

**WAGE DISTRIBUTION IN SPAIN, 1994-1999:  
AN APPLICATION OF A FLEXIBLE ESTIMATOR  
OF CONDITIONAL DISTRIBUTIONS**

**Antonia Febrer and Juan Mora\***

WP-EC 2005-04

Correspondencia: J. Mora, Universidad de Alicante, Dpto. Fundamentos del Análisis Económico, Apdo. 99, E-03080 Alicante, Spain / E-mail: [juan@merlin.fae.ua.es](mailto:juan@merlin.fae.ua.es)

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Febrero 2005

Depósito Legal: V-995-2005

*IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.*

---

\* A. Febrer y J. Mora: Universidad de Alicante.

**WAGE DISTRIBUTION IN SPAIN, 1994-1999:  
AN APPLICATION OF A FLEXIBLE ESTIMATOR OF  
CONDITIONAL DISTRIBUTIONS**

**Antonia Febrer and Juan Mora**

**ABSTRACT**

To investigate the trends in wages in Spain in 1994-1999, we propose a flexible estimator of conditional distributions. The estimator, based on a piecewise-linear specification of the conditional hazard function, allows us to capture almost any underlying relationship and is unaffected by the curse of dimensionality. Our results reveal that the main changes in the labor market involved graduate workers entering the labor market: the “overeducation” phenomenon intensified in Spain between 1994 and 1999, provoking a decrease in returns to schooling at higher levels of education.

*J.E.L. Classification Numbers:* C14, J31.

*Keywords:* Hazard-Based Estimation; Wage Distribution.

**RESUMEN**

En este trabajo proponemos analizar la evolución de los salarios en España entre 1994 y 1999 utilizando un estimador flexible de las distribuciones condicionales. Este estimador, basado en una especificación lineal a tramos de la función de razón de fallo condicional, permite captar casi cualquier relación subyacente, y no se ve afectado por la maldición de la dimensionalidad. Los resultados que obtenemos muestran que los cambios más importantes en el mercado laboral se han producido en el grupo de los trabajadores con estudios superiores que entran al mercado laboral. En concreto, se observa que el fenómeno de “sobreeducación” se intensificó en España entre 1994 y 1999, provocando un descenso de los rendimientos de la educación en los niveles de educación superiores.

*Palabras clave:* Estimación basada en la función de fallo; Distribución salarial.

## 1. INTRODUCTION

The analysis of wage distributions has always played a major role in Economics. Thanks to the availability of large data sets, the use of nonparametric statistical techniques to perform such analysis is now widespread. However, to understand better the sources of differences in wages, much recent research has focused on the estimation of the conditional distribution of wages given certain explanatory variables, also referred to as “covariates”, which may be continuous or discrete. For this kind of analysis, nonparametric methods are not entirely appropriate because they suffer from what has been referred to as “curse of dimensionality”: when a large number of covariates is included, the nonparametric estimation of the conditional distribution is extremely inefficient. To overcome this problem, various semiparametric procedures have been proposed, see e.g. Buchinsky (1994), Fortin and Lemieux (1995) or DiNardo, Fortin and Lemieux (1996). More recently, Donald, Green and Paarsch (2000), hereafter DGP, devised an estimation method which provides very flexible estimators of conditional wage distributions, in the sense that only a minimal number of restrictions on the shape of the conditional densities are imposed. In this paper we propose to use a generalization of the DGP estimation method which shares its advantages and yields better fits.

The DGP procedure is based on a semiparametric specification of the conditional hazard function. Hazard functions were first used as a device to specify models in the previously developed literature on spell duration (see e.g. Meyer 1990). The reason why the hazard function is used as a starting point is that it makes it easy to introduce flexible functions of the covariates with no great computational complexity, and it allows the covariates to affect not only the mean

or variance of the conditional density, but also its shape. The estimation method described in DGP assumes a step conditional hazard function. This assumption leads to easy-to-compute estimates which can capture the underlying shape of the true conditional hazard function, as long as a sufficiently large number of discontinuities are allowed. However, a step fitted hazard function leads to a step fitted density function, which is not an appropriate estimate in many cases. To circumvent this problem, DGP suggest smoothing the estimated densities after they are calculated. In this paper we propose a different solution: assuming a piecewise-linear conditional hazard function. This does not introduce any greater computational complexity since there is no increase in the number of parameters to be estimated. When the true underlying hazard function is continuous, as is the case when the dependent variable is wages, a continuous piecewise-linear hazard function will yield a better approximation than a step hazard function and, therefore, more accurate estimates should be obtained.

When modeling income-related variables, interest is often focused on the analysis of how equally the variable is distributed among the population. The most widespread statistical tools used to perform this kind of analysis are the Lorenz curve and the Gini index. The traditional approach for estimating them in the presence of covariates is to group the observations according to covariate values, and then construct a nonparametric Lorenz curve and a Gini index for each group. However, the methodology which we propose here allows us to obtain one estimated Lorenz curve and one estimated Gini index for each covariate value, with no need to group observations; thus, if the set of covariates includes any continuous variable, a much more precise analysis of inequality can be performed.

The methodology which we propose is applied here to analyze the conditional

distribution of wages in Spain between 1994 and 1999, a period in which several labor market reforms were approved, using data from the European Household Panel Data Survey (PHOGUE). With our procedure we obtain estimates of the entire conditional distribution or density functions, together with confidence bands. These flexible estimates reveal certain characteristics of these curves that would have remained hidden if a different methodology had been used, and allow us to analyze changes over time in returns to schooling, returns to experience and inequality. As we discuss below, the main changes in the labor market in the nineties involved graduate workers entering the labor market: our results provide evidence that the phenomenon labelled as “overeducation” (hiring of graduates for jobs that do not require a university degree) intensified in this period, resulting in a decrease in returns to moving from the middle level of education to the highest.

The rest of the paper is organized as follows. In Section 2 we describe the model, derive the conditional Lorenz curve and Gini index which stem from it, explain how to obtain estimators of the conditional distribution or density functions and their standard errors, and discuss the advantages and disadvantages of the methodology. In Section 3 we use Spanish data to estimate wage density and distribution functions conditional on education and experience, compute measures of returns to schooling, returns to experience and inequality and discuss the empirical results. Section 4 concludes.

## 2. METHODOLOGY

Although our interest is focused on estimating the conditional distribution of wages  $Y$  given certain covariates  $X$ , we describe our estimation procedure in a

general context. We first present the statistical features of the methodology and leave the discussion of its relationship with other procedures and its performance to the last two subsections.

## 2.1. Specification of the model

Let  $(X', Y)'$  be a random variable in  $\mathbb{R}^{K+1}$  such that  $Y$  is absolutely continuous with support  $[y_1, +\infty)$ , where  $y_1 \geq 0$ . For the reasons pointed out in the introduction, our starting point is the assumption that the conditional hazard function of  $Y$  given  $X = x$  is piecewise-linear. Specifically, we assume that it can be expressed as

$$h(y | x) = \left\{ \begin{array}{ll} \theta_j + (\theta_{j+1} - \theta_j) \frac{y - y_j}{y_{j+1} - y_j} & \text{if } y \in [y_j, y_{j+1}), \text{ for } j = 1, \dots, J, \\ \theta_{J+1} & \text{if } y \geq y_{J+1}, \end{array} \right\} \quad (1)$$

where  $[y_1, y_2), \dots, [y_J, y_{J+1}), [y_{J+1}, +\infty)$  are known intervals which are not allowed to depend on  $x$ , and  $\theta_1, \dots, \theta_{J+1}$  are positive values which may depend on  $x$ , though, for simplicity this dependence is not made explicit in the notation (this convention will also be followed with all quantities defined from  $\theta_j$  hereafter). The  $J + 1$  intervals  $[y_1, y_2), \dots, [y_J, y_{J+1}), [y_{J+1}, +\infty)$  will be referred to as “baseline intervals”. Following the literature on hazard functions, covariates are now introduced using a proportional-hazards model. However, to gain flexibility, the proportionality property is only assumed to hold within baseline intervals. Therefore, we assume that the coefficients  $\theta_j$  can be expressed as

$$\theta_j = \exp(\alpha_j + \beta_1^{(j)} x_1 + \dots + \beta_K^{(j)} x_K), \quad \text{for } j = 1, \dots, J + 1, \quad (2)$$

where  $x = (x_1, \dots, x_K)'$  is the observed vector of the covariates and  $\alpha_1, \dots, \alpha_{J+1}$ ,  $\beta^{(1)} \equiv (\beta_1^{(1)}, \dots, \beta_K^{(1)})', \dots, \beta^{(J+1)} \equiv (\beta_1^{(J+1)}, \dots, \beta_K^{(J+1)})'$  are unknown parameters.

