ESTUDIOS SOBRE LA ECONOMÍA ESPAÑOLA

Actual and potential returns to schooling in Spain

Carmen García Prieto
Angel Martín Román
Carlos Pérez Domínguez

EEE 157

February 2003

FEDEA
Fundación de Estudios de Economía Aplicada

Actual and potential returns to schooling in Spain.

Carmen García Prieto

Angel Martín Román

Carlos Pérez Domínguez

Departamento de Fundamentos del Análisis Económico
(University of Valladolid)

February 2003

The returns to formal schooling in Spain are estimated in this paper. The main difference between this and previous papers on this subject is that, here, a distinction is made between the increase in the worker's potential maximum wage due to schooling and the actual registered increase. This difference (or underpayment) can be justified on the basis of job search theory. We use the stochastic frontiers technique because it allows the estimation of variables (such as the potential wage) that cannot be directly observed.

One of the main results of this paper is that formal schooling clearly increases a worker's potential maximum wage. This increase is particularly noticeable for those workers who have completed at least a five-year university programme. It has also been estimated that schooling increases the degree of underpayment, which is also quite relevant in the case of long-term university education. In spite of this, the effect of formal schooling on actual wages is clearly positive.

1.- INTRODUCTION.

The analysis of the relationship between an individual's level of formal schooling and their labor income has been of concern to economists for a long time. The work of Cantillon (1755) justifies the payment of higher wages to workers with better qualifications. Still within the classical period, Smith (1776) takes up these ideas once more. It could be said that he is the most direct predecessor of the modern theory of human capital developed at the beginning of the 1960s with the pioneering work of Schultz (1961), Mincer (1962) and Becker (1964).

The systematic analysis of the effect of schooling on labor income has a point of reference, which is fundamental to recent research on the subject: the work of Mincer (1974). The importance of this work is such that, thereafter, the most orthodox income equations have been called Mincer equations. The economic literature derived from this seminal work is ample. The estimation of the returns provided by schooling, based on the econometric adjustment of Mincer equations, is a topic that has given rise to much research over recent years as quality micro-databases have become more generally available in non-anglo-saxon countries[1].

The work of Griliches (1977) looked at some problems that could appear when estimating Mincer equations using ordinary least squared (OLS). One of the most frequent criticisms of the OLS estimation is that an individual's schooling (which is one of the explanatory variables of labor income) is an endogenous variable. This would suggest the use of instrumental variables (IV) econometric methods.[2] Willis & Rosen (1979) pose a different econometric problem: the very samples used to estimate earnings equation may not fulfill the basic requirement of being representative of the whole population. This is known in the literature on the subject as the self-selection bias.

Although this paper is closely related to the entire bibliography on this topic, in the sense that it takes the work of Mincer (1974) as its starting point,

---

[1] Some recent works of research on this subject in the Spanish economy are: Alba-Ramírez & San Segundo (1995), De la Rica & Ugidos (1995), San Segundo (1997), Vila & Mora (1998), Barceinas et al. (2000), García et al. (2001) and Pons & Gonzalo (2001).

[2] In the Spanish case, in accordance with the research of Barceinas et al. (2002), the returns of schooling obtained with IV estimations are very similar to those obtained from the OLS estimations when the sample is adequately screened.

there are differences of both a methodological and conceptual nature which distinguish it from prior economic literature. To our understanding, this paper introduces an original element that sets it apart from other, previous papers. This original element lies in the estimation, not only of the *effective* returns obtained by individuals from their educational qualifications, but also of the *potential* returns associated with the various schooling levels. What is the reason for this difference between an individual's potential wage and the actual wage earned? Because in the job search process workers do not have perfect information, and acquiring these information concerning employment opportunities has a cost. This is the theoretical framework of the so-called *job-search theory* associated with the work of McCall (1970), Mortensen (1970) and Lippman & McCall (1976a & 1976b). According to this theory, individuals fix a critical wage (the reservation wage) and when they get an offer of employment with an associate wage higher than the said critical value, they accept it. This supposes that many individuals will end their search before achieving the maximum wage they could aspire to, given their level of schooling. The difference between that 'potential maximum' wage and the effective wage earned is what is called 'underpayment'.

The aim of this paper is, precisely, to measure that underpayment using Spanish data, and to see whether the difference between the potential and the effective wage increases in line with the individual's level of schooling, or whether the opposite is true. The econometric technique of stochastic frontiers is used to achieve this aim. This technique has habitually been used in the framework of studies concerning productive efficiency, ever since Aigner et al. (1977) and Meeusen & Van den Broeck (1977), defined and used the concept of stochastic frontier in a simultaneous yet independent way.

Nevertheless, there are some applications of this technique in labor economics, to be more precise, in the wage setting process. As far as we know, the pioneering paper was that of Robinson & Wunnava (1989). The technique of stochastic frontiers is used in this paper to measure the female wage discrimination. Other prior papers were those of Hofler & Murphy (1992 and 1994). The first of them estimates up to what point workers achieve an effective wage below the potential maximum they could earn, given their marginal productivity (that is, the wage inefficiency is measured); while the second estimates the worker's reservation wage. Both papers take the framework of the search theory as their setting. McClure, Girma & Hofler (1998) once more take up the question of wage inefficiency, but this time comparing stochastic frontier estimations for the United States and Canada. Polacheck & Robst (1998) take a similar line. Lang (2000) analyzes the question of wage discrimination among

German workers from different ethnic origins. Finally Watson (2000) relates wage inefficiency and the minimum wage in the United Kingdom, and deduces that the said minimum wage is not an effective economic policy in the fight against poverty.

The first new element of our work lies, precisely, in the use of this technique to measure the returns of schooling. It should also be pointed out that, despite the fact that there is already some literature on the subject of *underpayment* and labor income, we believe it to be the first time –and not only in Spain– that an analysis of how the latter is related to the different levels of schooling has been approached. Finally, it should be said that the use of this technique is not only a methodological novelty in the estimation of educational returns, but also that it has interesting properties from the strictly economic point of view. As shall be seen later, in addition to giving a measurement of the potential maximum wage an individual could reach with a particular level of schooling, it will also allow us to measure the efficiency of the employment search process of each educational group.

