INVESTIGACIONES ECONOMICAS, vol. XXI(2), 1997, 253-272

CHANGES IN SPANISH LABOR INCOME STRUCTURE DURING THE 1980'S: A QUANTILE REGRESSION APPROACH

ALBERTO ABADIE*
Massachusetts Institute of Technology

The objective of this study is to show how labor income structure changed in Spain during the 1980's. For this purpose quantile regression techniques are applied. The advantage of using quantile regressions resides in the possibility of a more accurate description of the conditional distribution of labor income. This allows us to analyze the changes in both tails of the distribution rather than just in the mean. The results show a decrease both in returns to schooling and in within groups dispersion. (*JEL* J31, C21)

1. Introduction

Changes in wage structure have given rise to an interesting debate in recent years. Bound and Johnson (1992), Katz and Murphy (1992) and Junh, Murphy, and Pierce (1993) document an increase in wage dispersion for the United States during the 1980's. This phenomenon is often attributed to a shift in the demand of labor towards highly-skilled workers. For Europe there is no conclusive evidence of a larger dispersion in wages but of a rise in unemployment. Bertola and Ichino

This paper is a revised version of my Master's thesis, presented in partial fulfillment for the 1993-1995 Graduate Program at the Centro de Estudios Monetarios y Financieros (CEMFI). I am indebted to Manuel Arellano for his excellent supervision and to Samuel Bentolila, Juan J. Dolado and other professors and students at CEMFI for their helpful comments. I would like to thank Pablo Antolín and Javier Ruiz-Castillo for providing me with the data. I also benefited from the hospitality of the Department of Economics in Universidad del País Vasco during the Summer of 1996 and from the comments made by the editor and two anonymous referees. All remaining errors are my own.

(1995) explain these differences between the US and Europe as the result of a high volatility of labor demand in two different scenarios: a heavily regulated labor market and a flexible one.

The Spanish case is specially interesting. During the 1980's, Spain pursued the opening up of her economy, became a member of the European Community (EC) and started a process of liberalization of her labor market. The direction in which the wage distribution has been affected by these events is still a controversial point. Nevertheless, empirical work on this matter is almost non-existent. The reason is that no good microdata about wages are available. So, this study focuses on labor income structure rather than on wage structure. Statistics used for the analysis and the subsample selected for estimation are chosen in order to minimize the effects of hours differentials and employment restrictions on the results.

Following the work of Chamberlain (1994) and Buchinsky (1994, 1995), this paper applies quantile regression to the study of the structure of payments in labor markets. If we divide the workers of the studied labor market into groups defined by observable characteristics such as schooling, age, etc., traditional OLS techniques can be used to describe parsimoniously the position of the mean earner in each group of the population. However, we may also want to describe the location of those who are low earners or high earners relative to their groups of population. Quantile regression provides a suitable framework in this context. In this article, quantile regression is used to estimate quantile functions of the conditional distribution of labor income. This allows us to analyze changes in both tails of the conditional distribution rather than just in the mean.

The use of quantile regression techniques has an additional advantage: it is a robust method in the presence of one type of outliers that will be present in our sample. If the value of the dependent variable is reduced for any set of observations that fall below a particular conditional quantile, then the estimate of that conditional quantile will not be altered.[1] The importance of this feature will become clear later.

[1]The properties of sample quantiles provide some insight about this. The median (that is, the quantile 0.5) is not affected by a reduction in value of any set of observations below it. The same principle applies to any other quantile.

The rest of the paper is structured as follows. In Section 2 the econometric methods are presented. Section 3 describes the data and discusses some issues about them. Section 4 contains the results and some alternative interpretations. Section 5 concludes.

2. Econometric methods

The original motivation for quantile regression was the extension to the linear model of the notion of robust statistic used in the simple location model.[2] For the simple location model, under non-normality, the sample mean may be dominated in terms of asymptotic efficiency by other statistics such as the median, the trimmed mean or the Gastwirth statistic.[3] These statistics are robust in the sense that they are not affected by extreme observations, this feature makes them very efficient when the distribution of the data is long-tailed. Koenker and Bassett (1978) introduced quantile regression, especially median regression, as a robust method for the linear model. Formally, by using quantile regression we estimate quantile functions from the conditional distribution of interest. Some good properties of sample quantiles are preserved in quantile regression. In particular, it provides estimates that are not affected by large residuals over the quantile regression line.