Note that the exponential function in (2) ensures that  $\theta_j$  is positive. To prevent the total number of parameters from becoming too large, we include an additional restriction in the specification. As we discuss below, the number of baseline intervals might be very large. Hence, parameters  $\beta^{(1)}, \dots, \beta^{(J+1)}$  will not always be allowed to be all different. Instead, we assume that there exist integers  $J_1, \dots, J_P$ , with  $1 = J_1 < J_2 < \dots < J_P \leq J + 1$ , such that

$$\beta^{(1)} = \dots = \beta^{(J_2-1)}, \beta^{(J_2)} = \dots = \beta^{(J_3-1)}, \dots, \beta^{(J_P)} = \dots = \beta^{(J+1)}. \quad (3)$$

The  $P$  intervals  $I_1 \equiv [y_1, y_2) \cup \dots \cup [y_{J_2-1}, y_{J_2}), \dots, I_P \equiv [y_{J_P}, y_{J_P+1}) \cup \dots \cup [y_{J+1}, +\infty)$  will be referred to as ‘‘covariate intervals’’. Note that the restrictions in (3) imply that the beta coefficients must be the same within each covariate interval, though they can vary across covariate intervals. To sum up, our specification is determined by equations (1), (2) and (3), that is, we assume a continuous piecewise-linear baseline hazard function, and introduce covariates at each covariate interval with a proportional-hazards model. Thus, the total number of parameters in the specification is  $J + 1 + KP$ , and the parameter vector is  $(\alpha_1, \dots, \alpha_{J+1}, \beta^{(J_1)'}, \dots, \beta^{(J_P)'})'$ , which is hereafter denoted as  $\varphi$ .

All statistical properties of the conditional distribution of  $Y$  given  $X = x$  can be derived from (1). Using the relationship  $h(y | x) = f(y | x) / \{1 - F(y | x)\}$ , it follows that  $F(y | x) = 1 - \exp\{-\int_{y_1}^y h(t | x) dt\}$ , for  $y \geq y_1$ . Hence, if we denote  $y_{J+2} \equiv +\infty$ , the conditional distribution function at  $y$ , when  $y \in [y_j, y_{j+1})$ , is

$$F(y | x) = 1 - \exp\{-\delta_j - \theta_j(y - y_j) - \lambda_j(y - y_j)^2\}, \quad (4)$$

where  $\delta_1 \equiv 0$ ,  $\delta_j \equiv \sum_{l=1}^{j-1} (\theta_{l+1} + \theta_l)(y_{l+1} - y_l)/2$ , for  $j = 2, \dots, J + 1$ ,  $\lambda_j \equiv (\theta_{j+1} - \theta_j) / \{2(y_{j+1} - y_j)\}$  for  $j = 1, \dots, J$ , and  $\lambda_{J+1} \equiv 0$ . Thus, the conditional

density function at  $y \in [y_j, y_{j+1})$  is

$$f(y | x) = \{\theta_j + 2\lambda_j(y - y_j)\} \exp\{-\delta_j - \theta_j(y - y_j) - \lambda_j(y - y_j)^2\}. \quad (5)$$

Finally, to facilitate inequality analysis, we derive the expression for the conditional Lorenz curve and Gini index that stem from our specification. The Lorenz curve of a non-negative random variable with finite expectation, strictly increasing distribution function  $F(\cdot)$  and density function  $f(\cdot)$  is defined as  $L(u) = \int_0^{F^{-1}(u)} tf(t)dt / \int_0^\infty tf(t)dt$ , for  $u \in (0, 1)$ . In our context, observe that for any  $b$  in  $[y_j, y_{j+1})$ ,  $\int_0^b tf(t | x)dt = \sum_{l=1}^{j-1} \int_{y_l}^{y_{l+1}} tf(t | x)dt + \int_{y_j}^b tf(t | x)dt$ , where the first term in the right-hand member only appears if  $j > 1$ . These integrals can be computed using (5); thus we derive that

$$\int_0^b tf(t | x)dt = y_1 - b \exp\{-\delta_j - \theta_j(b - y_j) - \lambda_j(b - y_j)^2\} + H_j(b | x) + \psi_j,$$

where  $H_j(z | x) \equiv \int_{y_j}^z \exp\{-\delta_j - \theta_j(y - y_j) - \lambda_j(y - y_j)^2\}dy$ ,  $\psi_1 \equiv 0$  and  $\psi_j \equiv \sum_{l=1}^{j-1} H_l(y_{l+1})$ , for  $j = 2, \dots, J + 1$ . Similarly,

$$\int_0^\infty tf(t | x)dt = y_1 + \sum_{l=1}^{J+1} H_l(y_{l+1} | x).$$

Finally, observe that  $F^{-1}(u | x) \in [y_j, y_{j+1})$  is equivalent to  $u \in [u_j, u_{j+1})$ , where  $u_j \equiv 1 - \exp(-\delta_j)$ , for  $j = 1, \dots, J + 1$ , and  $u_{J+2} \equiv 1$ . Moreover, from (4) we derive that, for  $u \in [u_j, u_{j+1})$ ,

$$F^{-1}(u | x) = \begin{cases} y_j + \frac{1}{2\lambda_j}(-\theta_j + [\theta_j^2 - 4\lambda_j\{\delta_j + \ln(1 - u)\}]^{1/2}) & \text{if } \lambda_j \neq 0, \\ y_j - \{\delta_j + \ln(1 - u)\}/\theta_j & \text{if } \lambda_j = 0. \end{cases}$$

Putting together these results, it follows that the conditional Lorenz curve at a point  $u \in [u_j, u_{j+1})$  is

$$L(u | x) = \frac{y_1 - (1 - u)F^{-1}(u | x) + H_j\{F^{-1}(u | x) | x\} + \psi_j}{y_1 + \sum_{l=1}^{J+1} H_l(y_{l+1} | x)}. \quad (6)$$



The Gini index  $G$  corresponding to a Lorenz curve  $L(\cdot)$  is  $G = 1 - 2 \int_0^1 L(u) du$ . In our case, after some algebraic manipulations, it follows from (6) that the conditional Gini index is

$$G(x) = 1 - \frac{y_1 + \sum_{j=1}^{J+1} H_j^{(2)}(y_{j+1} | x)}{y_1 + \sum_{j=1}^{J+1} H_j(y_{j+1} | x)}, \quad (7)$$

where  $H_j^{(2)}(z | x) \equiv \int_{y_j}^z \exp\{-2\delta_j - 2\theta_j(y - y_j) - 2\lambda_j(y - y_j)^2\} dy$ . For computational purposes, note that, if  $\lambda_j \geq 0$ , the integrals which appear in  $H_j(\cdot | x)$  and  $H_j^{(2)}(\cdot | x)$  can easily be computed using the exponential function or the standard normal distribution function; otherwise, these integrals can be approximated numerically.

## 2.2. Inference

Given a sample  $\{(X'_i, Y_i)'\}_{i=1}^n$  of independent and identically distributed observations, the parameters of the model can be estimated by maximum likelihood. From (5) it follows that the log-likelihood function can be expressed as

$$\begin{aligned} \ln L(\varphi) = & \sum_{i=1}^n \sum_{j=1}^{J+1} \{ \mathbf{I}(y_j \leq Y_i < y_{j+1}) [\ln \{ \theta_{i,j} + 2\lambda_{i,j}(Y_i - y_j) \} \\ & - \delta_{i,j} - \theta_{i,j}(Y_i - y_j) - \lambda_{i,j}(Y_i - y_j)^2] \}, \end{aligned}$$

where  $\mathbf{I}(\cdot)$  is the indicator function, and  $\theta_{i,j}$ ,  $\delta_{i,j}$ ,  $\lambda_{i,j}$  are defined as  $\theta_j$ ,  $\delta_j$ ,  $\lambda_j$ , but replacing  $x$  by  $X_i$ . Note that we add the subscript  $i$  in these values to emphasize that not only do they depend on the vector parameter, but also on the regressors. Maximization of  $\ln L(\varphi)$  yields a root- $n$ -consistent estimate of  $\varphi$ , denoted as  $\widehat{\varphi}$ . Thus, given any covariate vector  $x$ , we can define  $\widehat{\theta}_j \equiv \exp(\widehat{\alpha}_j + x' \widehat{\beta}^{(j)})$ , and  $\widehat{\delta}_j$ ,  $\widehat{\lambda}_j$  in the same way as  $\delta_j$ ,  $\lambda_j$ , replacing  $\theta_j$  by  $\widehat{\theta}_j$ . From here it is straightforward to obtain root- $n$ -consistent estimates of  $F(y | x)$  and  $f(y | x)$ , simply replacing  $\theta_j$ ,  $\delta_j$  and  $\lambda_j$  by  $\widehat{\theta}_j$ ,  $\widehat{\delta}_j$  and  $\widehat{\lambda}_j$  in (4) and (5), respectively. By maximum-likelihood

techniques we can also compute a consistent estimate  $\widehat{V}$  of the variance-covariance matrix of  $\widehat{\varphi}$ , so that the asymptotic distribution of  $\widehat{\varphi}$  is normal  $N(\varphi, \widehat{V})$ . Therefore, consistent standard errors of the estimates of  $F(y | x)$  and  $f(y | x)$  can be obtained with the delta method as follows: first note that