The rest of the paper is organized as follows: Section two offers the theoretical basis on which our estimations are supported. Section three gives the econometric specifications of the model, pointing out, firstly, the estimation technique used, and secondly, the data used. Section four gives the results obtained in the estimation, while section five, the last, summarizes the main conclusions. The paper ends with four appendices: the first is of a technical nature; the second explains the variables used and shows some statistics describing the sample used; the third shows the complete results of the estimations; and the fourth shows the validation test.

2.- ACTUAL AND POTENTIAL EARNINGS: THEORETICAL BASIS

As a starting point, let us suppose that there is an individual function for generating potential income as described by the following equation:

$$w_i^P = f(X_i) \tag{1}$$

It is a technological relation that determines the maximum wage earnings (w_i^P) that the ith worker can obtain given certain income generating inputs represented by the vector X_i. The said inputs are basically determined by the human capital the worker possesses[3]. It should be pointed out that expression (1) presupposes an 'efficient' behavior of the worker, in the sense that it fixes the maximum amount of money the worker could earn with the best use of his/her formal knowledge and work tenure. In other words, the above expression is an upper boundary for labor income, so the actual or effective income earned by the worker at any given time (w_i) must be lower than, or at best equal to, the maximum potential ($w_i \leq w_i^P$).

The main cause of *underpayment*, that is, that workers may not actually be earning their potential wage, is to be found in the existence of imperfect information. Thus, the job search process becomes costly for the worker. In this way, for a particular individual looking for a job, the best option may be to accept a post offering a wage below his/her maximum potential. This will always be so as long as the marginal cost of continuing to search for employment exceeds the expected marginal benefit of searching.[4]

To be more precise, the reasons why a certain individual's effective wage falls below the maximum potential, or wage frontier, can be put into two main categories.

The first of these categories has an essentially 'objective' nature. It is related to the wage distribution that a particular individual has to face when looking for a job, given his/her earnings generating inputs. It would be expected that workers with the lowest level of formation and tenure would have to face a concentrated wage distribution; that is, that their possible wage range will be fairly restricted. However, as the worker's human capital increases, so will her/his wage possibilities. Thus, individuals with the least qualifications, who face a concentrated wage distribution, should, on average, find themselves

---

[3] In accordance with the proposals of Mincer (1974)

[4] In accordance with the Job Search Theory, the worker's optimum strategy is to determine a reservation wage in such a way that any job offer with a wage below it is rejected, while the first offer of employment providing a wage equal to or higher than the reservation wage is accepted. In order to determine this acceptance wage, the worker must take into account precisely those costs and benefits associated with fixing a marginally higher reservation wage. See McCall (1970), Mortensen (1970) and Lippman & McCall (1976a & 1976b).

closer to their potential wage. A good example is the situation of a teenager, who can only realistically expect to earn the legal minimum wage. It is thus very likely that the wage frontier for such a worker will be very close to the minimum wage and that the effective wage of most workers in this collective is very close to the said frontier.

Other reasons that influence the wage distribution an individual must face concern the structure of the local labor market, that is, its professional and industrial structures.

The second of these categories is related to more 'subjective' questions. On the one hand, there are all those factors that determine an individual's reservation wage, and on the other, the elements associated with their efficiency in searching for employment. Given the same wage distribution, those with higher acceptance wages and who carry out the search mechanisms with greater efficiency will earn wages closer to their maximum potentials. For instance, let us consider two individuals A and B with identical 'objective characteristics''. Let us suppose that A is married, that the spouse works and that they have no children, while B is divorced or married, but that the spouse is unemployed, and that they have children. It would be expected that the individual A would fix a higher reservation wage (he or she will be more demanding in accepting employment), since there is a subsidiary income (that of the spouse) and because they have no children. On the other hand, the individual B would fix a much lower reservation wage for the reasons inversely opposite to those of A. The conclusion is that A, with a high probability, will be closer to their wage frontier than B, simply because he/she has been more demanding when selecting wage offers.

Figure 1 shows graphically the arguments expressed above. The abscissa measures wages and the ordinate measures the number of posts the individual can find for each pre-fixed wage level (thus, what it represents is not exactly a function of density). Two "bells" can be seen. The smaller one (solid line) is associated with the situation of an individual with few income-generating inputs (an unqualified youth). The larger "bell" (dotted line) corresponds to a worker with greater wage possibilities (an older person with a higher level of schooling). Such a figure shows the fact that an older person with better qualifications can always carry out the work of an unqualified youth, while the opposite is not true. The potential wage of the person with a higher level of schooling, w^{P1}, is greater than that of the youth, w^{P0}, and it is this that the estimation of our wage frontier shows.

Let us imagine that the 'subjective' circumstances of both individuals are such that they fix their reservation wages around the mean for the corresponding distributions (w^{r0} for the youth and w^{r1} for the qualified adult). The above supposition incorporates the fact that the person with a higher level of schooling, aware of the fact that she/he has access to a greater variety of job offers, will fix a reservation wage which is greater than that of the young uneducated person. However, the distance to the corresponding maximum potential is smaller (even in relative terms) in the case of the youth than in the case of the person with a higher level of schooling.

3.-   ECONOMETRIC TECHNIQUE

Estimation method

The stochastic frontier estimation techniques offer a plausible means for the estimation of the potential wage a worker could earn.

The method basically consists in considering that the effective wage of each individual is equal to, or lower than, the maximum level that can be reached on the market (the potential wage). The potential wage forms the upper limit of the observations and is obtained, as we have already seen, from a set of variables that give an estimate of the marginal productivity of each individual.

Let w_i be the effective wage earned by the ith worker. We suppose that this can be explained by the following model:

$$log\ w_i = log\ w_i^p - u_i \tag{2}$$

where w_i^p is the potential wage and u_i a random, non-negative disturbance.

The potential wage of each individual, which is their frontier, is obtained from a set of variables all of them reflecting their income generating inputs (basically, their human capital), according to the following specification:

$$log\ w_i^p = \beta' X_i + v_i \tag{3}$$

where β is a parameter vector to be estimated, X_i is the vector of the income generating inputs, and v_i a random disturbance term that gives the frontier a stochastic nature.

Substituting (3) in (2), we get:

$$log\ w_i = \beta' X_i + v_i - u_i \tag{4}$$

We explain the difference between the potential wage and the effective wage a worker earns by using a set of variables $z_i = (z_{1i}, ... z_{mi})'$ specific to each individual[5], and a new random disturbance term ξ_i, in accordance with:

$$u_i = \delta' z_i + \xi_i \tag{5}$$

where $\delta = (\delta_1, ..., \delta_m)'$ is a parameter vector to be estimated.

