AN UNBIASED ESTIMATION OF WAGE FRONTIERS

Pérez Villadóniga María Jose[1]

Rodríguez-Alvarez, Ana[1,2 *]

(1) Departamento de Economía. Universidad de Oviedo

(2) Oviedo Efficiency Group

(*) Corresponding author:

Departamento de Economía

Facultad de Economía y Empresa

Universidad de Oviedo.

Campus del Cristo s/n

33006 Oviedo.

e-mail: ana@uniovi.es

Tfno: 34-985-10-48-84

Abstract

The analysis of the determinants of differences in wages across workers has traditionally been based on the estimation of mean earnings functions following Mincer (1974). As this methodology presents several econometric problems, in this paper, we propose an alternative technique based on the estimation of wage frontiers. First, we propose a new theoretical model of workers' behaviour, where workers choose the amount of investment in human capital, as well as marginal productivity in order to achieve their maximum earnings. Second, both human capital and earnings are likely to be influenced by worker's (unobserved) ability, leading to endogeneity problems. The empirical implementation of the theoretical model allows us to obtain consistent estimates. Using data from the Spanish Wage Structure Survey 2006, we find a positive effect on wages of human capital variables. Finally, we explain workers' inefficiency to obtain their maximum potential wage. Results show that having a temporary contract increases inefficiency. Interestingly, females seem to be less efficient than males in achieving their maximum potential wage.

Jel codes: C51, J24, J31

Keywords: Production (P), Input distance functions, endogeneity, earnings functions, frontier analysis

1. Introduction

Earnings vary widely across individuals. According to the human capital theory, most of these differences in wages are due to differences in individuals' human capital, such as education, training, and labour market experience. As with other forms of investment, the decision to acquire skills to enhance one's productivity requires the outlay of resources now for returns in the future. Then, differences between the wages of individuals with different levels of human capital must reflect differences in the returns necessary to compensate the costs of acquiring these skills. In a perfectly competitive labour market, wages equal the value of the marginal product, and there should be no differences in wages among workers with identical characteristics. However, labour markets are far from perfectly competitive. For instance, the role of institutions, as trade unions and collective bargaining systems, is especially relevant, making equilibrium wages differ from the competitive outcome. Also, incomplete information and search costs, discrimination and other market imperfections may play an important role in the determination of wages, and may generate differences in wages across individuals with identical endowments.

Then, the market equilibrium (or potential) wage is determined both by individuals' human capital and the nature of the job, and by labour market conditions. However, not all workers manage to attain their potential wage. Several researchers (see, for instance, Polachek and Robst 1998) have recognized that we should distinguish between the observed wage that the worker receives and the potential wage, i.e. the maximum wage attainable given the worker's human capital endowments and market conditions.

In production economic theory, production functions provide the maximum possible output for given inputs. Translating this idea to labour economic theory, the worker's production function provides the maximum wage that the worker can obtain given his human capital and the characteristics of the labour market where he is employed (i.e. technology). The gap between the observed and the potential wage is known in the literature as worker's inefficiency (for instance, Hunt-McCool and Warren 1993 or Jensen, Hermann and S. Rässler 2010). Then, the efficient workers will get the equilibrium market wage while the inefficient ones will be below their frontier.

Robinson and Wunnava (1989), Dawson, Hinks and Watson (2001), and Lang (2005) have used frontier functions to analyse wage discrimination by gender and/or nationality. The underlying assumption of this approach is that the wage an individual receives is equal to the maximum level that he could attain in the labour market (potential wage) minus an error term that captures inefficiency. The potential wage constitutes the upper frontier of the observations, and is obtained from a set of variables that proxy the marginal productivity of each individual and other market characteristics.

In this paper, we link labour and production economic theory to model the transformation of workers' human capital into marginal productivity. In particular, we present a model where investment in schooling and labour market experience are inputs, and the marginal productivity (MP) of the worker is the output. The contribution of this paper is twofold. First, and in contrast to previous literature, we develop a theoretical model that introduces assumptions on workers' behaviour. This theoretical contribution is important both from an economic and an econometric perspective. Workers are economic agents and so are maximizers. The amount of human capital inputs (education or experience) is determined by the worker's decisions as a function of his ability. Then, the chosen amounts of inputs could be endogenously decided. In this case, theoretical endogeneity will cause econometric problems, and the estimated coefficients will be biased.

The second contribution of this paper is methodological. In order to adapt the empirical specification to the theoretical model proposed, we proxy technology using an input oriented distance function. In contrast to the traditional production function approach usually used to model wage frontiers, where inputs are assumed to be exogenous, by using an input distance function we will be able to estimate technology consistently, even under the assumption of endogeneity both in inputs and output.

The paper is organized as follows. In section 2 we introduce the concept of wage frontier, comparing it to the notion of average wage function as in Mincer (1974). In section 3 we present a theoretical model of the behaviour of the worker and in section 4 we describe the corresponding empirical implementation. After describing the data in section 5, in section 6 we present the econometric specification and the results of the estimation of a wage frontier model, where the wage frontier is estimated together with the equation of the determinants of wage inefficiency. Finally, we present the main conclusions regarding the determinants of wage inefficiency.

2. Earnings, productivity and inefficiency

To explain how much of the differences in observed earnings can be explained by the human capital theory, economists have relied on the estimation of earnings functions, starting with the seminal work of Mincer (1974). Despite its wide application, several problems arise when estimating Mincerian equations using OLS. Below, we address these issues.

a) Average equation versus frontier equation

We begin by defining the standard Mincerian (average) earnings equation. Let us denote E_i individual *i*'s observed earnings, which depend on human capital. The Mincerian earnings function can be written as:

$$\ln E_i = \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_s x_{si} + \alpha_e x_{ei} + \alpha_{ee} x_{ei}^{\ 2} + \varepsilon_i \qquad (1)$$

where α_0 is a constant term; P_I and P_F are vectors of individual and firm characteristics, respectively; x_s represents a measure of schooling and x_e stands for labour market experience. The motivation for allowing experience to enter quadratically is that it permits a nonlinear pattern in the lifetime earnings profile. Finally, ε_i is a disturbance term, assumed to be distributed normally and independently of the human capital variables.