$$\frac{\partial \theta_j}{\partial \varphi_s} = \begin{cases} \theta_j \mathbf{I}(s = j) & \text{if } s \leq J + 1, \\ \theta_j x_{r(s)+1} \mathbf{I}\{q(s) = p(j) - 1\} & \text{if } s > J + 1, \end{cases} \quad (8)$$

where  $q(s)$  and  $r(s)$  are defined as the quotient and remainder after dividing  $s - (J + 1) - 1$  by  $K$  and, for a given  $j$ ,  $p(j)$  denotes the integer in  $\{1, \dots, P\}$  such that  $[y_j, y_{j+1}) \subset I_p$ . Using (8),  $\partial \delta_j / \partial \varphi_s$  and  $\partial \lambda_j / \partial \varphi_s$  are readily derived and then, given  $y \in [y_j, y_{j+1})$ , we can also obtain

$$\frac{\partial F(y | x)}{\partial \varphi_s} = \left\{ \frac{\partial \delta_j}{\partial \varphi_s} + (y - y_j) \frac{\partial \theta_j}{\partial \varphi_s} + (y - y_j)^2 \frac{\partial \lambda_j}{\partial \varphi_s} \right\} \{1 - F(y | x)\};$$

finally, if  $\widehat{\Gamma}$  denotes the  $1 \times (J + 1 + KP)$  matrix whose  $s$ -th element is  $\partial F(y | x) / \partial \varphi_s$ , but replacing  $F(y | x)$ ,  $\theta_j$ ,  $\delta_j$  and  $\lambda_j$  by their estimates, then the asymptotic distribution of  $\widehat{F}(y | x)$  is normal  $N(F(y | x), \widehat{\Gamma} \widehat{V} \widehat{\Gamma}')$ . Similar reasoning also applies for  $f(y | x)$ . However, the delta method does not apply for  $L(u | x)$  and  $G(x)$ , because these quantities are not differentiable functions of  $\varphi$ , since the values  $u_j$  depend on  $\delta_j$ .

An alternative method for computing consistent standard errors is bootstrap resampling. From the expression for  $F^{-1}(\cdot | x)$  derived in the previous subsection, it is concluded that a bootstrap sample can be obtained as follows: define  $\widehat{\theta}_{i,j} \equiv \exp(\widehat{\alpha}_j + X_i' \widehat{\beta}^{(j)})$ ; define  $\widehat{\delta}_{i,j}$ ,  $\widehat{\lambda}_{i,j}$  in the same way as  $\delta_j$ ,  $\lambda_j$  but replacing  $\theta_j$  by  $\widehat{\theta}_{i,j}$ ; define  $\widehat{u}_{i,j}$  in the same way as  $u_j$  but replacing  $\delta_j$  by  $\widehat{\delta}_{i,j}$ ; generate a random number from a uniform distribution in  $(0, 1)$ , say  $U_i^*$ , and let  $j^*$  be the number in  $\{1, \dots, J + 1\}$  such that  $U_i^* \in [u_{i,j^*}, u_{i,j^*+1})$ ; finally, define  $X_i^* = X_i$

and  $Y_i^* = y_{j^*} + ([\widehat{\theta}_{i,j}^2 - 4\widehat{\lambda}_{i,j}\{\widehat{\delta}_{i,j} + \ln(1 - U_i^*)\}]^{1/2} - \widehat{\theta}_{i,j})/(2\widehat{\lambda}_{i,j})$  if  $\widehat{\lambda}_{i,j} \neq 0$ , or  $y_{j^*} - \{\widehat{\delta}_{i,j} + \ln(1 - U_i^*)\}/\widehat{\theta}_{i,j}$  if  $\widehat{\lambda}_{i,j} = 0$ . Given this bootstrap sample  $\{(X_i^{*l}, Y_i^*)'_{i=1}^n$ , a consistent standard error of  $F(y | x)$  can then be computed as follows: i) generate a bootstrap sample, obtain the maximum-likelihood estimate for that bootstrap sample  $\widehat{\varphi}^*$ , and compute with it  $\widehat{\theta}_j^* \equiv \exp(\widehat{\alpha}_j^* + x'\widehat{\beta}^{(j)*})$ ,  $\widehat{\delta}_j^*$ ,  $\widehat{\lambda}_j^*$  and the corresponding bootstrap estimate  $\widehat{F}(y | x)^*$ ; ii) repeat the previous step  $B$  times to obtain  $B$  bootstrap estimates; the sample standard deviation of the  $B$  bootstrap estimates provides a bootstrap approximation of the standard error of  $\widehat{F}(y | x)$ . This bootstrap procedure also applies for obtaining bootstrap approximations of the standard errors of  $f(y | x)$ ,  $L(u | x)$  and  $G(x)$ .

### 2.3. Discussion of the methodology

In principle, our methodology should be considered as purely parametric, because if the number of baseline intervals is treated as fixed, our specification leads to a conditional distribution function which is completely known except for a finite set of parameters. However, if the number of baseline intervals is large enough, the assumption that the hazard function is piecewise-linear is, in essence, a nonparametric assumption, since any underlying hazard function can be approximated in this way. On the other hand, covariates are introduced parametrically, but the parameters which determine their influence are partly allowed to vary along the support of the dependent variable, so that the parametric component of the model gains flexibility. Thus, our specification as a whole can be considered as semiparametric, but flexible enough to capture almost any possible underlying relationship. One should keep in mind, however, that the good asymptotic behavior of the maximum-likelihood estimators that are proposed here is ensured only for

a fixed number of baseline intervals; this means that, in practice,  $n$  must be large in relation to  $J$ .

This discussion also shows that the key point of the methodology is the choice of baseline and covariate intervals. Baseline intervals play the same role here as bins in a histogram: too many will lead to a very wiggly estimate, whereas too few may lead to an excessively flat estimate that masks the underlying shape. DGP describe a rule-of-thumb for choosing the number of baseline intervals, though they suggest that a simple graphical inspection of the results may well be of great help in this choice. The importance of the number of covariate intervals depends on how sensitive conditional distributions are to changes in the covariates.

As our methodology yields a fully parameterized conditional distribution function, we can check the specification of the model using any of the recently developed nonparametric specification tests that are consistent versus any alternative, see e.g. Andrews (1997), Zheng (2000) or Bai (2003). Additionally, the p-values obtained with any of these tests for different values of  $J$  and  $P$  can help us to decide how many baseline and covariate intervals must be used. From a theoretical point of view, this may well be a much more satisfactory way to choose  $J$  and  $P$  than the rule-of-thumb proposed in DGP or the simple graphical inspection of the estimates, but the computational burden of the aforementioned specification tests makes their use much less appealing in practice.

#### **2.4. Comparison with other procedures**

Let us analyze first the differences between our methodology and the one considered in DGP. With our notation, their estimation procedure amounts to assuming that: i)  $h(y | x) = \theta_j$  when  $y \in [y_j, y_{j+1})$ ; and ii) equation (2) holds with the

reparameterization  $\alpha_j = \gamma_j - \ln(y_{j+1} - y_j)$ . Additionally, DGP discretize the dependent variable, considering all observations within the same baseline interval as equivalent. Hence, to derive their likelihood function, which is equation (2.5) in their paper, they only have to compute the probability that the dependent variable falls within each baseline interval, say  $P(y_j | x)$ . Note that their approach does not allow their parameter  $\gamma_{J+1}$  to be estimated. Finally, given  $y \in [y_j, y_{j+1})$  for  $j = 1, \dots, J$ , DGP propose estimating the conditional density  $f(y | x)$  with the histogram-like estimator  $\tilde{f}(y | x) \equiv \tilde{P}(y_j | x)/(y_{j+1} - y_j)$ , where  $\tilde{P}(y_j | x)$  is defined as  $P(y_j | x)$ , but replacing unknown parameters by maximum-likelihood estimates. Thus, we can summarize the main advantages of our methodology with respect to the DGP procedure as follows: i) a continuous conditional hazard function is assumed, which should lead to a better fit when the true hazard function is continuous, as is the case in most applications; ii) no discretization is performed and, thus, there is no loss of information when constructing the likelihood function; iii) our methodology provides a continuous estimate of the conditional density function, whereas the DGP procedure only provides a histogram-like estimate, which may be less appealing in most contexts; and iv) in the DGP procedure, the discontinuities of the step hazard function lead to excessively spiky density estimates, even after smoothing; this undesirable property is lessened when a continuous piecewise-linear hazard function is assumed, as is proven by the graphs that we report in the next subsection.