To guarantee that $u_i \geq 0$, we consider that ξ_i is distributed identically among the sample as a normal with zero mean and variance σ_u^2, truncated at the point $-\delta' z_i$, in such a way that, $\xi_i \geq -\delta' z_i$. Thus, u_i is distributed as a normal variable whose mean depends on the specific explanatory variables of the individuals and truncated at zero, $N^+(\delta' z_i, \sigma_u^2)$. On the other hand, we suppose that v_i is distributed $N(0, \sigma_v^2)$, independent of u_i and of the regressors. Thus, we have a model with a composed error $\varepsilon_i = v_i - u_i$, whose likelihood function, considering the existence of N individuals in the sample, is shown in Appendix 1.

Using maximum likelihood, we get consistent estimators of the frontier parameters, so individuals' potential wage estimation is consistent. The estimators of the parameters that accompany the explanatory variables of the *underpayment*[6] are also consistent. However, the estimation does not give a

---

[5] Following the specification of Huang & Liu (1994) and Battese & Coelli (1995).

[6] In some papers –e.g. Lang (2000)– the previous estimations are carried out in two stages. The determining parameters of the frontier are estimated in the first, while those of the inefficiency are estimated in the second. This procedure is inconsistent (see, for instance Kumbhakar and Lovell, 2000).

value for every u_i, as this is integrated in the compound error term ε_i. To find a specific value for each individual of the amount that separates her/him from their potential wage, it is necessary to consider the conditioned density function $f(u_i|\varepsilon_i)$ [7]

Data used and description of the variables

The data source used to carry out the estimations was the Household Panel of the European Union (PHOGUE) in its third edition, corresponding to the year 1996, with data referring to Spain, which offers individualized information on 15,643 people and 6,268 households. We selected from the sample, wage-earning males working 15 or more hours per week[8], thus reducing the number of observations to a total of 2,780. Likewise, the data concerning the households of these individuals was also used to obtain some variables.

The sample of wage-earning males used gives us a homogeneous and numerous group in the population which minimizes possible problems of self-selection bias. As suggested by San Segundo (1997), bias can come from the use of a sample made up of workers from both sexes[9] or because of the incorporation of self-employed workers whose reported income can be less reliable than wage-earners. Workers clocking up few hours per week have also been excluded from the chosen sample.

---

[7] This was proposed by Jondrow et al. (1982). It would be possible to use the mean or the mode of this conditioned distribution to obtain an unbiased, though inconsistent estimator of the underpayment.

[8] The PHOGUE offers two ways of finding out the hours worked by an individual: their personal statement and the objective classification technique suggested by the Current Population Survey (Spanish National Institute of Statistics). We have chosen the second option.

[9] Pons & Gonzalo (2001), also using Spanish data, justify the use of an exclusively male sample in order to avoid the possible complications derived from the fact that, with females, many professional careers are interrupted by the birth and care of children.

The dependent variable considered is the log of the net hourly wage.[10] The variables of human capital used for the estimation of the wage frontier (X_i) refer to the level of formal schooling reached by the worker, his/her labor tenure and age.[11] The variables used to pick up the difference between the real and the potential wage of each worker (z_i) refer to their age and level of schooling, their marital status and dependent relatives, the individual's non-labor income, the geographical area of residence (NUT), the industry in which the individual works and their labor mobility capacity.

Appendix 2 gives a detailed explanation of the way in which all these variables were made, as well as a summary of the descriptive statistics corresponding to the variables associated with schooling.

4.- RESULTS

Table 1 shows the results obtained in the different estimations we have carried out concerning the variables of formal schooling. Appendix 3 shows the complete set of results. The estimation of the stochastic frontiers was carried out using the computer program FRONTIER 4.1, developed in the Centre for Efficiency and Productivity Analysis (CEPA) of the University of New England (Australia) to study the productive efficiency through frontier functions. A brief guide to how it works can be found in Coelli (1996).

[Insert Table 1]

Appendix 4 shows the results of the validation tests of the estimated models, the main results of which are summarized below.

Columns (I) and (II) of Table 1 shows the results of the estimations made using Ordinary Least Squared which constitute our first reference point. The first of these columns incorporates the number of years of schooling as a continuous variable, obtaining a mean annual return to schooling of 5%. On the other hand, the estimation of column (II) incorporates schooling in a discrete way; in this case, the coefficients show the growth rate of the wage associated

---

[10] The said variable is made up of the quotient between the current net monthly income derived from work as an employee and the mean number of hours worked per month. The use of net wages in our study must be stressed. This fact could explain the slightly lower value of the estimated returns with respect to other papers which use the gross wage, as is the case with Barceinas et al. (2002).

[11] These last two variables try to reflect the worker's specific and generic experience.

to each stage of formal schooling, taking the least qualified group (less primary) as the reference point. The growth of the rate as the subject covers higher levels of schooling can be appreciated. In both cases the estimated returns are in line with those obtained in prior research carried out for Spain.[12]

Columns (III) to (VIII) show the results of the estimations of the stochastic frontiers. In the case of column (III) it can be seen how the maximum potential wage increase associated with an extra year's formal schooling rises to 5.6%. However, the underpayment (difference between the real and potential wage) also increases with schooling by 1% per year –column (V)-, that is, each year of formal schooling separates the maximum potential wage from the effective wage earned by that percentage. As a result of both phenomena, the effective mean wage of a worker grows annually by 4.6% -column (VII)– a figure slightly below that obtained in the OLS estimations.

The results obtained on incorporating levels of schooling as a discrete variable are graphically summarized in figure 2. The continuous growth of the potential returns corresponding to higher levels of schooling and, most especially, that associated with 'long cycle' university studies can be appreciated –column (IV) of table 1–. However, the degree of underpayment – column (VI)– shows a less regular behavior although, in general, it can be seen to be reduced in the groups with the lowest levels of schooling and noticeably higher in the case of 'long cycle' university studies. As a result of these two effects, the increases in the effective mean wage is shown in column (VIII); as it can be seen, in general, the values are slightly lower than those obtained from the OLS estimation –column (II)–.