However, as Lovell (2001) points out, a valuable extension of this literature would be the use of production frontiers methodology to construct an earnings frontier, this is, the maximum attainable wage, given the individual's human capital endowments and other individual characteristics that may influence earnings. This information is more accurate than that derived from previous research (based on average functions) given that it includes the possibility of being inefficient in the achievement of the objective of maximizing earnings given the stock of human capital. Moreover, this would allow computing the "efficiency" with which individuals or groups of individuals approach their earnings frontier. Then, the distance to the frontier would indicate individuals' inefficiency to attain their potential wages once we have taken into account individuals' or groups' characteristics.

If workers are inefficient in the transformation of human capital into earnings, we can rewrite the Mincerian wage equation as:[1]

$$\ln E_i = \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_s x_{si} + \alpha_e x_{ei} + \alpha_{ee} x_{ei}{}^2 + v_i - u_i \qquad (2)$$

where equation (2) differs from (1) in the way the error term is modeled. Concretely, the error term in (2) has two components: v_i, which is a normally distributed error term with zero mean and variance σ_v^2; and u_i, that follows a one-tail distribution (so workers can be on the frontier or below it). Thus u_i, that is the difference between observed earnings and

[1] See, for instance, Kumbhakar and Lovell (2000) for a survey on frontiers.

the potential wage[2], captures a worker's "inefficiency" to obtain the maximum wage attainable given his productivity.

b) The functional form

Mincer (1974) uses a semilogarithmic model, more specifically, a log-linear model given that the distribution of income is log-normal. However, these not-flexible models impose restrictions on technology. To address this issue, researchers have estimated non-parametric models or added higher-order terms in the schooling or experience variables. However, as Card (1999) acknowledges, these models are not always satisfactory and we "need more flexible interactions between education and experience". In this paper, we propose a translog frontier model, which is a second order approximation to the true but unknown production function, where human capital variables are the inputs and marginal productivity is the output. The translog function, which has been largely used in production economics studies, but less frequently in labour economics, is a more flexible functional form and does not impose as many restrictions on the parameters as the log-lin model. A translog production function stochastic frontier can be defined as follows:

$$\begin{aligned}\ln E_i &= \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_s \ln x_{si} + \alpha_e \ln x_{ei} + \\ &+ \frac{1}{2}\alpha_{ss} \ln {x_{si}}^2 + \frac{1}{2}\alpha_{ee} \ln {x_{ei}}^2 + \alpha_{se} \ln x_{si} \ln x_{ei} + v_i - u_i\end{aligned} \tag{3}$$

As in equation (2), to allow for the existence of technical inefficiency, we have added a composed error term, where u_i represents the "inefficiency" of worker i in the transformation of human capital variables into wages. With this specification, we allow both schooling and experience to enter quadratically, as well as interactions between education and labour market experience.

[2] The concept of frontier is relative, i.e. it is not the theoretical frontier, but the one obtained from the observations. Each worker is compared to the most efficient one in the sample, but this does not mean that the latter reaches his theoretical frontier.

c) *Endogeneity problem*

In the Mincerian earnings function, earnings are endogenous and the human capital variables are assumed to be exogenous this is, uncorrelated with the error term. However, it is likely that both human capital inputs and output are influenced by the ability of the worker, which goes largely unobserved. If so, conventional least squares estimates of (1), (2), or (3) lose their usual desirable properties, including consistency. This potential bias has long been of interest to labour econometricians (for example Griliches 1977). To address the issue of endogenity in the estimation of earnings functions, researchers have used one of four methods.[3] The first one is to find a proxy for ability and include it in the earnings equation (Griliches and Mason 1972; Blackburn and Neumark 1995). However, it is difficult to obtain ability measures that are not determined by schooling. A second approach exploits differences between twins in the level of schooling and earnings based on the assumption that they share the same unobserved ability (Griliches 1979). This method is subject to many criticisms, as measurement errors in schooling may lead to larger bias and the results may not be easily generalized to the non-twin population. Third, the ability bias can be eliminated by using panel data and treating ability as a fixed effect. Still, the rate of return to schooling can only be obtained for individuals who return to school. Fourth, several researchers have taken advantage of exogenous variation in factors that affect schooling decisions to obtain instruments for schooling that are uncorrelated with ability (Angrist and Krueger 1991; Harmon and Walker 1995). However, the instruments used in the literature have been challenged (Carneiro and Heckman 2002). In this paper, we propose a theoretical model and its corresponding empirical implementation that will allow us to obtain consistent estimates, even under the assumption of endogeneity of the human capital variables.

[3] See Harmon and Walker (1995) or Ashenfelter, Harmon and Oosterbeek (1999) for a review.

3. The theoretical model

Let us consider a worker who has to decide on human capital investment in order to maximize earnings, given the costs of acquiring those skills and given the transformation process of human capital into productivity.[4] The worker's objective function can be expressed as:

$$\begin{aligned} &\max_{x,MP} \ \frac{MP_i \times P_{NC}}{(w'x)_i} \\ &s.t \quad D(MP_i, x_i, P_I, P_F) = 1 \\ &s.t \quad E_i \le E_i^* \end{aligned} \tag{4}$$

Under perfect competition, the wage received by the worker would be equal to the value of the marginal product, i.e. the marginal productivity multiplied by the market price of the product sold. However, most labour markets are far from perfectly competitive, and therefore earnings will be determined by a (non-competitive) coefficient, P_{NC}, on the marginal product. This coefficient will depend on factors related to the market and the firm, such as union's relative bargaining power and the size of the firm among others. In this sense, earnings ($E_i = MP_i \times P_{NC}$) are a function of both the worker's marginal productivity (MP_i) and the conditions relative to the firm and the (non-competitive) market where the individual is employed. Finally, w is the input price vector and $w'x$ represents the cost of investment in human capital undertaken by the worker. Under these assumptions, the worker will choose a certain combination of schooling and experience to maximize earnings, given the environmental restrictions and taking into account that the wage cannot exceed the equilibrium market wage (E_i^*).