Many other methods have been developed to estimate entire conditional distribution or density functions. In comparison with purely parametric methods, the main advantage of hazard-based estimators is their flexibility, since they do not impose any prior functional form. In comparison with nonparametric methods,

observe that the latter are extremely inefficient if many covariates are included, and this is not the case with hazard-based estimators since they are derived by maximizing a likelihood function. On the other hand, an alternative widespread procedure that is close in spirit to the one we propose here is quantile regression, see e.g. Koenker and Hallock (2001). This procedure also yields flexible estimators of conditional distributions under relatively mild assumptions. However, the following characteristics of our methodology might make it more appealing than quantile regression for a practitioner: i) our procedure eventually leads to a fully parameterized conditional distribution function; this allows us to derive a conditional Lorenz curve and a conditional Gini index for any covariate vector with no loss of information; and this also allows us to check the appropriateness of the procedure using nonparametric specification tests; ii) in our procedure, the number of beta parameters may be large if the number of observations is large enough; thus, as the sample size grows our procedure provides greater flexibility in the parametric component than quantile regression; iii) from a computational point of view, our procedure is much easier to implement since it only requires us to solve one optimization problem, with no restrictions, to obtain continuous estimates of the conditional distribution and density functions; with quantile regression, one has to estimate a large number of quantiles to derive an accurate estimate of the conditional distribution and density functions, and many restrictions have to be imposed to ensure that, for any covariate vector, the  $p_1$ -th quantile is not larger than the  $p_2$ -th quantile if  $p_2 > p_1$  (if these restrictions were not imposed, the resulting estimate of the conditional distribution function might not be a distribution function).

## 2.5. Monte Carlo evidence on the performance of the estimator

To provide evidence on how well the estimator behaves in practice, and also to shed light on the extent to which our methodology is an improvement on the DGP procedure, we perform a Monte Carlo experiment, similar in spirit to the one described in Section 2.4 of DGP<sup>1</sup>. We generate  $n = 3000$  observations from a standard normal distribution, say  $\{U_i\}_{i=1}^n$ ;  $n$  observations from a uniform  $(0, 1)$  distribution, say  $\{X_i\}_{i=1}^n$ ; and  $n$  observations from a Bernoulli distribution with  $p = 0.5$ , say  $\{Z_i\}_{i=1}^n$ . All observations are independent from one another. Define

$$Y_i = \begin{cases} \exp(0.5 + 0.1U_i) & \text{if } Z_i = 1, \\ \exp(0.5 + 0.5X_i + 0.1U_i) & \text{if } Z_i = 0. \end{cases}$$

Then, the conditional distribution  $Y \mid X = x$  is a mixture between two lognormal distributions. The conditional density is unimodal if  $x$  is close to 0, but bimodal if  $x$  is close to 1. With these artificial data, we can check the ability of the estimates to detect meaningful changes in the conditional densities induced by changes in the covariates. We generate 50 samples of data  $\{(X'_i, Y_i)'\}_{i=1}^n$  and, with each sample, we estimate the conditional densities at  $x = 0.7$  and  $x = 0.2$  using both the estimate based on a step hazard-function and the estimate based on a piecewise-linear hazard-function, for various  $J$  and  $P$ . To compute the estimates, at the tails we choose baseline intervals with left endpoints  $y_1 = 1$ ,  $y_2 = q_{0.01}$ ,  $y_3 = q_{0.02}$ ,  $y_4 = q_{0.03}$ ,  $y_{J-1} = q_{0.97}$ ,  $y_J = q_{0.98}$  and  $y_{J+1} = q_{0.99}$ , where  $q_p$  denotes the  $p$ -th percentile of the unconditional distribution of  $Y$ ; the remaining  $J - 6$

---

<sup>1</sup>We do not replicate their experiment exactly because their description of the artificial data contains several mistakes; e.g., the 99-th percentile of the unconditional distribution of  $Y$  is much greater than 2.80. Also observe that the curves in their Figures 1-3 are not densities, since they do not integrate to one.

left endpoints of baseline intervals are equally spaced between  $y_4$  and  $y_{J-1}$ . On the other hand, covariate intervals are chosen uniformly among baseline intervals, approximately; e.g., when  $J = 20$  and  $P = 5$ , covariate intervals are defined taking  $J_i = 1 + 4(i - 1)$ , for  $i = 1, \dots, 5$ . We always take  $K = 2$  and consider as vector of covariates  $(X_i, X_i^2)'$ . To evaluate the goodness of the fit, for each sample we compute the difference between the fitted and true conditional densities at 156 equally-spaced points running from 1.35 to 2.90, and then the root-mean-squared-error of the sample. Finally, we average the 50 root-mean-squared-errors to obtain an overall measure of the goodness of the fit obtained with each procedure, for given  $J$  and  $P$ . The results are reported in Table 1.

The results in Table 1 indicate that our methodology leads to a substantial improvement in terms of mean-squared-error. It is also observed that, as expected, increasing the number of baseline or covariate intervals improves the fit of the estimates only up to a point; in this case the  $J = 15$ ,  $P = 5$  specification seems to be the preferred one. Also observe that  $J$  plays a more crucial role than  $P$ , as long as enough covariate intervals are included -note that specifications with  $P = 1$  do not yield satisfactory results, but very similar results are obtained with  $P = 5$  and  $P = 10$ . For a visual depiction of how the procedure performs and the effect of  $J$  on the estimates, in Figures 1 and 2 we plot the true conditional densities at  $x = 0.7$  and  $x = 0.2$ , respectively, with “typical” estimates obtained with the specifications  $J = 15$ ,  $P = 5$  and  $J = 25$ ,  $P = 5$ .

### **3. WAGE DISTRIBUTION IN SPAIN BETWEEN 1994 AND 1999**

In this section we apply our methodology to analyze the conditional distribution of wages in Spain between 1994 and 1999, a period in which several labor



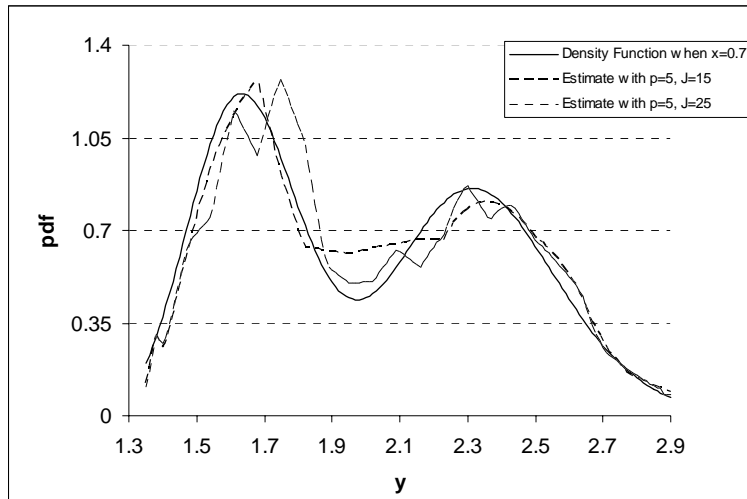
**TABLE 1:**

**Average Root-Mean-Squared-Error of Conditional Density Estimates  
Based on the Hazard Function (HF)**

Conditional Density Function at $x=0.7$						
	Estimate with step HF			Estimate with piecewise-linear HF		
	P=1	P=5	P=10	P=1	P=5	P=10
J=10	0.2471	0.1812	0.1792	0.2347	0.1743	0.1324
J=15	0.1895	0.1318	0.1345	0.2141	0.0883	0.0976
J=20	0.1990	0.1378	0.1456	0.1723	0.0902	0.0971
J=25	0.1934	0.1390	0.1435	0.1711	0.1087	0.1124
J=30	0.1908	0.1394	0.1561	0.1798	0.1203	0.1298
Conditional Density Function at $x=0.2$						
	Estimate with step HF			Estimate with piecewise-linear HF		
	P=1	P=5	P=10	P=1	P=5	P=10
J=10	0.3245	0.2797	0.2561	0.2457	0.1829	0.1672
J=15	0.2456	0.1992	0.1980	0.1934	0.0842	0.0994
J=20	0.2109	0.1765	0.1789	0.1852	0.1443	0.1268
J=25	0.2054	0.1716	0.1731	0.1843	0.1499	0.1354
J=30	0.2087	0.1704	0.1728	0.1803	0.1458	0.1562

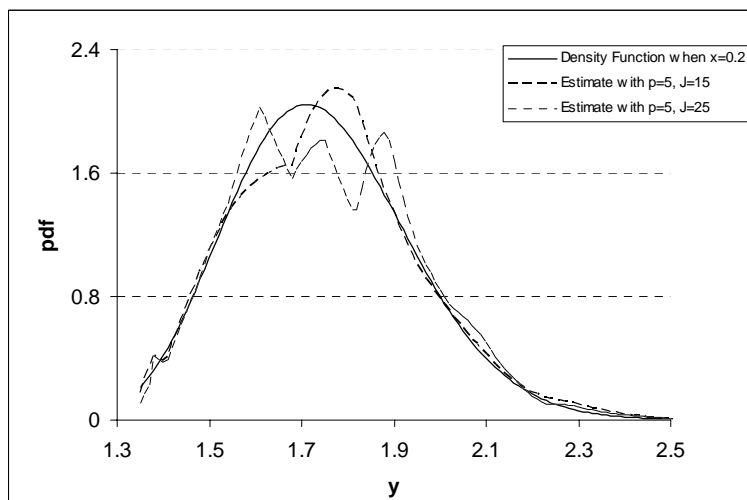
**FIGURE 1:**

**Conditional Density Function at  $x = 0.7$  with Estimates Based on a Piecewise-Linear Hazard Function**



**FIGURE 2:**

**Conditional Density Function at  $x = 0.2$  with Estimates Based on a Piecewise-Linear Hazard Function**



market reforms were approved. In the early eighties, the Spanish labor market was characterized by strong rigidities, which were partly alleviated with the 1984 reform. In the nineties, two major reforms took place in 1994 and 1997. The main changes introduced with these reforms were focused on: i) increasing the topics under control in collective bargaining; ii) decreasing the possibilities of short-term hiring and introducing another kind of unlimited contract with lower dismissal costs; iii) extending the possibilities of individual or collective dismissals for objective causes; iv) introducing more flexibility on part-time contracts; v) introducing incentives for permanent contracts; and vi) creating firms for temporary work.