[Insert Figure 2]

To conclude with the analysis of the results we have put together Table 2, based on the estimations of the stochastic frontier with discrete levels of schooling. It shows information on the estimated hourly wage levels for each schooling group. Column IV of the said table calculates the percentage of wage achievement for each group as the quotient between the 'de facto' earned wage –column II– and the potential maximum –column I–. It is interesting to see that the said percentage reaches its lowest value for the collective with the highest level of schooling. An individual belonging to this group achieves, on average, 73% of their maximum potential wage, while, for the rest of the groups, the

---

[12] See, for instance, the work of San Segundo (1997) and Pons & Gonzalo (2001).

percentage is, on average, equal to or higher than 80%. The dispersion of this wage achievement is also significantly higher in the group with the highest level of schooling, as shown by the variance, which reveals the presence in this collective of a wide range of professions with varying remuneration.

[Insert Table 2]

Looking further into this phenomenon, it can be appreciated how the worst paid individual with 'long cycle' university studies hardly reaches 12% of their potential wage, a value much lower than that corresponding to the rest of the schooling groups. Nevertheless, the degree of wage achievement of the best-paid individual hardly differs in each respective group. This result may be related to the phenomenon of *over-education*, which is especially important in Spain[13].

Finally, it should be pointed out that the lower percentage of wage achievement by the highest qualified individuals is also true for all the deciles of the distribution, the difference being more acute the lower the decile being considered.

5.-   SUMMARY AND CONCLUSIONS

The returns to formal schooling in Spain are estimated in this paper. The main difference between this and previous papers on this subject is that, here, a distinction is made between the increase in the worker's potential maximum wage due to schooling and the actual registered increase. This difference (or underpayment) can be justified on the basis of job search theory

The stochastic frontiers technique was used to carry out the estimations because it allows the approximation of the potential wage, a variable that cannot be directly observed.

The main results obtained are as follows. The increase in the potential wage associated with an extra year of formal schooling is 5.6%. However, the *underpayment* (difference between the real and the maximum potential) also increases with schooling by 1% per year, that is, each year of formal schooling separates the maximum potential wage from that 'de facto' earned by this percentage. As a result of these phenomena the worker's real mean wage increases annually by 4.6%.

---

[13] On this subject, see Dolado et al. (2002) and the related bibliography mentioned there.

If we separate schooling into discreet sections (in accordance with the highest qualification attained by the worker) the increase in the returns associated with higher levels of schooling and, most especially, that corresponding to 'long cycle' university studies can clearly be seen. The degree of underpayment shows, in this case, a less regular behavior. Nevertheless, it can be seen that the magnitude is smaller in the groups with the lowest qualification and noticeably higher in the case of 'long cycle' university studies. In spite of this, the effect on the effective wage of a higher level of schooling is clearly positive.

TABLE 1: Schooling Equation

	OLS		Potential Wage		Underpayment		Actual Wage	
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Years of schooling	5.0%		5.6%		1.%		4.6%	
Primary		4.0%		10.0%		6.9%		2.9%
Secondary (1st. Level)		16.6%		23.7%		8.9%		13.5%
FP (1st. Level)		19.0%		35.2%		16.8%		15.7%
FP (2nd Level)		33.8%		41.8%		11.0%		27.8%
Secondary (2nd Level)		39.5%		49.9%		12.3%		33.5%
University (Short cycle)		82.%		82.9%		5.3%		73.7%
University (Long cycle)		96.3%		145.5%		26.5%		94.1%

- Source: Appendix 3.
- The coefficients associated with the different schooling levels represent growth rates respect to the reference group "less primary and no schooling". These rates have been elaborated in accordance with the following expression: $\exp(\beta)-1$, where β is the coefficient obtained in the estimation (see appendix 3).
- The values of the underpayment and the actual wage are calculated from the estimations of appendix 3 in accordance with the specifications of appendix 1.

TABLE 2: Underpayment by schooling

	Mean Potential Wage (I)	Wage / hour Mean Actual Wage (II)=(I)–(III)	Mean Underpay. (III)	Degree of wage achievement (II) / (I)												
				Mean	Var.	Min.	Max.	10%	20%	30%	40%	50%	60%	70%	80%	90%
Less primary	839.56	708.55	131.01	0.85	0.008	0.50	0.95	0.71	0.81	0.83	0.86	0.88	0.89	0.90	0.91	0.92
Primary	897.16	735.65	161.51	0.82	0.012	0.25	0.96	0.69	0.77	0.81	0.84	0.85	0.87	0.89	0.90	0.92
Secondary (1st. Level)	901.20	747.79	153.41	0.83	0.009	0.31	0.95	0.71	0.78	0.81	0.84	0.86	0.87	0.89	0.90	0.92
FP (1st Level)	976.90	773.93	202.97	0.79	0.017	0.24	0.94	0.66	0.72	0.75	0.80	0.82	0.85	0.87	0.89	0.91
FP (2nd Level)	1023.14	856.80	166.34	0.84	0.008	0.26	0.96	0.73	0.77	0.82	0.84	0.86	0.87	0.89	0.90	0.92
Secondary (2nd Level)	1175.84	975.25	200.58	0.83	0.009	0.23	0.95	0.70	0.76	0.80	0.83	0.85	0.87	0.88	0.90	0.92
University (Short cycle)	1582.39	1371.13	211.27	0.87	0.005	0.46	0.95	0.77	0.83	0.86	0.88	0.89	0.90	0.91	0.92	0.93
University (Long cycle)	2046.52	1491.32	555.20	0.73	0.025	0.12	0.94	0.51	0.61	0.68	0.74	0.77	0.80	0.83	0.86	0.88