In order to capture technology, that is to say, "the process of production" of inputs (education and experience) into output (MP), we use an input distance function (IDF)

[4] This idea stems from the seminal paper of Georgescu-Roegen's (1951), based on the maximisation of the return to the outlay (return to the dollar), and that has been recently applied to firms' behaviour by Kumbhakar, (2011)

widely used in production economics (see for example, Färe and Primont 1990 or Atkinson, Färe, and Primont 2003). The IDF equals one when the worker is on the frontier. Besides, we include environmental (P_F) and individual (P_I) variables, that may affect the process of transformation of human capital into MP.[5]

Under these assumptions, the Lagrangian function associated to (4) is:

$$L = \frac{MP_i \times P_i}{(w' x)_i} - \mu(1 - D(.)) - \theta(E_i^* - E_i) \tag{5}$$

And the first order conditions with respect to the decision variables are:

$$\frac{\partial L}{\partial x_s} = MP \times P \times \frac{-1}{(w'x)^2} w_s + \mu \frac{\partial D(.)}{\partial x_s} = 0 \tag{6a}$$

$$\frac{\partial L}{\partial x_e} = MP \times P \times \frac{-1}{(w'x)^2} w_e + \mu \frac{\partial D(.)}{\partial x_e} = 0 \tag{6b}$$

$$\frac{\partial L}{\partial MP} = \frac{P}{(w'x)} + \mu \frac{\partial D(.)}{\partial MP} = 0 \tag{6c}$$

Multiplying (6a), (6b) and (6c) by x_s, x_e and MP, respectively, we can rewrite these equations as:

$$\frac{\partial L}{\partial x_s} = \frac{-MP \times P \times w_s x_s}{(w'x)^2} + \mu \frac{\partial \ln D(.)}{\partial \ln x_s} D = 0 \tag{7a}$$

$$\frac{\partial L}{\partial x_e} = \frac{-MP \times P \times w_e x_e}{(w'x)^2} + \mu \frac{\partial \ln D(.)}{\partial \ln x_e} D = 0 \tag{7b}$$

$$\frac{\partial L}{\partial MP} = \frac{P \times MP}{(w'x)} + \mu \frac{\partial \ln D(.)}{\partial \ln MP} D = 0 \tag{7c}$$

[5] The IDF is dual of the cost function. For details and empirical applications see for example Baños-Pino *et al.* (2002).

Rearranging the above equations, we get the following condition:

$$\frac{\partial \ln D(.)}{\partial \ln x_s} + \frac{\partial \ln D(.)}{\partial \ln x_e} + \frac{\partial \ln D(.)}{\partial \ln MP} = 0 \tag{8}$$

Where equation (8) includes the first order conditions obtained from the theoretical model (4). In sum, the idea of introducing assumptions on workers' behaviour is important both from an economic and an econometric perspective. Workers are economic agents and so are maximizers. Hence, given the restrictions faced by the worker, the amount of education, experience and MP is determined through the worker's objective function, so they can be endogenously decided. Moreover, it is feasible to assume that the worker's choice of MP and human capital investment is the outcome of other unobservable factors, such as ability. Under this interpretation it is ability that is exogenous, whereas education, experience and MP are jointly endogenously determined by maximisation people's decisions (based on their ability).

In this view, the Mincerian earnings functions (equations 1-3) may be subject to the simultaneity problem: education and experience may be endogenous explanatory variables. If so, estimates of the earnings equation lose their usual desirable properties including unbiasedness. In the following section we propose an empirical model based on Kumbhakar (2011) that allows us to estimate technology circumventing these econometric problems, by imposing the first order conditions obtained in (8).

4. The empirical model

Based on the idea that Kumbhakar (2011) applied to the theory of the firm, we propose an estimation strategy that allows us, still recognizing theoretical endogeneity, to estimate wage frontier technology consistently. To do so, the transformation process of human capital into MP is represented by an input-oriented distance function stochastic frontier instead of the traditional production function stochastic frontier usually used to estimate

Mincerian wage equations. Besides, we will use a flexible functional form, in particular a translog. Hence, the translog IDF stochastic frontier we propose is:

$$\begin{aligned}
\ln D_i &= \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_y \ln(MP)_i + \frac{1}{2}\alpha_{yy} (\ln MP)_i^{\,2} + \\
&+ \alpha_s \ln(x_s)_i + \alpha_e \ln(x_e)_i + \frac{1}{2}\alpha_{ss} \ln(x_s)_i^{\,2} + \frac{1}{2}\alpha_{ee} \ln(x_e)_i^{\,2} + \\
&+ \alpha_{se} \ln(x_s)_i \ln(x_e)_i + \alpha_{sy} \ln(x_s)_i \ln(MP)_i + \alpha_{ey} \ln(x_e)_i \ln(MP)_i + v_i - u_i
\end{aligned} \tag{9}$$

Where again, x_s and x_e represent the schooling and the labour market experience inputs, respectively, MP is the marginal productivity and P_I and P_F represent individual and firm characteristics, respectively, and u_i represents the “inefficiency” of worker *i* in the transformation of human capital variables into MP.