The different behavior of conditional quantile functions relative to conditional mean functions was exploited soon, providing alternative uses for quantile regression. Koenker and Bassett (1982) constructed a test for heteroskedasticity based on quantile regressions. The basic idea of that test is that for the linear model, the slope coefficients of different quantile functions will be identical if the distribution of the regression error, $u = y - x'\beta$, is independent of x. Then, significant

[2]Suppose that we want to characterize the distribution of some random variable y. In the simple location model we try to estimate $\mu = E[y]$, that is the unconditional mean of y whose existence is assumed. The linear model considers a linear specification for the conditional mean of y, that is $E[y \mid x] = x'\beta$ where x is a $k \times 1$ vector of random variables and β is a $k \times 1$ vector of unknown parameters to be estimated.

[3]The α-trimmed mean is defined as the sample mean after the proportion α of larger and smaller observations has been removed. The Gastwirth statistic is defined as $0.3 \times quantile(1/3) + 0.4 \times quantile(1/2) + 0.3 \times quantile(2/3)$.

differences between the estimates of the slope coefficients will support the hypothesis of heteroskedasticity when testing for it. Powell (1984, 1986) proposed a quantile regression model for the censored case. That model is based on the simple fact that censoring does not affect those quantiles located outside the censored region while those quantiles located in that region are moved to the censoring point.

However, the capability of quantile regression to describe the entire conditional distribution has turned out to be its most characteristic application. This feature makes quantile regression especially suitable when we are interested not only in the mean but also in the tails of some conditional distribution. In this spirit, quantile regression has been recently applied to the study of changes in the conditional distribution of wages.[4] With this technique we can describe how the wages of those who are low earners or high earners within their groups of population vary with changes in the covariates that define those groups (e.g. schooling, age).

The standard quantile regression model can be described as follows. Denote by $Q_\tau(y|x)$ the τ-th quantile of the conditional distribution of y given x. Assume that $Q_\tau(y|x) = x'\beta_\tau$ and define $u_\tau \equiv y - Q_\tau(y|x) = y - x'\beta_\tau$. Our goal is to estimate the unknown parameters, β_τ, that define the conditional quantile functions.

It can be shown (Manski, 1988) that $\beta_\tau = \arg\min_\beta E\left[\rho_\tau\left(y - x'\beta\right)\right]$ where

$$\rho_\tau(\lambda) = \left(\tau \mathbf{1}(\lambda > 0) + (1-\tau)\mathbf{1}(\lambda < 0)\right)|\lambda| = \left(\tau - \mathbf{1}(\lambda < 0)\right)\lambda,$$

and $\mathbf{1}(\mathcal{A})$ denotes the indicator function of the event $\mathcal{A}$. Using the analogy principle, a quantile regression estimator is naturally defined by

$$\hat{\beta}_\tau = \arg\min_\beta \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(y_i - x_i'\beta\right),$$

where $\{(y_i, x_i')\}_{i=1}^n$ is a random sample of the studied population. Denote by $f_{u_\tau}(\cdot|x)$ the conditional density of u_τ given x and suppose that $f_{u_\tau}(0|x) = f_{u_\tau}(0)$. Under certain conditions, it can be shown

[4] See Chamberlain (1994) or Buchinsky (1994, 1995).

that $\sqrt{n}\left(\hat{\beta}_\tau - \beta_\tau\right) \xrightarrow{d} N\ (0, \Lambda_\tau)$,[5] where

$$\Lambda_\tau = \sigma_\tau^2\ \left(E\left[xx'\right]\right)^{-1} \qquad \text{and} \qquad \sigma_\tau^2 = \frac{\tau(1-\tau)}{f_{u_\tau}^2\ (0)}.$$

An alternative strategy when covariates are discrete has been proposed by Chamberlain (1994). Such strategy consists of estimating sample quantiles within each cell, or probability mass point of covariates, and imposing a parametric form via minimum-distance (MD). Let $\phi_j = \Pr\left(x = x^j\right),\ j \in \{1, .., J\}$, $Q_\tau^j = Q_\tau\left(y \mid x = x^j\right)$ and $Q_\tau = \left(Q_\tau^1, .., Q_\tau^J\right)'$. Assume that $Q_\tau = G\beta_\tau$ where $G = \left(x^1, .., x^J\right)'$. An MD estimator of β_τ is given by

$$\hat{\beta}_\tau^{MD} = \arg\min_\beta\ \left(\hat{Q}_\tau - G\beta\right)' A_n \left(\hat{Q}_\tau - G\beta\right),$$

where $\hat{Q}_\tau$ is the sample counterpart of Q_τ and A_n is a positive semidefinite matrix. Applying standard MD results,