- Source: Frontier estimations.
- (I) Mean estimated potential wage for each group of schooling.
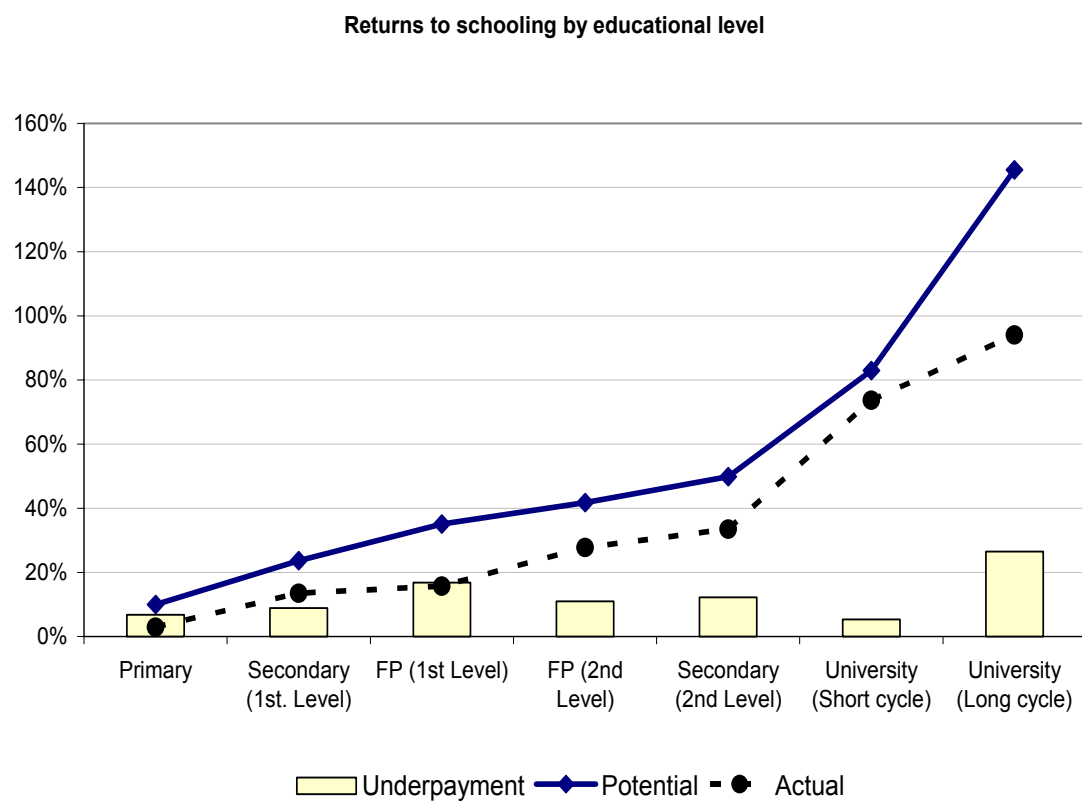- (III) Mean estimated underpayment for each group of schooling.

FIGURE 1

FIGURE 2

Returns to schooling by educational level
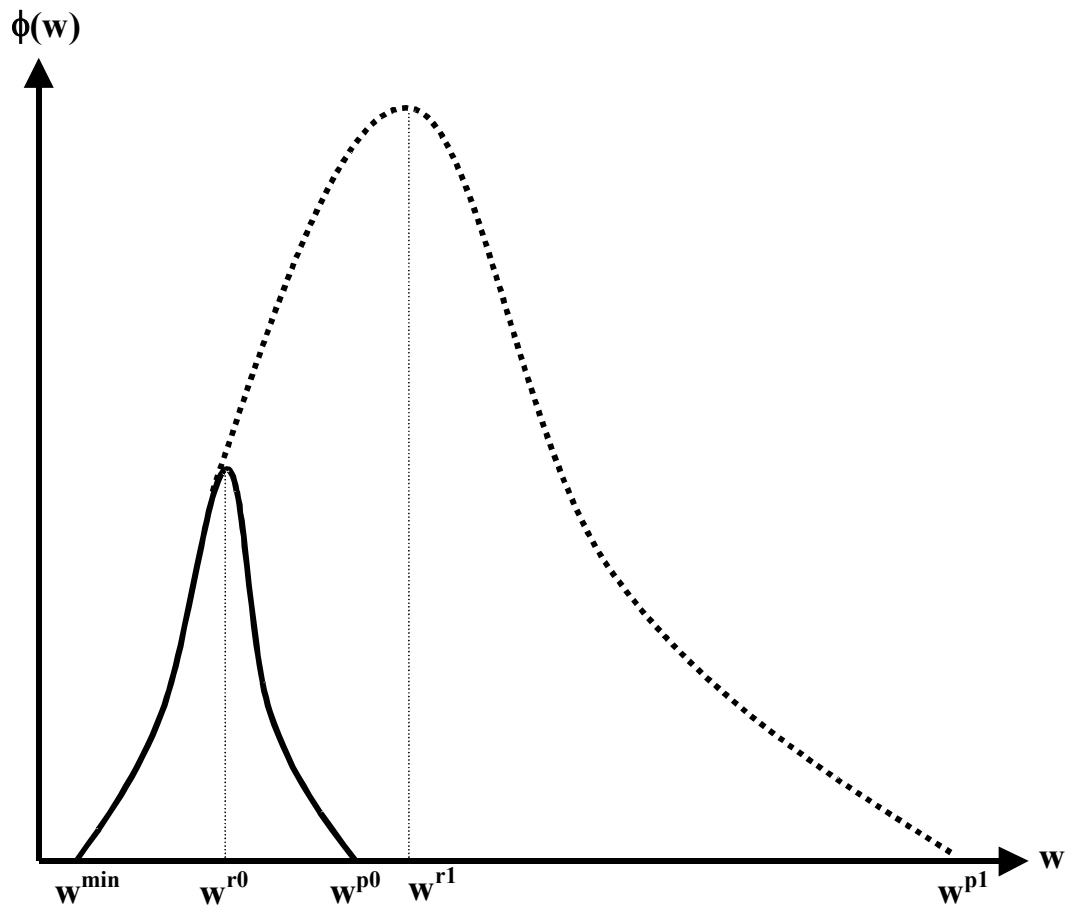
APPENDIX 1: Likelihood function.

The likelihood function used to obtain the frontier estimation is as follows:

$$\ln L = cte - \frac{N}{2}\ln\sigma^2 - \sum_{i=1}^{N}\ln\Phi(d_i) + \sum_{i=1}^{N}\ln\Phi(d_i^*) - \frac{1}{2}\sum_{i=1}^{N}\frac{(\varepsilon_i + \delta'z_i)^2}{\sigma^2}$$

where $d_i = \dfrac{\delta'z_i}{(\gamma\sigma^2)^{1/2}}$ and $d_i^* = \mu_i^*/\sigma^*$, been $\mu_i^* = (1-\gamma)\delta'z_i - \gamma\varepsilon_i$, and $\sigma^* = [\gamma(1-\gamma)\sigma^2]^{1/2}$.

$\sigma^2 = \sigma_u^2 + \sigma_v^2$ is approaching the total residual variance of ε_i, and $\gamma = \sigma_u^2/\sigma^2$ indicates the relative contribution of u_i to this residual variance.