By imposing the first order condition (8) in equation (9) we get:

$$\begin{aligned}
&\alpha_s + \alpha_{ss} \ln x_s + \alpha_{se} \ln x_e + \alpha_{sy} \ln MP + \alpha_{ee} \ln x_e + \alpha_{se} \ln x_s + \\
&\alpha_{ey} \ln MP + \alpha_y + \alpha_{yy} \ln MP + \alpha_{sy} \ln x_s + \alpha_{ey} \ln x_e = 0
\end{aligned} \tag{10}$$

Finally, by adding the condition of homogeneity of degree one in inputs to (10), which is a property of the IDF, we get: [6]

$$1 + \alpha_y + \alpha_{yy} \ln MP + \alpha_{sy} \ln(\frac{x_e}{x_s}) = 0 \tag{11}$$

This restriction will hold for any $\ln MP$ and $\ln(x_e / x_s)$ if and only if:

[6] H(1) in inputs implies $\alpha_s + \alpha_e = 1; \quad \alpha_{ss} + \alpha_{se} = 0; \quad \alpha_{es} + \alpha_{ee} = 0; \quad \alpha_{sy} + \alpha_{ey} = 0$.

$$\alpha_y = -1; \alpha_{yy} = 0; \alpha_{sy} = 0 \tag{12}$$

By imposing conditions of equation (12) together with the conditions of homogeneity of degree one in inputs in equation (9), we get the following equation to estimate:

$$ln\left(\frac{MP}{x_s}\right)_i = \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_e \, ln\left(\frac{x_e}{x_s}\right)_i + \frac{1}{2}\alpha_{ee} \, ln\left(\frac{x_e}{x_s}\right)_i^2 + v_i - u_i \tag{13}$$

Note that in the right-hand side of the above equation both education and experience appear as regressors in a ratio form. This property of equation (13) is especially relevant for the objectives of this paper. In our theoretical model, both education and experience are considered as endogenous variables, as they are influenced by individuals' unobserved ability. To deal with this issue we model the relationship between human capital inputs and ability as follows.

First, the schooling input of individual *i*, x_{si}, i.e. the educational human capital the employee is paid for, depends on the level of formal schooling, x_{is}^{obs}, and on the ability A_i of the worker. Then, different individuals may have spent the same time at school, but with different efficiency:

$$x_{is} = x_{is}^{obs} A_i \tag{14}$$

Similarly, the effect of an individual's experience on the production process also depends on the time spent in the labour market, x_{ie}^{obs}, and on his ability. Then the experience input, x_{ie}, can be expressed as:

$$x_{ie} = x_{ie}^{obs} A_i \tag{15}$$

Substituting (14) and (15) in (13) we have that:

$$ln\left(\frac{MP}{x^{s^{(obs)}}\times A}\right)_i = \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_e ln\left(\frac{x^{e^{(obs)}}\times A}{x^{s^{(obs)}}\times A}\right)_i + \frac{1}{2}\alpha_{ee} ln\left(\frac{x^{e^{(obs)}}\times A}{x^{s^{(obs)}}\times A}\right)_i^2 + v_i - u_i \qquad (16)$$

Then,

$$\ln\left(\frac{MP}{x^{s^{(obs)}}}\right)_i = \alpha_0 + \sum_I \alpha_I P_I + \sum_F \alpha_F P_F + \alpha_e \ln\left(\frac{x^{e^{(obs)}}}{x^{s^{(obs)}}}\right)_i + \frac{1}{2}\alpha_{ee} \ln\left(\frac{x^{e^{(obs)}}}{x^{s^{(obs)}}}\right)_i^2 + \phi - u_i \qquad (17)$$

where $\phi = \ln A_i + v_i$. Note that in equation (17) the explanatory variables do not depend on ability, so the omission of this variable does not cause endogeneity problems. Hence, regressors in (17) will be independent of the random error term ϕ (for details see Coelli 2000 or Kumbhakar 2011). In conclusion, by imposing the first order condition (8) derived from the theoretical model to the empirical specification, we are able to obtain consistent estimates, despite recognizing the endogenetiy of the human capital variables. Besides, equation (17) will capture the inefficiency of the worker to obtain the maximum potential wage given his human capital and the characteristics of the environment where he operates. This will allow us to compute Wage Efficiency Indexes (WEI) for each worker in the sample, using the following expression:

$$\text{WEI} = \exp(-u) \qquad (18)$$

The values of the WEI indexes range between zero and one. If the WEI takes the value 1 the worker is on the frontier of his potential wage; values below 1 imply inefficiency. In the next subsection we specify how to analyze the determinants of this inefficiency.

4.a. Determinants of the inefficiency

To explain workers' inefficiency in obtaining their maximum potential wage, equation (17) will be estimated simultaneously with an equation that specifies the determinants of inefficiency. Traditionally, the analysis of the determinants of inefficiency has been carried out by means of a second stage analysis, i.e. after the efficiency indexes have been obtained they are regressed against a set of explanatory variables. However, several researchers have acknowledged problems of inconsistency with this methodology (see Wang and Schmidt 2002).

By using the model of Batesse and Coelli (1995), the inconsistency of the second stage analysis is avoided. This model assumes that the term u_i follows a truncated normal distribution with mean μ_i, and common variance σ_u^2, $u_i \rightarrow$ iid $N^+(\mu_i, \sigma_u^2)$. Then, μ_i is modeled as a function of a set of variables that may affect wage inefficiency:

$$\mu_i = \delta z_i + W_i \tag{19}$$

where z_i is a px1 vector of variables that may influence wage inefficiency, and δ is the 1xp parameter vector to be estimated. Finally, W_i is a random variable obtained from the truncation of a normal distribution, where ($-z_i\delta$) is the point of truncation. Therefore, the relevance of the proposed methodology is that it allows us to specify economic inefficiency in terms of a set of explanatory variables without resorting to second stage analysis.