The main objective of our empirical analysis is to examine how these reforms affected wage distribution, paying especial attention to changes over time in returns to schooling, returns to experience and inequality. Previous empirical analyses on the influence of these reforms in Spanish labor market were limited by the lack of availability of representative samples for the whole period. Abadie (1997) analyzed how the distribution of labor income in Spain was affected by the process of liberalization that took place during the eighties, when Spain became a member of the European Community; using quantile regression and data from the Spanish Expenditure Survey for 1980/81 and 1990/91, he concluded that returns to schooling declined sharply in Spain during the eighties, in contrast to what had been detected in the USA. He also observed that income dispersion decreased remarkably within each education level. Alba-Ramírez and San Segundo (1995) analyzed returns to education using least-squares and data from the second quarter of the 1990 Spanish labor force survey, and obtained that an additional year of education yields approximately an 8.5% increase in earnings, though this average

figure varies substantially when distinguishing by class of worker (men/women, private sector/public sector, and so on). More recently, Del Río and Ruiz-Castillo (2001) analyze the evolution of labor income in Spain with an innovative methodology. They conclude that income inequality has dropped continuously since 1973, and that returns to schooling showed a decreasing trend in the eighties and early nineties.

In this section our aim is to determine how all these observed characteristics changed in the second half of the nineties, using the methodology described in the previous section and data from the European Household Panel Data Survey (PHOGUE), which compiles information on wages and demographic characteristics for a wide range of individuals and households from 1994 to 1999. This sample contains a large set of individuals (over 17000 in Spain), with information about income sources and demographic variables (for a description of the database see e.g. Andrés and Mercader-Prats, 2001). To avoid problems with sample selection, here we use an extract that contains all males who were employed in the private or the public sector. Since we want to examine what the Spanish labor market paid for education and experience, with these data we estimate the conditional distribution and density functions of gross wages, with the level of education and the years of experience as covariates. As a dependent variable we consider “real gross hourly wage”, obtained by dividing nominal-net-monthly-wage by four times weekly-hours-worked, and deflating the result by the 1992-based Spanish Consumer Price Index. We consider three covariates: a variable for years of experience (defined as the difference between current age and the age at which the individual started his working life) and two dummy variables to pick up the level of education (one for individuals who finished high-school and another for individ-

uals who completed a university degree). To simplify the presentation of results, we report only estimations corresponding to 1994 and 1999. Mean wages for the whole samples and for various subsamples are presented in Table 2.

We estimate the conditional distribution of wages given these three covariates using the methodology described in Section 2. Taking into account the formula given by Scott (1979) to select the number of bins in histogram estimation, the graphical depiction of some preliminary estimates, and the results of our Monte Carlo experiments, we opt for  $J + 1 = 24$  baseline intervals with left endpoints  $y_1 = 1$ ,  $y_2 = q_{0.025}$ ,  $y_j = q_{0.05(j-2)}$  for  $j = 3, \dots, 19$ ,  $y_{20} = q_{0.88}$ ,  $y_{21} = q_{0.91}$ ,  $y_{22} = q_{0.94}$ ,  $y_{23} = q_{0.96}$  and  $y_{24} = q_{0.98}$ , where  $q_p$  denotes the  $p$ -th sample percentile of the dependent variable. Note that we choose comparatively more baseline intervals for the highest wages; in this way, we try to prevent an excess of smoothing from masking relevant characteristics of the conditional distributions at the upper tail, where important information is contained, especially when conditioning on individuals with a university degree. Finally, we consider  $P = 4$  covariate intervals, constructed using  $J_i = 1 + 6(i - 1)$ , for  $i = 1, \dots, 4$ . All the results reported below are based on these specifications.

Before proceeding to discuss the results of our estimations, it is worth emphasizing that the problem of endogeneity, which typically arises when estimating functional relationships between wages and education in parametric models, does not appear with our approach. In a parametric context, the problem of endogeneity may lead to inconsistent estimates of the regression coefficients and, thus, the subsequent estimates of the conditional distributions are invalid. In our procedure, however, the beta parameters are introduced simply to obtain a flexible enough specification; thus, as long as baseline and covariate intervals are appro-

**TABLE 2:**  
**Sample Mean Wages for Spanish Male Workers**

		1994		1999	
		Sample size	Mean wage	Sample size	Mean wage
Whole sample		3461	4.797	2702	5.140
Education	Less than high school	2247	3.931	1662	4.292
	High school	686	5.056	589	5.274
	University	528	8.144	451	8.087
Experience	$\leq 1$ year	117	2.767	179	3.326
	$>1 \leq 10$ years	774	3.916	614	4.167
	$>10 \leq 20$ years	856	4.943	728	5.106
	$>20 \leq 40$ years	1360	5.375	997	6.067
	$>40$ years	354	4.824	184	5.260

priately chosen, the beta parameters lead to accurate estimates of the underlying conditional distributions, irrespective of whether there is endogeneity or not.

### **3.1. Conditional Wage Densities and Distributions**

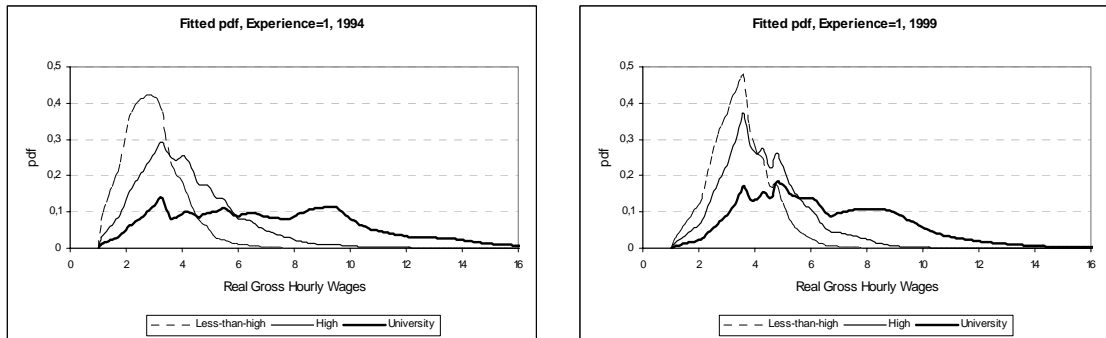
In Figures 3, 4 and 5 we plot the estimated probability density functions (pdf) conditional on the three levels of education (university/high school/less-than-high school) and three possible situations for years of experience (1, 20 and 40 years of experience). With these levels of experience, we try to summarize the beginning, the middle and the end of the working lives of individuals. In each case we report estimates for both 1994 and 1999.

Figures 3-5 show that important differences arise between university-educated workers and workers with a lower level of education. The pdf for workers with a university degree who enter the labor market (1 year of experience) has a shape with no clear main mode. Instead, we observe a flat shape for a wide range of real wages. As expected, the pdf's for the other two levels of education display a shape with a clear mode and a long right tail. Comparing the plots in Figures 4 and 5 with those in Figure 3, we observe that as experience increases the differences between the shapes of all pdf's decrease though, as expected, mean and variance grow with experience. Thus, the covariate "years of experience" seems to have an important effect on the shape of the pdf only in the first years of working life.

Comparing the fitted densities for 1994 and 1999, at first sight we only observe meaningful differences for workers with 1 year of experience. To further explore this issue, in Figures 6, 7 and 8 we plot the estimated conditional cumulative distribution functions (cdf) for workers with 1 year of experience, with 95 percent confidence bands.