Method to obtain the elasticities:

Following Huang and Liu (1994), the effect on the individual expected effect of a variable x_j that simultaneously explains both the frontier and the underpayment is: $\dfrac{\partial \ln w_i}{\partial x_{ij}} = \beta_j - \psi_i\dfrac{\partial(\delta'Z_i)}{\partial x_{ij}}$, where:

$$\psi_i = 1 - \frac{1}{\sigma_u}\left[\frac{\phi\left(\dfrac{\delta'Z_i}{\sigma_u} - \sigma_u\right)}{\Phi\left(\dfrac{\delta'Z_i}{\sigma_u} - \sigma_u\right)} - \frac{\phi\left(\dfrac{\delta'Z_i}{\sigma_u}\right)}{\Phi\left(\dfrac{\delta'Z_i}{\sigma_u}\right)}\right]$$, ϕ and Φ are the standard normal

density and cumulative distribution functions. This expression is composed of two terms: the effect of the variable on the frontier, β_j, and the effect on the underpayment, $\psi_i\dfrac{\partial(\delta'Z_i)}{\partial x_{ij}}$.

APPENDIX 2: Description of the variables used in the estimation

Sex: Dummy variable that takes the value one if the worker is a woman and zero if it is a man.

Age: Worker's age.

AgeSq: Square of the worker's age.

Ten: Tenure, number of years the worker has been employed in his/her present job.

TenSq: Tenure squared.

Ed1:   Dummy variable that takes the value one if the worker is illiterate or has no studies (2 years) and zero otherwise.

Ed2:   Dummy variable that takes the value one if the worker completed primary school (5 years) and zero otherwise.

Ed3:   Dummy variable that takes the value one if the worker has completed a first level of secondary education (8 years) and zero otherwise.

Ed4:   Dummy variable that takes the value one if the worker has completed a cycle of further education (9 years) and zero otherwise.

Ed5:   Dummy variable that takes the value one if the worker has completed a second cycle of further education (11 years) and zero otherwise.

Ed6:   Dummy variable that takes the value one if the worker has

completed a second level of secondary education (12 years) and zero otherwise.

Ed7:   Dummy variable that takes the value one if the worker has a short-cycle university degree or equivalent (15 years) and zero otherwise.

Ed8:   Dummy variable that takes the value one if the worker has a long-cycle university degree or equivalent (17 years) and zero otherwise.

Marrdep: Dummy variable that takes the value one if the worker is married and has, at least, a dependent child, and zero otherwise.

Marrnodep: Dummy variable that takes the value one if the worker is married and has not dependent children, and zero otherwise.

Singledep: Dummy variable that takes the value one if the worker is single and has, at least, a dependent child, and zero otherwise.

Famincome: Total income of a household where the worker lives, minus that worker's labour income.

Nutma: Dummy variable that takes the value one if the worker lives in Madrid and zero otherwise.

Nutca: Dummy variable that takes the value one if the worker lives in Canarias and zero otherwise.

Nutce: Dummy variable that takes the value one if the worker lives in Castilla y León, Castilla-La Mancha or Extremadura and zero otherwise.

Nutes: Dummy variable that takes the value one if the worker lives in

20

Cataluña, Comunidad Valenciana or Baleares and zero otherwise.

Nutne: Dummy variable that takes the value one if the worker lives in País Vasco, Navarra, La Rioja or Aragón and zero otherwise.

Nutno: Dummy variable that takes the value one if the worker lives in Galicia, Asturias or Cantabria and zero otherwise.

Nutsu: Dummy variable that takes the value one if the worker lives in Andalucía, Murcia, Ceuta or Melilla and zero otherwise.

Agricult: Dummy variable that takes the value one if the worker's professional activity is in the farming sector and zero otherwise.

Energy: Dummy variable that takes the value one if the worker's professional activity is in the energy and mining industries and zero otherwise.

Manufact: Dummy variable that takes the value one if the worker's professional activity is in the manufacturing sector and zero otherwise.

Mineral: Dummy variable that takes the value one if the worker's professional activity is in the metallic and non-metallic mineral sector and zero otherwise.

Machinery: Dummy variable that takes the value one if the worker's professional activity is in the machinery and equipment sector and zero otherwise.

Construction: Dummy variable that takes the value one if the worker's

professional activity is in the construction sector and zero otherwise.

Saleserv: Dummy variable that takes the value one if the worker's professional activity is in the sector of sales services and zero otherwise.

Finaserv: Dummy variable that takes the value one if the worker's professional activity is in the sector of banking and financial services and zero otherwise.

AAPP: Dummy variable that takes the value one if the worker's professional activity is in the Public Administration and Social Services sector and zero otherwise.

Eduhealth: Dummy variable that takes the value one if the worker's professional activity is in the education and health care sector and zero otherwise.

Others: Dummy variable that takes the value one if the worker's professional activity is in other sectors of the NACE and zero otherwise.

Immobility: Dummy variable that takes the value one if the worker was born in Spain and has resided in the same region ever since and zero otherwise.

Main descriptive statistics

	Mean	Max.	Min.	St. Deviation	N obs.
Wage	931,35	5576,32	81,72	516,21	2780
Age	38,76	69,00	17,00	11,15	2780
Tenure	7,83	15,00	0,00	6,20	2780
Education	8,97	17,00	2,00	4,12	2780
Wage by educational level					
Less primary	725,56	1953,53	250,33	266,27	131
Primary	759,54	3037,38	84,42	315,19	794
Secondary (1st. Level)	780,50	2955,61	116,82	358,34	647
FP (1st Level)	798,00	2076,84	108,22	336,88	214
FP (2nd Level)	906,92	3119,16	245,33	446,27	226
Secondary (2nd Level)	1034,69	3178,41	132,86	504,53	319
University (Short cycle)	1431,20	2951,30	400,53	534,04	179
University (Long cycle)	1570,57	5576,32	81,72	777,92	270

APPENDIX 3: Complete estimations

Ordinary Least Squares (OLS)