5. Data

To carry out our empirical analysis we use data from the Wage Structure Survey 2006, which was conducted by the Spanish Statistics Institute. The selection of the sample follows a stratified two-stage sampling. In the first stage, establishments, which were previously stratified by region and size, are selected. In the second stage, workers at each establishment are selected randomly. The survey includes 19,308 firms and 147,616 salary workers, and provides information on individuals' personal characteristics, their wages, and the firm where they work.

Our empirical analysis requires data on the amount of inputs, as well as other characteristics both of workers (including marginal productivity) and firms. Regarding the input variables, we have transformed the school level variable to obtain years of formal education. On the other hand, the sample does not provide information on individuals' experience in the labour market, so we compute potential experience as age minus years of schooling minus 6, as is standard in labour economics. A fortunate feature of this dataset is that we have disaggregated information on professional category. In particular, workers are classified to the two-digit groups of the International Standard Classification of Occupations 1988 (ISCO-88). Then, we have included a dummy variable for each category. We have also included dummy variables that indicate whether the worker has responsibility on the job, the length of the contract, as well as gender and nationality of the individual.

In order to control for firm-specific factors, we include a dummy variable for each firm. Finally, as happens with most datasets, we have no information on individuals' marginal productivity. However, given that our model is defined in logs and includes controls for professional category and firm, we can approximate marginal productivity using earnings: if we assume that two workers with equal marginal productivity, employed in the same firm and in the same category receive the same wage (which seems a plausible assumption), the measurement error will not affect our results.[7]

The hourly wage (in logs) is calculated by dividing the monthly wage by the number of hours worked. The monthly wage is obtained as the sum of the base wage, payments for

[7] Note that in our model $E=MPxP_{NC}$, i.e. $MP=E/P_{NC}$. Then, $\ln MP=\ln E-\ln P_{NC}$. If P_{NC} is considered to be constant for each firm and category, then P_{NC} will only affect the constant term (in particular the firm and category dummy variables) and we will be able to use E as a proxy of MP without biasing results.

extraordinary hours and wage complements, which include seniority payments, pluses for activity, productivity, attendance, incentives, languages and qualifications, from which we deduct complements for shift work, work at the weekend or on holidays, and night work.

Table 1 provides information on the number of firms and workers selected within each sector of activity. Given the large number of firms in most sectors, we restrict our analysis to the energy industry, where the number of firm dummies is tractable. Then our sample consists of 59 firm and 843 salary workers, classified into nearly forty occupational categories.

Descriptive statistics are shown in Table 2. On average workers in the sample have completed over 12 years of formal schooling, and have over 24 years of potential experience. Nearly 80% of the individuals in the sample are male. About 30% of the workers hold a job with responsibility, and less than 8% have a temporary contract. The proportion of immigrants in the sample is less than 1%. The average log hourly wage is 2.683 (around 16.8 Euros). With respect to the occupational distribution, as shown in Appendix A.I, about half of the workers in the sample are employed in professional occupations and nearly one third are in skilled blue-collar jobs.

6. Econometric specification and results

To carry out the empirical analysis we estimate the system of equations (17)-(19). First, in the wage frontier specification, equation (17), the vector of individual variables, $P_{I,}$ includes job category, nationality, length of the contract and whether the individual has responsibility in the workplace, which may affect marginal productivity. We also include a dummy variable for each firm, $P_{E,}$ as firm's characteristics affect wages. Second, to analyse the determinants of inefficiency jointly with the frontier, in equation (19) we include a set of dummy variables that capture differences across individuals, such as nationality, responsibility in the job, length of the contract and gender. In this way, we can explain the distance to the frontier, i.e. the workers' inefficiency to achieve the maximum attainable MP given their human capital endowments.

With respect to the gender dummy variable, we need a more detailed analysis. In principle, in the frontier we should *only* include those variables that affect workers' frontier and not inefficiency. From this perspective, while responsibility in the workplace, length of the contract, and nationality can affect MP and, therefore, wages there are no *a priori* reasons to expect that gender may influence productivity. In other words, we assume that both men and women are potentially equally productive, so they can both be on the same wage frontier and gender should not be included in the specification of the frontier (equation 17). The case is somewhat different when we analyze the determinants of wage inefficiency. Moreover, it is especially relevant for the aim of this paper to include the gender of the worker in order to determine whether it affects wage inefficiency, as well as the implications. If the coefficient is not significantly different from zero, this means that gender does not contribute to explain inefficiency. A different result would lead to interesting conclusions and would require a more exhaustive analysis.

Estimated maximum likelihood parameters of the equations system (17-19) are presented in Table 3. In this translog model, the variables have been divided by their geometric mean, so the first order coefficients can be interpreted as elasticities at the mean value of the data. All the first order parameters have the expected signs and are highly significant, which implies that the estimated technology complies with the theoretically expected monotonicity condition (increasing in inputs).[8] Likelihood ratio tests show that, for the data used, the translog model, which incorporates interaction terms between education and experience, is a better representation of the production technology than the traditional linear model.[9]

[8] The coefficients estimated in the distance function do not have a direct interpretation, so we would have to resort to the duality between the distance function and the cost function, to interpret them as a normalized shadow price of each input. However, this analysis is beyond the scope of this paper.

[9] The value of this test was 40.6, higher than the critical value of the *chi*-square distribution for 1degree of freedom at the usual levels of significance.

With respect to nationality, we observe that workers from Latin America are, on average, 0.84% less productive than national workers, while the effect for the other nationalities is not significant. Having a temporary contract does not affect the wage frontier. As expected, responsibility in the workplace has a positive and significant effect on the frontier, that is to say, workers with more responsibility likely to be more productive and hence, receive a higher wage. In particular, a worker who has no responsibility in the workplace receives a wage that is 0.24% lower relative to a worker with the same human capital but who has responsibility.