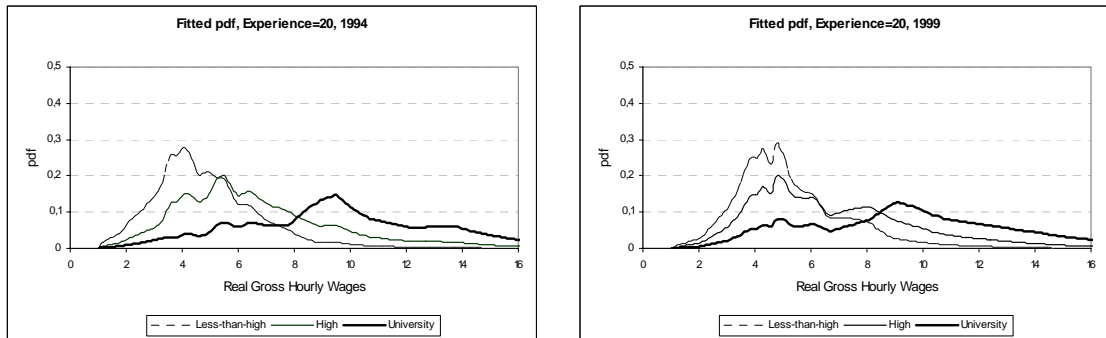
**FIGURE 3:**

**Fitted Densities Conditional on Experience = 1 year and Education;  
1994 (left) and 1999 (right)**



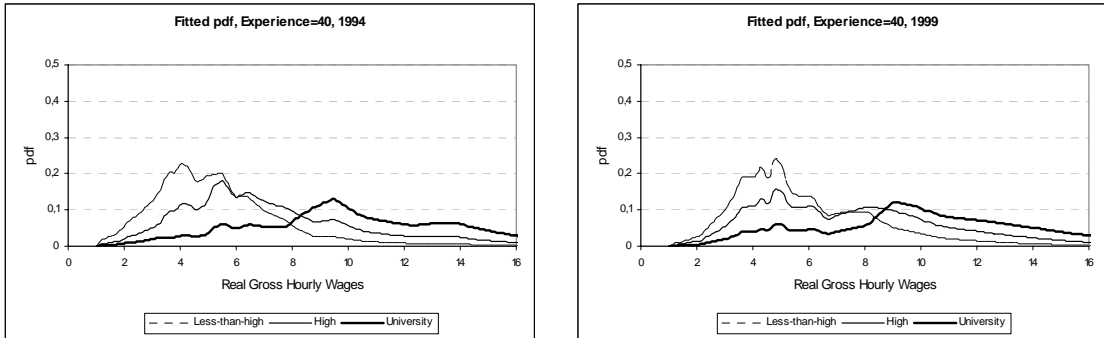
**FIGURE 4:**

**Fitted Densities Conditional on Experience = 20 years and  
Education; 1994 (left) and 1999 (right)**

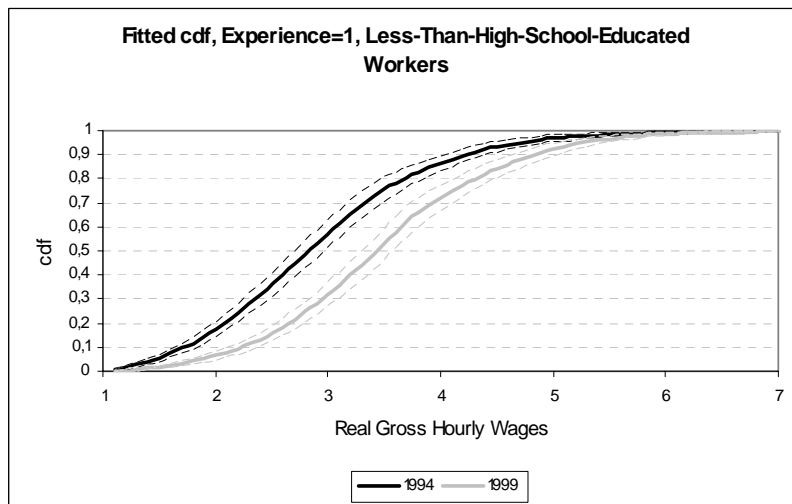




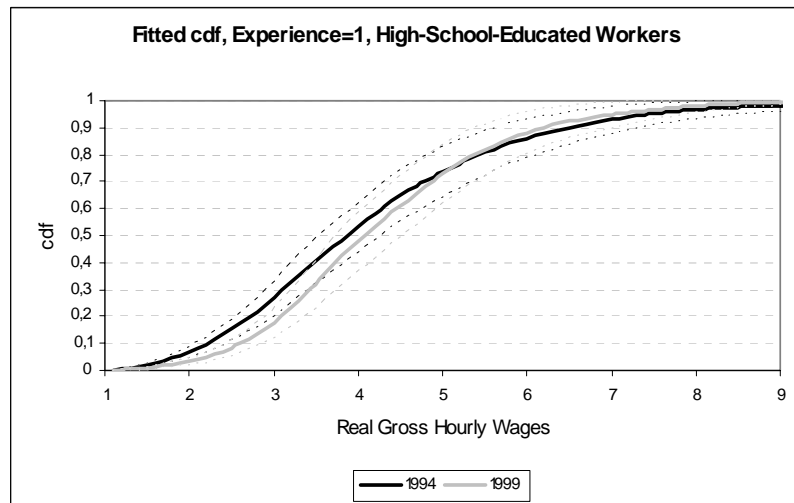
**FIGURE 5:**  
**Fitted Densities Conditional on Experience = 40 years and**  
**Education; 1994 (left) and 1999 (right)**



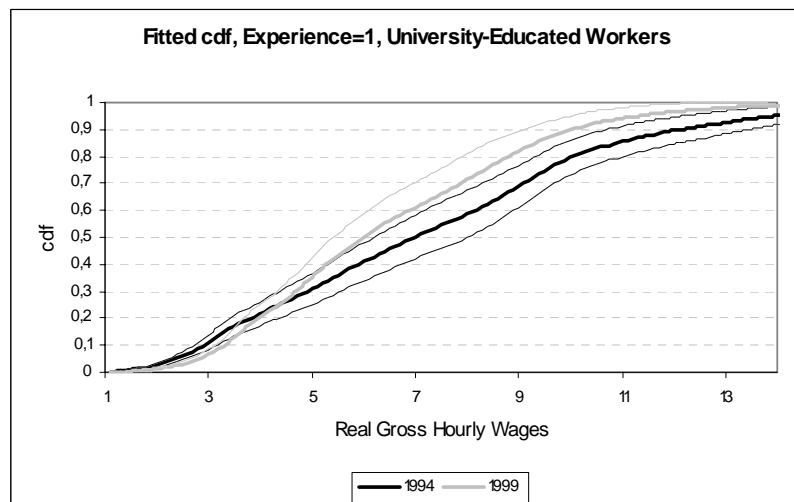
**FIGURE 6:**  
**Fitted Distributions Conditional on Experience = 1 year, and Level of**  
**Education = Less-than-High School**



**FIGURE 7:**  
**Fitted Distributions Conditional on Experience = 1 year, and Level of**  
**Education = High School**



**FIGURE 8:**  
**Fitted Distributions Conditional on Experience = 1 year, and Level of**  
**Education = University**



In these figures we observe that workers with the lowest level of education (less-than-high school) improve their real wages from 1994 to 1999, and this improvement affects all wages proportionally. However, for workers who completed high school the improvement only proves to be significant for low wages (note that the 1999 cdf falls below the confidence band for the 1994 cdf for workers in the lower three deciles). Finally, for workers with a university degree there is no improvement at all; in fact, there is a significant worsening for workers with high wages (now the 1999 cdf falls above the confidence band for the 1994 for workers in the upper four deciles). We have also estimated the cdf's for workers with 20 and 40 years of experience, but we do not report the results here, since in these cases no significant differences are detected between 1994 and 1999.

### 3.2. Returns to Schooling

With the method we propose in this paper we can estimate the whole distribution function of wages conditional on education and experience. In this way, using the inverse of this CDF, we can define a measure of the returns to education that does not impose linearity on schooling and can vary across the wage distribution. Observe that, for  $p \in (0, 1)$ ,  $F^{-1}(p|\text{educ.} = i, \text{exp.} = x) - F^{-1}(p|\text{educ.} = i - 1, \text{exp.} = x)$  represents the wage increase which the  $p$ -th worker with  $x$  years of experience would obtain if his/her level of education changed from  $i - 1$  to  $i$ , where the “ $p$ -th worker” is defined in terms of the ordering induced by wages. Hence, the relative wage increase which this worker would obtain is

$$\frac{F^{-1}(p|\text{educ.} = i, \text{exp.} = x) - F^{-1}(p|\text{educ.} = i - 1, \text{exp.} = x)}{F^{-1}(p|\text{educ.} = i - 1, \text{exp.} = x)}.$$

