". *Econometrica*, 45 (1), pp. 1-22.
18. Guillotin, Y. and Sevestre, P. (1994), "Estimations de fonctions de gains sur données de panel: endogénéité du capital humain et effets de la sélection", *Économie et Prévision*, 116, pp. 119-35.

19. Harmon, C. and Walker, I. (1995): "Estimates of the economic return to schooling for the United Kingdom". *American Economic Review* 85(5), pp. 1278-86.
20. Hausman, Jerry A. (1978), "Specification Tests in Econometrics", *Econometrica*, 46(6), pp. 1251-72.
21. Hausman, J. and Taylor, W. (1981): "Panel data and unobservable individual effects". *Econometrica*, 49(6), pp. 1377-99.
22. Heckman, J. (1979): "Sample selection bias as a specification error". *Econometrica*, 47(1), pp. 153-61.
24. Jovanovic, B. (1979): "Job matching and the theory of turnover". *Journal of Political Economy* 87(5), pp. 972-90.
25. Kunze, A. (2000): "The determination of wages and the gender wage gap: a survey". IZA Discussion Paper n° 193.
26. Lassibille, G. (1998): "Wage gaps between the public and private sectors in Spain". *Economics of Education Review*, 17(1), pp. 83-92.
27. Lassibille, G and Navarro, L. (1998): "The evolution of returns to education in Spain, 1981/1991". *Education Economics* 5(3), pp. 3-10.
28. Marin, A. and Psacharopoulos, G. (1976): "Schooling and income distribution". *Review of Economics and Statistics* 58(3), pp. 332-38.
30. Mincer, J. (1958): "Investment in human capital and personal income distribution". *Journal of Political Economy* 66(1), pp. 281-302.
31. Mincer, J. (1974): *Schooling, experience and earnings*. NBER, New York, EE.UU.
32. Murphy, K. and Welch, F. (1990): "Empirical age-earnings profiles". *Journal of Labor Economics* 8(2), pp. 202-29.
33. Pereira, P. and Martins, P. (2001): ""Returns to education and wage equations". IZA Discussion Paper n° 298.
34. Pons, E. and Gonzalo, M. (2002), "Returns to schooling in Spain: how reliable are IV estimates?" *Labour* 16(4), pp. 747-70.
35. Psacharopoulos, G. (1985): "Returns to education: a further international update and implications". *Journal of Human Resources* 20(4), pp. 583-604.
36. Psacharopoulos, G. (1994): "Returns to investment in education; a global update". *World Development* 22(9), pp. 1325-43.
37. Ram, R. (1989): "Can educational expansion reduce income inequality in less-developed countries?" *Economics of Education Review* 8(2), pp. 185-95.

38. Ris, C. (2001): "Unobservable Heterogeneity in Wage Equations, Application to the Ivorian Manufacturing Sector". Document de Recherche n° 2001-6, Centre de Recherches Economiques de l'Université de Saint-Etienne.
39. San Segundo, M. and Valiente, A. (2003): "Family background and returns to schooling in Spain". *Education Economics*, 11(1), pp. 39-52.
40. Shultz, T.W. (1961): "Investment in human capital". *American Economic Review* 51, pp. 1-77.
41. Trost, R. and Lee, L. (1984): "Technical training and earnings: a polytomous model with selectivity". *Review of Economics Statistics* 66(1), pp. 151-56.
42. Vila, L. and Mora, J. (1998): "Changing returns to education in Spain during the 1980s". *Economics of Education Review* 17(2), pp. 173-78.
43. Wright, R. (1999): "The rate of return to private schooling". IZA Discussion Paper n° 92.

Appendix

Table A1. Earnings functions estimates using actual year of education as a regressor

Variables	Men			Women		
	HT ^a	OLS	Random effects	HT	OLS	Random effects
Constant	8.968***	11.067***	10.979***	8.477***	11.044***	10.904***
Human capital						
Years of education	0.134***	0.024***	0.027***	0.150***	0.018***	0.024***
Experience	0.029***	0.016***	0.023***	0.021***	0.012***	0.016***
Experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
<i>Specific training</i>						
Self-financed	0.007	0.021	0.005	0.025**	0.042*	0.026*
Financed by employer	0.026***	0.154***	0.046***	0.009	0.123***	0.026***
Personal characteristics						
Stable relationship	0.039***	0.158***	0.107***	0.032**	0.064***	0.044***
Self-assessed health						
High	0.001	0.019***	0.004	0.003	0.029***	0.008
Acceptable	-0.006	-0.043***	-0.017***	-0.013*	-0.063***	-0.022**
Job characteristics						
Public sector	0.009	0.084***	0.071***	0.052***	0.282***	0.231***
Part-time contract	-0.300***	-0.450***	-0.346***	-0.259***	-0.504***	-0.349***
<i>Tenure</i>						
6 to 12 months	0.002	0.014	0.013	-0.013	-0.002	-0.004
1 to 2 years	0.027***	0.027*	0.032***	0.000	0.014	0.008
2 to 5 years	0.043***	0.099***	0.075***	0.026*	0.116***	0.070***
5 to 10 years	0.033***	0.158***	0.117***	0.025	0.212***	0.132***
More than 10 years	0.045***	0.288***	0.213***	0.017	0.343***	0.221***
Not working the previous year	-0.036***	-0.139***	-0.078***	-0.068***	-0.103***	-0.093***
Unemployment regional rate	-0.005***	-0.010***	-0.009***	0.005	-0.009***	-0.001***
Observations:						
N	4492	4492	4492	2487	2487	2487
NT	11582	11582	11582	5971	5971	5971
R²		0.531	0.524		0.538	0.518
σ_α	0.806		0.290	1.313		0.329
σ_ε	0.173		0.173	0.163		0.163
ρ	0.956		0.737	0.984		0.803
Hausman Test	10.570		359.190	5.85		507.680
(degrees of freedom)	10		16	10		16

(a) The reference is an individual without partner, with good health, working full-time in the private sector, not training and with less than 6 month of tenure.

(b) As the computed R² for instrumental variables models is not bound between 0 and 1, it is not suitable as a goodness of fit measure and thus not presented.

(***) Statistically significant at 1 percent, (**) at 5 percent and (*) at 10 percent.