Table 4 displays the mean elasticities for the various occupations considered relative to the reference category (Managers and Senior Officials). On average, workers in any occupation other than the reference category are less productive and receive a lower wage, and the differences are highest for less skilled workers, as expected. In particular, ceteris paribus, a worker in an elementary occupation receives an hourly wage that is 53% lower that an individual in a manager position.

Regarding the determinants of inefficiency, equation (19), we observe that the coefficient of temporary contract is positive and significant, showing that increases in job insecurity explain why workers may lie below their wage frontier. This may reflect the fact that having a temporary contract implies a weaker attachment of the worker with the firm and, therefore, less on-the-job training and promotion possibilities. Then temporary contracts are associated to a higher wage gap. On the other hand, while responsibility shifts the frontier upward, workers with responsibility on the job seem to be less efficient in reaching their potential wage. However, the contribution of this variable to explain inefficiency is only significant at the 10 per cent level. While the variable nationality is not significant, we find that gender is a significant determinant of inefficiency.

The positive and significant coefficient on the gender variable means that female workers *systematically* receive a lower wage, relative to their male counterparts who are employed *in the same sector of activity, firm and occupation* (disaggregated up to two digits). This is, being a female increases the difference between the maximum potential wage and the actually perceived wage. A potential explanation of this result lies in the occupational classification used in this analysis. Professional categories, although quite disaggregated, does not include the third and fourth digits of the ISCO-88. Once we have controlled for the characteristics of the workers, firms and sector, the wage differential will be due to the fact that, *ceteris paribus*, within each professional category, women tend to be employed in the lowest paid jobs. Then, given similar human capital endowments, women tend to be systematically situated below their frontier. This result highlights the existence of occupational segregation in the energy sector in Spain. Within each occupational category, women tend to hold the lower status jobs, although given their human capital they could have accessed to higher ranking positions. Then, we observe the existence of a glass ceiling in this sector.

In order to analyse inefficiency in more detail, we can compute Wage Efficiency Indexes (WEI) that capture the distance of each worker from his potential wage as defined in equation (18). These indexes can take values between 0 and 1. If the index takes the value 1, this means that the worker reaches the maximum potential wage, given his human capital, so he is completely efficient. Conversely, a value close to 0 would indicate that the worker is very inefficient in approaching his potential wage.

WEI by broad occupations, as displayed in Graph 1, shows that efficiency is lower in the highest rank occupations for both genders. This result can be explained if we take into account the fact that dispersion within these positions is higher, which may give rise to higher wage variation. Interestingly, females are significantly less efficient than males in managerial and professional occupations. Again, this may be evidence of the existence of a glass ceiling as, within these occupations, women seem to be concentrated in the lowest paid jobs.

Finally, in Graph 2 we plot WEI against years of schooling. Efficiency decreases slightly with years of schooling for both men and women, although males are more efficient for most levels of education. This means that less educated workers manage to get closer to their potential wage. In consonance with the previous results, the reason may be that less educated workers can access to jobs where dispersion is lower, so differences in wages tend to be small. In contrast, more educated workers are in more qualified occupations, where the range of wages is much larger. This result is similar to that of De la Rica (2010) who finds that occupational segregation within firms is a major determinant of the observed gender wage gap in Spain.

6. Conclusions

To analyze the determinants of differences in wages across workers, economists have traditionally relied on the estimation of Mincerian earnings functions. Although this methodology has been widely used, it presents several estimation problems. In this paper, we study the determinants of wages from a different perspective. Our contribution is twofold. First, we propose a theoretical model of workers' behaviour. In this model, workers are assumed to maximize earnings, given the costs of acquiring the necessary human capital. Workers will choose the amount of human capital (education and labour market experience), as well as marginal productivity in order to achieve their maximum earnings. Second, we tackle the endogeneity problem that results from the theoretical model in order to obtain unbiased parameters.

The empirical analysis consists of the estimation of a system of two equations: a wage frontier, which yields the maximum attainable wage given the worker's human capital endowments, and an equation of the determinants of workers' inefficiency, i.e. the distance between the maximum potential wage and the wage the worker receives.

The estimation of the wage frontier shows that human capital variables – education and potential experience – as well as having a job that entails some kind of responsibility have a positive effect on productivity and, hence, on wages. Coming from Latin America have the opposite effect. Having a temporary contract does not shift the earnings frontier. With respect to the determinants of inefficiency, having a temporary contract increases inefficiency. Also, females seem to be less efficient than males in achieving their maximum potential wage.

References

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

Angrist J. D. and A. B. Krueger (1991). Does Compulsory Schooling Attendance Affect Schooling and Earnings?, *Quarterly Journal of Economics*, 106 (4), 979-1014.

Ashenfeter, O., Harmon, C. and H. Oosterbeek (1999). A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias, *Labour Economics*, 6, 453-470.

Atkinson, S., Färe, R. and D. Primont (2003). Stochastic estimation of firm inefficiency using distance functions, *Southern Economic Journal*, 69 (3), 596-611.

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

Blackburn, M. L. and D. Neumark (1995). Are OLS Estimates of the Returns to Schooling Biased Downward? another look. *Review of Economics and Statistics*, 77 (2), 217-229.

Baños-Pino, J., Fernández-Blanco, V. and Rodríguez-Álvarez, A. (2002). The Allocative Efficiency Measure by Means of a Distance Function: the Case of Spanish Public Railways. *European Journal of Operational Research*, 137-1, 191-205

Card, D. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, vol. 3, Amsterdam: North Holland, pp. 1801-1863.

Carneiro, P. and J. J. Heckman (2002). The Evidence on Credit Constraints in Post--secondary Schooling. *Economic Journal*, 112 (482), 705-734.

Coelli, T. J. (2000). On the Econometric Estimation of the Distance Function Representation of a Production Technology. *Discussion paper 2000/42*, Center for Operations Research and Econometrics, University Catholique de Louvain.