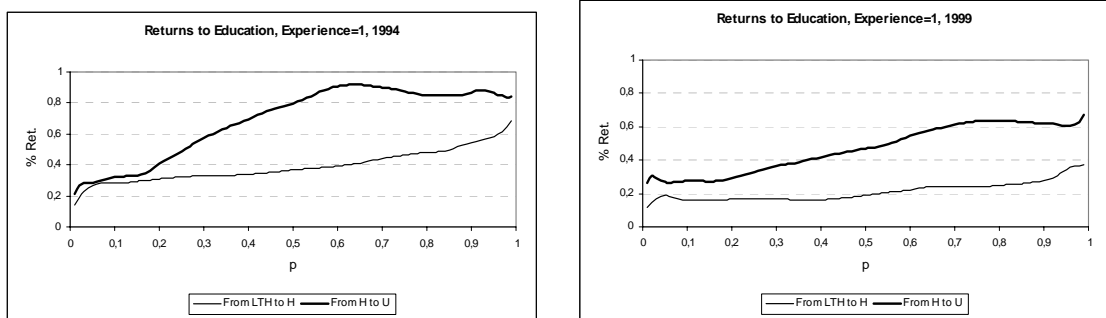
This quantity provides a measure of the wage incentive to higher schooling for the  $p$ -th worker, and can be readily estimated from an estimate of the conditional distribution function. Observe that with this definition the term “returns to schooling” simply refers to the effect of education on the conditional distribution of real wages. We do not address identification issues here; of course, this does not mean that we ignore the causal effect interpretation introduced in Heckman and Robb (1985), but this question lies outside the scope of this study. In Figures 9, 10 and 11 we plot the estimates of these quantities for  $x = 1, 20$  and 40 years of experience. Each figure contains four curves: two corresponding to the relative increase for changing from the less-than-high school level (LTH) to the high school level (H) in 1994 and 1999, and two corresponding to the relative increase for changing from the high school level (H) to the university level (U).

From Figures 9, 10 and 11 we deduce that moving from high-school to university gives more profits than moving from less-than-high-school to high-school for any value of years of experience, in both 1994 and 1999. Since the number of years of schooling for both movements is the same, we can conclude that the labor market values university time investment more highly than that spent in high school.

Since most people invest in human capital before entering the labor market, Figure 9 (entrants to the labor market) is the most important one. In this figure we also observe dissimilarities among the percentiles of the distribution: returns to education are higher at the upper quantiles. As experience grows, the returns to education become more stable through the entire distribution, but major important differences between moving from LTH to H and from H to U still arise. When comparing 1994 and 1999, we observe that there is a strong decrease in the returns to education, especially for workers with 1 year of experience, i.e.,

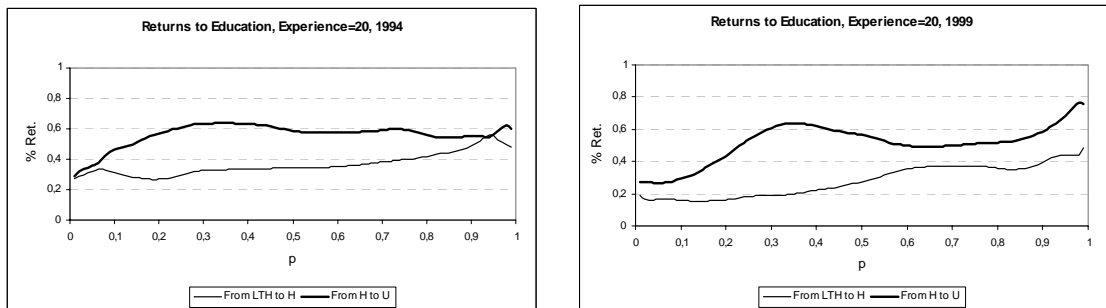
**FIGURE 9:**

**Relative Increase in Wage of the  $p$ -th Worker with Changes in Education Level (Workers with Experience = 1 year)**



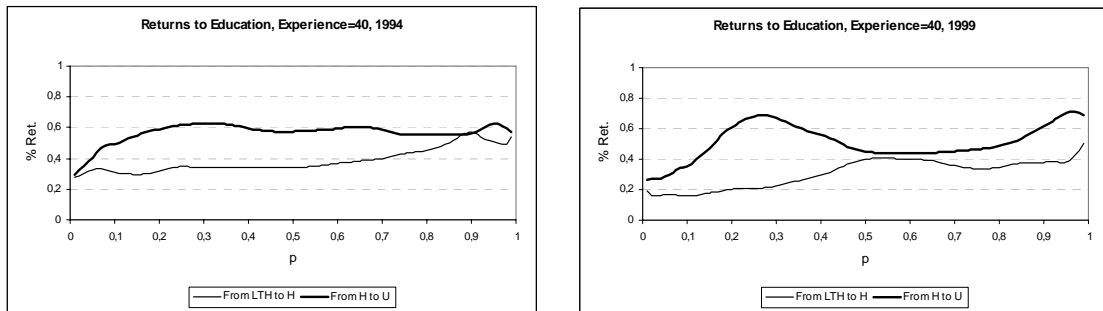
**FIGURE 10:**

**Relative Increase in Wage of the  $p$ -th Worker with Changes in Education Level (Workers with Experience = 20 years)**



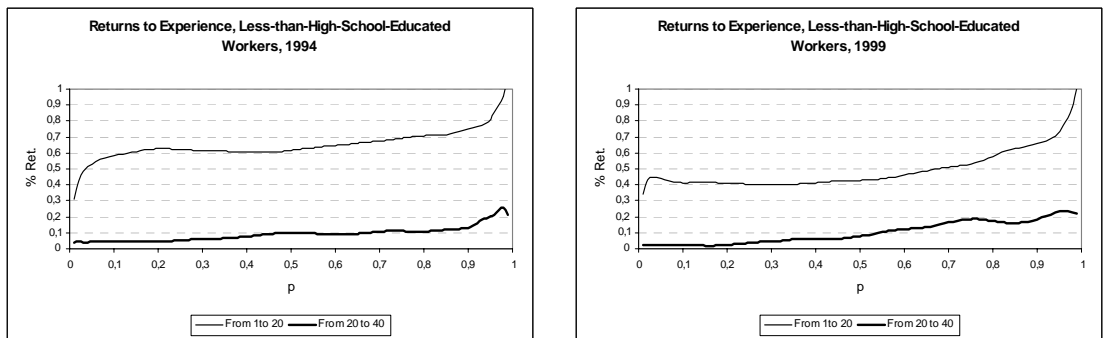
**FIGURE 11:**

**Relative Increase in Wage of the  $p$ -th Worker with Changes in Education Level (Workers with Experience = 40 years)**



**FIGURE 12:**

**Relative Increase in Wage of the  $p$ -th Worker with Changes in Years of Experience (Workers with Education = Less-Than-High School)**



incentives to higher education decrease substantially in 1999 with respect to 1994; a possible explanation for this is that the hiring of graduates for jobs that do not require a university degree (“overeducation”) intensified in this period. Finally, comparing returns to education by years of experience, we observe that moving from LTH to H yields very similar returns at all levels of experience; moving from H to U at low levels of experience yields greater returns in 1994 than in 1999. In general, as experience increases less differences in the returns to schooling in 1994 and 1999 are observed.

### 3.3 Returns to Experience

As in the previous subsection, we can define a measure of the returns to moving from  $x_1$  to  $x_2$  years of experience for the  $p$ -th worker at the  $i$ -th level of education as follows:

$$\frac{F^{-1}(p|\text{educ.} = i, \text{exp.} = x_2) - F^{-1}(p|\text{educ.} = i, \text{exp.} = x_1)}{F^{-1}(p|\text{educ.} = i, \text{exp.} = x_1)}.$$

In Figures 12, 13 and 14 we plot the returns when moving from 1 to 20 years of experience and from 20 to 40 years of experience, at all three levels of education.

As expected, these figures show that there are huge differences between the returns to moving from 1 to 20 years of experience and the returns to moving from 20 to 40 years: the former are much larger and less homogeneous. The returns to moving from 20 to 40 years show an increasing pattern with the percentiles of the distribution; they are very similar for all levels of education and no significant changes are detected between 1994 and 1999. The returns to moving from 1 to 20 years of experience show similar patterns for workers at the two levels of education; when comparing them between 1994 and 1999 we observe that these

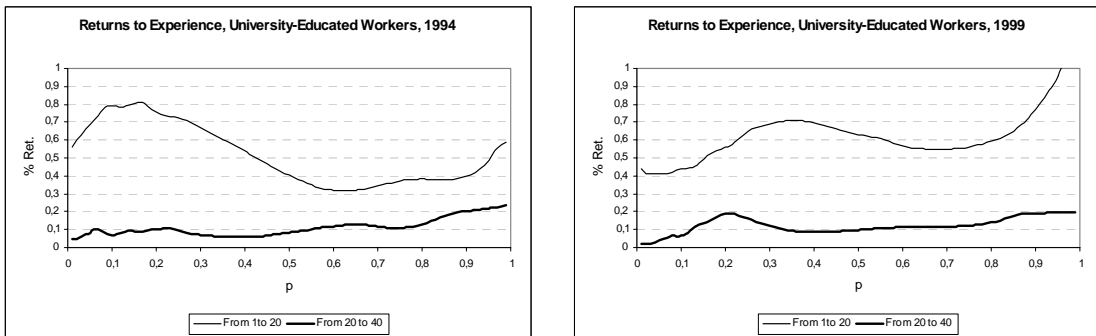
**FIGURE 13:**

**Relative Increase in Wage of the  $p$ -th Worker with Changes in Years of Experience (Workers with Education = High School)**