	Coefficient	St. Deviation.	t-statistic	Coefficient	St. Deviation.	t-statistic
Intercept	4,994	0,086	57,817	5,265	0,092	57,031
Age	0,045	0,004	9,942	0,044	0,004	9,678
Age Sq	-0,000	0,000	-7,8805	-0,000	5,481	-8,012
Tenure	0,039	0,005	7,267	0,041	0,005	7,515
Tenure Sq	-0,000	0,000	-2,782	-0,000	0,000	-2,836
Education	0,050	0,001	28,793			
Ed2				0,039	0,035	1,130
Ed3				0,153	0,036	4,175
Ed4				0,173	0,042	4,092
Ed5				0,291	0,041	6,933
Ed6				0,333	0,039	8,407
Ed7				0,599	0,043	13,922
Ed8				0,674	0,040	16,865
R^2	0,451			0,457		
DW	1,976			1,990		
σ^2	0,368			0,366		
Log-likelihood	-1168,832			-1150,921		

Stochastic Frontiers

	Coefficient	St. Deviation.	t-statistic	Coefficient	St. Deviation.	t-statistic
			FRONTIER			
Intercept	5,343	0,087	61,086	5,581	0,095	58,861
Age	0,035	0,004	7,854	0,034	0,004	7,541
Age Sq	0,000	0,000	-5,191	0,000	0,000	-5,214
Tenure	0,034	0,005	6,567	0,034	0,005	6,681
Tenure Sq	-0,001	0,000	-2,262	-0,001	0,000	-2,056
Education	0,056	0,002	26,577			
Ed2				0,095	0,039	2,414
Ed3				0,212	0,041	5,134
Ed4				0,301	0,049	6,125
Ed5				0,349	0,050	6,984
Ed6				0,405	0,045	8,914
Ed7				0,604	0,052	11,700
Ed8				0,898	0,054	16,744
			UNDERPAYMENT			
Intercept	-2,482	0,627	-3,957	-2,584	0,758	-3,411
Age	0,029	0,006	4,477	0,029	0,007	4,345
Age Sq	-0,512	0,107	-4,778	-0,469	0,093	-5,050
Marrnodep	-0,430	0,118	-3,647	-0,393	0,095	-4,137
Singledep	0,214	0,077	2,783	0,213	0,110	1,943
Famincome	0,000	0,000	-0,320	0,000	0,000	0,132
Education	0,078	0,013	5,825			
Ed2				0,491	0,229	2,142
Ed3				0,641	0,252	2,540
Ed4				1,163	0,339	3,431
Ed5				0,782	0,320	2,446
Ed6				0,865	0,280	3,095
Ed7				0,385	0,403	0,955
Ed8				1,769	0,389	4,553
Nutca	0,501	0,127	3,954	0,498	0,119	4,186
Nutce	0,380	0,109	3,499	0,346	0,097	3,581
Nutes	0,162	0,079	2,048	0,111	0,091	1,216
Nutne	-0,058	0,110	-0,529	-0,071	0,081	-0,880
Nutno	0,732	0,152	4,824	0,713	0,154	4,640
Nutsu	0,467	0,122	3,824	0,475	0,121	3,927
Agricult	0,590	0,117	5,039	0,667	0,159	4,204
Energy	-3,063	1,422	-2,154	-2,718	1,802	-1,508
Manufact	-0,154	0,078	-1,978	-0,151	0,070	-2,171
Mineral	-0,635	0,149	-4,247	-0,667	0,151	-4,406
Machinery	-0,759	0,215	-3,538	-0,804	0,223	-3,604
Construction	-0,670	0,178	-3,767	-0,652	0,208	-3,129
Finaserv	-0,849	0,203	-4,181	-0,845	0,219	-3,866
AAPP	-1,066	0,250	-4,260	-0,980	0,254	-3,861
Eduhealth	-1,271	0,329	-3,858	-0,916	0,217	-4,230
Others	-0,348	0,098	-3,555	-0,346	0,111	-3,126
Immobility	0,258	0,067	3,837	0,247	0,064	3,882
σ^2	0,360	0,064	5,579	0,349	0,058	5,996
γ	0,772	0,040	19,056	0,769	0,037	21,061
Log-likelihood	-980,107			-946,838		
LR-test of one-side error	377,450			408,164		

APENDIX 4: Hypothesis established in order to validate the results

The following hypothesis are established:

- We have considerer a simpler model in wich the underpayment is not dependent on the variables suggested. In such a way, the u_i component would have a constant mean, equal for all the individuals. The hypothesis will be $H_0: \delta_1 = ... = \delta_{23} = 0$ when education is a continuous variable, and $H_0: \delta_1 = ... = \delta_{29} = 0$ when it is discrete.

- If the mean δ_0 is equal to cero, the hypothesis will be $H_0: \delta_0 = \delta_1 = ... = \delta_{23} = 0$ when education is a continuous variable, and $H_0: \delta_0 = \delta_1 = ... = \delta_{29} = 0$ when it is discrete.

- Finally, we contrast the existence of a frontier, $\gamma = 0$ [14]. In this way, the variables that explain the underpayment will became explicative of the effective wage, and the estimation becomes an OLS regression. The hypothesis will be in this case $H_0: \gamma = \delta_0 = \delta_1 = \delta_6 = 0$ when education is a continuous variable, and $H_0: \gamma = \delta_0 = \delta_1 = \delta_6 = ... = \delta_{12} = 0$ when it is discrete, because there is an intercept in the frontier as well as age and education variables.

All these hypothesis are rejected as it can be seen from the next table:

---

[14] Where, $\gamma = \sigma_u^2 / \sigma^2$ indicates the relative contribution of u_i to the total residual variance.

Education as a continuos variable:

NULL HYPÓTHESIS	LOGLIKELIHOOD	STATISTIC (*)	DECISIÓN
H_0: $\delta_1 = \delta_2 = ... = \delta_{23} = 0$	-1120.6	281.1	Reject
H_0: $\delta_0 = \delta_1 = ... = \delta_{23} = 0$	-1137.8	315.5	Reject
H_0: $\gamma = \delta_0 = \delta_1 = \delta_6 = 0$	-1026.5	92.8	Reject

Education with dummies

NULL HYPOTHESIS	LOGLIKELIHOOD	STATISTIC (*)	DECISIÓN
H_0: $\delta_1 = \delta_2 = ... = \delta_{29} = 0$	-1101.5	309.2	Reject
H_0: $\delta_0 = \delta_1 = ... = \delta_{29} = 0$	-1119.5	345.3	Reject
H_0: $\gamma = \delta_0 = \delta_1 = \delta_6 = ... = \delta_{12} = 0$	-1006.6	119.4	Reject

(*)The statistic was calculated for all cases, as $\lambda = -2[\text{loglikelihood}(H_0) - \text{loglikelihood}(H_A)]$. It is distributed as a χ^2 with as many degrees of freedom as parameters considered to be zero in the null hypothesis. In the last test, the statistic is distributed as a mix of χ^2, and the critical values can be found in Kodde and Palm (1986). If it works out to be greater than the value given by the tables at 95%, then the hypothesis is rejected.