Cornes, R. (1992). *Duality and Modern Economics*, Cambridge, University Press.

Dawson, P., Hinks, T. and D. Watson (2001). German Wage Underpayment: an Investigation into Labor Market Inefficiency and Discrimination, *Vierteljahrshefte zur Wirtschaftsforschung*, 70, 107-114.

De la Rica, S. (2010). Segregación Ocupacional y diferencias salariales por género en España: 1995- 2006, in *Mujeres y mercado laboral en España*, FBBVA-IVIE, pp: 21-47.

Färe, R. and Primont, D. (1990). A Distance Function Approach to Multioutput Technologies. *Southern Economic Journal*, 56, 4, 879-891.

Färe R. and D. Primont (1995). *Multi-Output Production and Duality: Theory and Applications*, Kluwer Academic Publishers, Norwell, Massachusetts.

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistics Society*, *Serie A*, 120 (3), 253-281.

Georgescu-Roegen, N. (1951). The aggregate linear production function and its applications to von Neumann's economic model, in: Koopmans, T. (Ed.), *Activity Analysis of Production and Allocation*, Wiley, New York. 98–115.

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

Griliches, Z. (1979). Sibling Models and Data in Economics: Beginnings of a Survey. *Journal of Political Economy*, 87 (2), 37-64.

Griliches, Z. and W. M. Mason (1972). Education, Income and Ability. *Journal of Political Economy*, 80 (2), 74-103.

Harmon, C. and I. Walker (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *American Economic Review*, 85 (5), 1279-1286.

Hunt-McCool, J. C. and R. S. Jr. Warren (1993). Earnings Frontiers and Labor Market Efficiency. in H. Fried, K. Lovell and S. Schmidt (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, 197-209.

Jensen, U., Hermann, G. and S. Rässler (2010). "Estimating German Overqualification with Stochastic Earnings Frontiers. *Advances in Statistical Analysis*, 94, 33–51.

Kumbhakar, S. (2011). Estimation of Production Technology when the Objective is to Maximize Return to the Outlay. *European Journal of Operational Research*, 208, 170-176.

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

Lang, G. (2005). The Difference between Wages and Wage Potentials: Earnings Disadvantages of Immigrants in Germany. *Journal of Economic Inequality*, 3, 21-42.

Lovell, C.A.K. (2001). "Mirando Hacia Delante: Oportunidades de Investigación Futura en el Análisis de Eficiencia y Productividad, in A. Álvarez Pinilla (Ed.), *La medición de la eficiencia y productividad*, Editorial Pirámide. Madrid. 331-343.

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

Mincer (1974). *Schooling, Experience and Earnings*, NBER, 1972, New York.

Polachek, S. W. and J. Robst (1998). Employee Labor Market Information: Comparing Direct World of Work Measures of Workers' Knowledge to Stochastic Frontier Estimates, *Labour Economics*, 5 (2), 231–242.

Robinson, M. D. and P. V. Wunnava (1989). Measuring Direct Discrimination in Labor Markets Using a Frontier Approach: Evidence from CPS Female Earnings Data. *Southern Economic Journal*, 56 (1), 212-216.

Shephard, R. W. (1953). *Cost and Production Functions*, Princeton, NJ, Princeton University Press.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*, Princeton, NJ, Princeton University Press.

Wang, H. and P. Schmidt (2002). One Step and Two Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 18, 129-144.

Table 1. Firms and workers by sector of activity

Sector of Activity	No. firms	No. workers
Extractive industries	157	880
Manufacturing	8777	69591
Energy industry	59	843
Construction	1905	12876
Wholesale & retail trade	1831	12523
Hotels & restaurants	981	6547
Transport & communication	798	5176
Financial intermediation	571	5398
Real state, renting & business activities	1878	14831
Education	716	6096
Health, veterinary & social services	541	7529
Other social activities	1094	5326

Table 2. Descriptive statistics

	Mean	Std. Dev.
Hourly wage (in logs)	2.683	0.503
Years of education	12.488	3.156
Potential experience	24.512	11.092
Male	0.795	0.404
Responsibility in the job	0.308	0.462
Temporary contract	0.078	0.269
Spanish nationality	0.993	0.084
Number of observations	843	

Table 3. Estimation of the wage frontier and the determinants of inefficiency

	Coefficient
Ln schooling	0.663***
	(0.152)
Ln schooling square	0.167***
	(0.027)
Ln experience	0.337***
	(0.015)
Ln experience square	0.167***
	(0.027)
Ln schooling*Ln experience	0.167***
	(0.027)
UE nationality	0.017
	(0.282)
South American nationality	-0.834***
	(0.202)
Asian nationality	-0.072
	(0.261)
Temporary contract	0.124
	(0.078)
No Responsibility on the job	-0.235***
	(0.032)
Determinants of inefficiency	
Temporary contract	3.208***
	(1.133)
Female	1.240**
	(0.622)
No responsibility on the job	-1.106*
	(0.669)
UE nationality	-1.279
	(11.824)
South American nationality	-29.637
	(111.396)
Asian nationality	-29.569
	(166.141)