**FIGURE 14:**

**Relative Increase in Wage of the  $p$ -th Worker with Changes in Years of Experience (Workers with Education = University)**





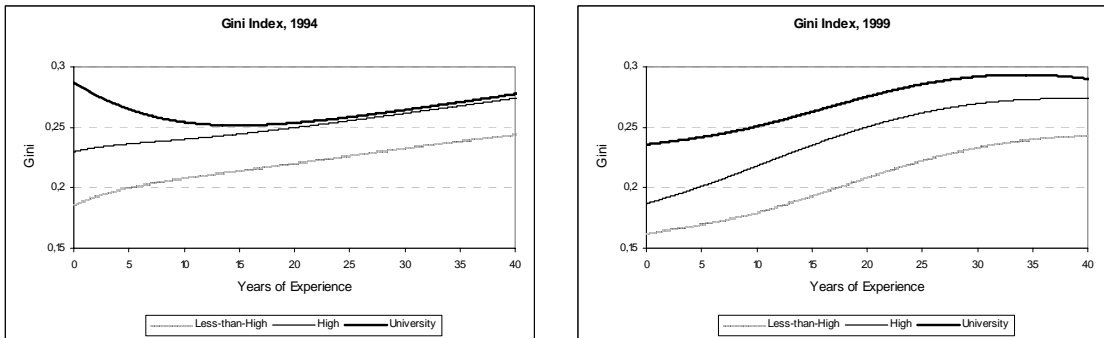
returns decrease for almost all workers at these levels. Finally, the returns to moving from 1 to 20 years of experience are less homogeneous: the better-paid the worker is, the less important role his years of experience plays, though this feature lessens in 1999.

### 3.4. Inequality Analysis

As discussed in Section 2.2, our methodology also allows us to obtain a conditional Gini index for each value of the covariates. Hence, in our case we can construct and plot  $G(x, i)$ , the conditional Gini index for real wages when workers have  $x$  years of experience and their level of education is  $i$ . In Figure 15 we plot these Gini indices as a function of  $x$  for each education level, in order to analyze how inequality changes with experience.

It is well-known that income inequality in Spain followed a decreasing trend from 1973 to the early nineties (see e.g. del R  o and Ruiz-Castillo, 2001). Comparing the plots that we obtain for 1994 and 1999 we can conclude that, in general, this trend continues, but we observe that this decrease does not affect all workers equally, since there is a major decrease in inequality at low levels of experience, whereas similar indices are obtained for workers with more than 30 years of experience. The former fact might be explained by the introduction of labor market reforms, which may have affected entrants into the labor market particularly. On the other hand, the curves in Figure 15 show that both in 1994 and 1999 wage inequality in Spain grows with education level and, in almost all cases, also with years of experience.

**FIGURE 15:**  
**Gini Indices for Wages, Conditional on Education and Experience;**  
**1994 (left) and 1999 (right)**



#### 4. CONCLUDING REMARKS

In this paper we describe a flexible estimator of conditional distributions which stems from a parametric specification of the conditional hazard function. The estimator is similar in spirit to that proposed in DGP, but whereas their starting point is a step hazard function, we propose a piecewise-linear specification. We derive how to estimate the conditional distribution and density functions with this specification. The resulting estimates continue to share the good properties of the DGP procedure: they are flexible, easy to compute and unaffected by the curse of dimensionality. The Monte Carlo experiments that we report show that our estimation procedure outperforms the one proposed in DGP: it produces smooth estimates and yields better fits in terms of mean-squared-error. Another contribution of this paper is that we obtain the conditional Lorenz curve and Gini index which are derived from the model, thus providing a valuable additional tool for analyzing inequality issues.

The application of our methodology to the analysis of wages in Spain in 1994 and 1999 also reveals some important characteristics of Spanish labor market. We find that the conditional densities of wages have very different shapes for workers with different levels of education, especially for workers with low levels of experience. The density for unexperienced workers with a university degree displays a flat shape in 1994, but in 1999 it is closer to the shape of the densities for less-skilled workers. A possible explanation for this is that the phenomenon labelled as “overeducation” (workers with a university degree who are hired for jobs which do not require such a qualification) intensified in this period; this might also explain the relative worsening of the situation of these workers that is

observed in Figure 8.

We also propose a measure of the returns to schooling based on the inverse of the conditional distribution function, and find that these returns decreased substantially between 1994 and 1999, especially for entrants into the labor market. A similar measure is used to analyze the returns to experience, and again a different pattern for workers with a university degree is observed. Finally, the inequality curves which we derive show that for workers with less-than-high school or high school level of education inequality decreased between 1994 and 1999 for entrants into the labor market, but no significant changes are detected for high-experienced workers; for workers with a university degree, inequality also decreased for entrants into the labor market, but increased for highly-experienced workers.

To further explore these results, we report in Table 3 the relative changes in employed and unemployed workers and active population in Spain at the three levels of education that we consider (data extracted from the Active Population Survey<sup>2</sup>).

Our results, together with the data in Table 3, point out that the economic structure pattern has generated shifts to the right for both the labor demand and the labor supply of university workers; but eventually labor supply has shifted further to the right than labor demand, motivating an increase in the amount of university degree holders working for lower real wages in 1999. This interpretation is also supported by the detailed descriptive figures reported in Grañeras et al. (2000), which show that individuals with a high level of education hugely increased their representation in the total active population during the nineties. As a consequence, the Spanish labor market reflects the overeducation effect men-

---

<sup>2</sup>EPA 1994 and 1999, Instituto Nacional de Estadística, Spain.

**TABLE 3:****Active Population in Spain: Relative Changes from 1994 to 1999**

---

Relative changes from 1994 to 1999			
	Active Population	Employed	Unemployed
Less-than-high-school	-22.22%	-15.45%	-45.64%
High-school	22.16%	40.56%	-25.93%
University	41.23%	47.34%	9.73%

---

tioned above. Whether overeducation is a long-term phenomenon at population level but only a short-term one at individual level, as pointed out by Rubb (2003), is an issue yet to be explored.

## REFERENCES

- Abadie, A., 1997, Changes in Spanish labor income structure during the 1980's: a quantile regression approach, *Investigaciones Económicas* 21, 253-272.
- Alba-Ramírez, A. and M.J. San Segundo, 1995, The returns to education in Spain, *Economics of Education Review* 14, 155-166.
- Andrés, L. and M. Mercader-Prats, 2001, Sobre la fiabilidad de los datos de renta en el Panel de Hogares de la Unión Europea (PHOGUE, 1999), *Estadística Española* 148, 241-280.
- Andrews, D.W.K., 1997, A conditional Kolmogorov test, *Econometrica* 65, 1097-1128.
- Bai, J., 2003, Testing parametric conditional distributions of dynamic models, *Review of Economics and Statistics* 85, 531-549.
- Buchinsky, M., 1994, Changes in the U.S. wage structure 1963-1987: Application of quantile regression, *Econometrica* 62, 405-458.
- Del Río, C. and J. Ruiz-Castillo, 2001, Intermediate inequality and welfare: the case of Spain, 1980-81 to 1990-91, *Review of Income and Wealth* 47, 221-238.
- DiNardo, J., N. Fortin, and T. Lemieux, 1996, Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach, *Econometrica* 64, 1001-1044.
- Donald, S., D. Green, D. and H. Paarsch, 2000, Differences in wage distributions between Canada and the United States: an application of a flexible estimator of

distribution functions in the presence of covariates, *Review of Economic Studies* 67, 609-633.

Fortin, N. and T. Lemieux, 1995, Rank regression, wage distributions, and the gender gap, *Journal of Human Resources* 33, 610-643.

Grañeras, M., J.L. Gordo, R. Lamelas, N. Villa and A. De Regil, 2000, *Las Desigualdades en la Educación en España II* (Area de Estudios de Investigación del Ministerio de Educación y Cultura, Madrid).

Heckman, J., and R. Robb, 1985, Using longitudinal data to estimate age, period and cohort effects in earnings equations, in: W.M. Mason and S.E. Feinberg, eds., *Cohort Analysis in Social Research*, (Springer-Verlag, New York).

Koenker, R. and K. Hallock, 2001, Quantile regression, *Journal of Economics Perspectives* 15, 143-156.

Meyer, B., 1990, Unemployment insurance and unemployment spells, *Econometrica* 58, 757-782.

Rubb, S., 2003, Overeducation: a short or long run phenomenon for individuals?, *Economics of Education Review* 22, 389-394.

Scott, D., 1979, On optimal and data-based histograms, *Biometrika* 66, 605-610.

Zheng, J.X., 2000, A consistent test of conditional parametric distributions, *Econometric Theory* 16, 667-691.