References

Aigner, D.J., Lovell, C.A.K. and Schmidt, P.J. (1977) Formulation and Estimation of Stochastic Frontier Production Function Models, **Journal of Econometrics**, 6, pp. 21-37.

Alba, A. and San Segundo, M.J., (1995) The Returns of Schooling in Spain, **Economics of Schooling Review**, 14, pp. 155-166.

Barceinas, F. Oliver, J., Raymon, J.L. and Roig, J.L. (2000) Spain. **In**: Harmon, C., Walker, I. and Westergrad-Nielsen, N. (Ed.), **Schooling and Earnings in Europe: a Cross Country Analysis of the Return to Schooling** (E. Elgar).

Barceinas, F. Oliver, J., Raymon, J.L. and Roig, J.L. (2002) Los rendimientos de la educación en España, mimeo.
http://www.minhac.es/ief/Seminarios/EconomiaPublica/20020214.pdf

Battese, G.E. and Coelli T. (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, **Empirical Economics**, 20, pp. 325-332.

Becker, G. S. (1964) **Human Capital: a Theoretical and Empirical Analysis, with Special Reference to Schooling**, New York: Columbia University Press.

Cantillon, R. (1755) **Essai sur la nature du commerce en général**, translated by Henry Higgs, New York, Augustus Kelley, 1964.

Coelli, T. (1996) A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation, **CEPA working paper** 96/07.

Dolado,J. J., Jansen, M. and Jimeno, J. F. (2000) A Matching Model of Crowding-Out and On-the-Job Search (with an application to Spain)", **Documento de trabajo FEDEA** 02-16. FEDEA, Madrid.

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.

García, J., Hernández, P.J. and López A. (2001) How Wide is the Gap? An Investigation of Gender Wage Differences Using Quantile Regression, **Empirical Economics**, 26(1), pp. 149-167.

Griliches, Z. (1977) Estimating the Return to Schooling: Some Econometric Problems, **Econometrica**, 45, pp. 1-22.

Hofler, R. A and Murphy K., J. (1992) Underpaid and Overworked: Measuring the Effect of Imperfect Information on Wages, **Economic Inquiry**, 30 (3), pp. 511-529.

Hofler, R. A and Murphy K., J. (1994) Estimating Reservation Wages of Employed Workers Using a Stochastic Frontier, **Southern Economic Journal**, 60, 4, pp.961-976.

Huang, C. J. and Liu, J. T. (1994) Estimation of a Non-Neutral Stochastic Frontier Production Function, **The Journal of Productivity Analysis**, 5, pp. 171-180.

Jondrow, J., Lovell, C.A.K., Materov, I.S. and Schmidt P. (1982) On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, **Journal of Econometrics**, 19, 233-238.

Kodde, D.A. y F.C. Palm (1986) Wald criteria for jointly testing equality and inequality restrictions, **Econometrica**, 54, 1243-1248.

Kumbhakar, S. C. and Lovell, C. A. K. (2000) **Stochastic Frontier Analysis**, Cambridge University Press. New York

Lang, G. (2000) Native-immigrant Wage Differentials in Germany - Assimilation, Discrimination, or Human Capital?, **Universitaet Augsburg, Institute for Economics, discussion paper,** n. 197

Lippman, S. A. and McCall, J. J. (1976a) The Economics of Job Search: a Survey", Part I, **Economic Inquiry**, 14, pp. 155-189.

Lippman, S. A. and McCall, J. J. (1976b) The Economics of Job Search: a Survey", Part II, **Economic Inquiry**, 14, pp. 347-368.

McCall, J. J. (1970) Economics of Information and Job Search, **Quarterly Journal of Economics**, 84, 113-126.

McClure, K. G., Girma, P. B. y Hofler, R. A. (1998) International Labour Underpayment: a Stochastic Frontier Comparison of Canada and the United States, **Canadian Journal of Regional-Science**,21(1), pp. 87-109.

Meeusen, W. and Van den Broeck, J. (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, **International Economic Review** 18, 2, pp. 435-444.

Mincer, J. (1962) On-the-job Training: Costs, Returns and Some Implications, *Journal of Political Economy*, 70 (5), part 2.

Mincer, J. (1974) **Schooling, Experience, and Earnings**, Columbia University Press. NBER, New York.

Mortensen, D. T. (1970) Job Search, the Duration of Unemployment and the Phillips Curve, **American Economic Review**, 60, pp. 847-862.

Polacheck, S.W. and Robst J. (1998) Employee Labor Market Information: Comparing Direct World of Work Measures of Workers' Knowledge to Stochastic Frontier Estimates, **Labour Economics**, 5, 231-242.

Pons, E and Gonzalo, M. T (2001) Return to Schooling in Spain. How Reliable Are IV Estimates?, **IV Jornadas de Economía Laboral**, Valencia.

Robinson, M.D. and Wunnava, P. V. (1989) Measuring Direct Discrimination in Labor Markets Using a Frontier Approach: Evidence from CPS Female Earnings Data, **Southern Economic Journal**, 56, pp. 212-218.

San Segundo, M J. (1997) Educación e ingresos en el mercado de trabajo español, **Cuadernos Económicos de ICE**, 63, pp. 105-123.

Schultz, T. W. (1961) Invesment in Human Capital, **American Economic Review**, 51: pp. 1-17.

Smith, A. (1776) **An Inquiry into the Nature and Causes of the Wealth of Nations**, eds. R.H Campbell and A.S Skinner (1976).

Vila, L. and Mora, J.G. (1998) Changing Returns to Schooling in Spain During the 80's, **Economic of Schooling Review**, 17 (2), pp. 173-178.

Watson, D. (2000) UK Underpayment: Implications for the Minimun Wage, **Applied Economics**, 32, pp. 429-440.

Willis, R. J. and Rosen, S. (1979) Schooling and self-selection, **Journal of Political Economy**, 87 (5), pp. 817-836.