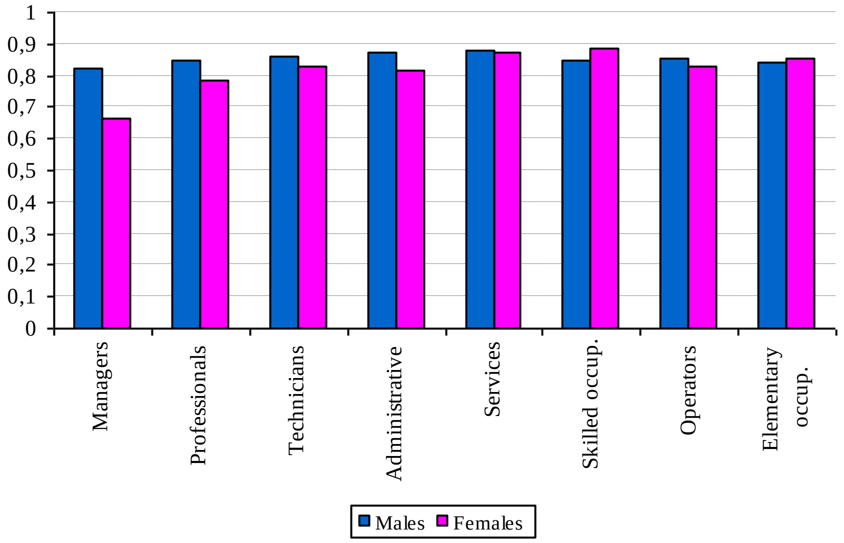
Notes: Standard errors in parenthesis. *significant at 10%;

** significant at 5%; *** significant at 1%.

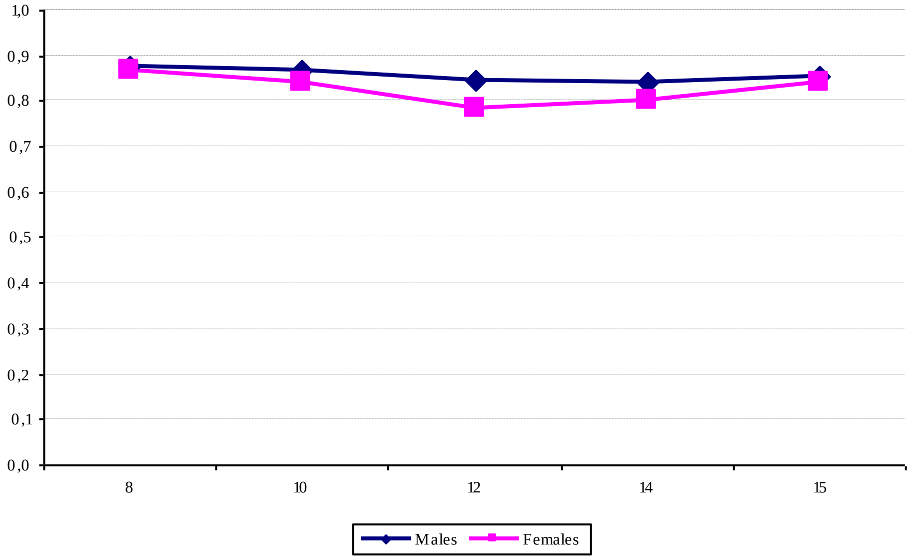
Table 4. Average elasticity by broad occupation

Category	Average elasticity
Professional occupations	-0.297
Technical occupations	-0.366
Administrative occupations	-0.805
Services occupations	-1.102
Skilled occupations	-0.431
Operators	-0.671
Elementary occupations	-0.533

Graph 1. Wage Efficiency Indexes by occupation



Graph 2. Wage Efficiency Indexes by years of schooling



Appendix I. Distribution of the sample by occupation

	Mean	Std dev
1. Managers & senior officials		
11. Corporate managers	0.060	0.239
2. Professional occupations		
20. Science & engineering professionals	0.026	0.160
23. Legal professionals	0.007	0.084
24. Business & administration professionals	0.023	0.149
25. Artists, writers & related occupations	0.002	0.049
26. Science & engineering associate professionals (3-year college)	0.018	0.132
27. Health associate professionals (3-year college)	0.002	0.049
29. Other associate professionals (3-year college)	0.004	0.060
3. **Associate professional & technical occupations**		
30. Science & engineering technicians	0.203	0.402
31. Health associate professionals	0.004	0.060
33. Business, sales & finance associate professionals	0.057	0.232
34. Administrative management associate professionals	0.109	0.312
4. Administrative & secretarial occupations		
40. Accounts & wages clerks, book-keepers, other financial clerks; production & transport clerks	0.043	0.202
43. Non-client information workers not elsewhere classified	0.045	0.208
44. Other information workers not elsewhere classified	0.025	0.156
45. Travel consultants & clerks, telephonists & receptionists	0.001	0.034
5. Customer service, personal service & trades occupations		
50. Personal services occupations	0.001	0.034
52. Protective services workers	0.002	0.049
53. Sales and related workers	0.002	0.049
6. Skilled agricultural & fishing occupations		
60. Skilled agricultural workers	0.008	0.091

7. Craft workers & skilled workers in manufacturing, construction & mining		
70. Building frame & related trades workers	0.007	0.084
71. Building frame & related trades workers not elsewhere classified	0.008	0.091
72. Building finishers & related trades workers, painters & related trades workers.	0.046	0.210
73. Metal, machinery and related trades workers	0.006	0.077
75. Sheet and structural metal workers, moulders and welders, and related workers	0.001	0.034
76. Machinery mechanics and repairers	0.065	0.247
8. Plant & machine operators, & assemblers		
80. Team leader and stationary plant responsible	0.009	0.097
81 Stationary plant & machine operators & related workers	0.146	0.353
83. Stationary machine operators	0.001	0.034
84. Mechanics and assemblers	0.006	0.077
85. Locomotive engine drivers, farm plant and mobile plant operators	0.005	0.069
86. Locomotive engine drivers and related workers; and heavy truck and bus drivers	0.012	0.108
9. Elementary occupations		
91. Domestic, hotel & office cleaners & helpers	0.002	0.049
92. Building caretakers, window cleaners & security guards	0.001	0.034
93. Other elementary workers in other services	0.015	0.123
96. Construction labourers	0.009	0.097
97. Manufacturing labourers	0.012	0.108
98. Transport and storage labourers	0.004	0.060