$\sqrt{n}\left(\hat{\beta}_\tau^{MD} - \beta_\tau\right) \xrightarrow{d} N\left(0, \Lambda_\tau^{MD}\right)$

where

$\Lambda_\tau^{MD} = (G'AG)^{-1}\, G'\, A\, \Lambda_{Q_\tau}\, A\, G\, (G'AG)^{-1}$,
$A = plim\ A_n$, $\sqrt{n}\left(\hat{Q}_\tau - Q_\tau\right) \xrightarrow{d} N\ (0, \Lambda_{Q_\tau})$,
$\Lambda_{Q_\tau} = diag\left(\sigma_{\tau_1}^2/\phi_1, \ldots, \sigma_{\tau_J}^2/\phi_J\right)$
and

$$\sigma_{\tau_j}^2 = \frac{\tau\,(1-\tau)}{f_{u_\tau}^2\ (0|x = x^j)}.$$

Chamberlain (1994) found that if $Q_\tau \neq Q_\tau^* = G\beta_\tau$, where Q_τ^* is the optimal linear predictor of Q_τ, and we choose

$A_n = diag\,(n_1/n, .., n_J/n)$,

with n_j being the number of observations within each cell, then

$$\sqrt{n}\left(\hat{\beta}_\tau^{MD} - \beta_\tau\right) \xrightarrow{d} N\ \left(0, \Lambda_\tau^{MD} + \Lambda_\tau^{M}\right),$$

[5]See Koenker and Bassett (1978).

where $\Lambda_\tau^M = (G'AG)^{-1}G'A\, diag\left(\delta_{\tau 1}^2/\phi_1, .., \delta_{\tau_J}^2/\phi_J\right) AG(G'AG)^{-1}$ and $\delta_{\tau_j} = Q_\tau^j - x^{j\prime}\beta_\tau$.

Chamberlain (1994) and Buchinsky (1994, 1995) use these techniques to describe the relationship between wage distribution characteristics and variables related to schooling, experience, etc. Note that the interpretation of coefficients should not be causal but merely descriptive because of the problems the estimation of this type of effects involves.[6]

3. The data

The data used in this study come from the Spanish Family Expenditure Survey (Encuesta de Presupuestos Familiares (EPF)) for 1980/81 and 1990/91. This survey is carried out approximately every ten years and about 20,000 households are included. The EPF contains individual data of annual (net of taxes) labor earnings, age, educational level and other relevant variables. Unfortunately, the EPF does not provide wage data and information about hours is very limited.

Using labor income data could involve two types of distortions. The first one is due to hours differentials in labor supply, that is, labor income can differ between two, otherwise identical, individuals just because one of them supplies more hours than the other. The second type of distortion is related to employment experience. Labor income depends on how the worker has been affected by unemployment during the relevant year. These distortions make the interpretation of results difficult when using data on labor income, since we cannot identify whether changes in labor income are due to variations in wages, hours or in the number of days that workers are employed

[6] For example, the presence of unobservable individual skills might make us overstate the causal effect of schooling on average wages. The reason is that unobservable skills are likely to induce more schooling but also higher wages given the educational level. On the other hand, measurement error in schooling variables would lead us to understate the effect of schooling on wages. So, the overall sign of the bias is theoretically ambiguous (and highly controversial). Grilliches (1977) provides a comprehensive discussion of the problems involved in the estimation of returns to schooling. Abadie (1995) proposes an instrumental variable model for quantile regression when covariates are endogenous.

during the relevant year.

In order to minimize the effects of hours differentials and employment restrictions on the results, I selected a sample of male workers who meet the following criteria: a) main earners of the household, b) aged 18-65, c) not self-employed, d) working more than 1/3 of the usual weekly hours, e) not studying or having received transfer income, and f) non-immigrants. The resulting sample includes 9956 observations for 1980/81 and 7083 for 1990/91. In this way we extract a subsample of individuals for which realized labor income approximates potential labor income, that is the income that those individuals would have earned if they had worked the usual weekly hours without being affected by unemployment during the relevant year.[7] It is likely that, because of the low flexibility in working hours in the Spanish economy, this inclusion criteria accounts for the first mentioned problem. Nevertheless, we still face some outliers in the low tail of labor income distribution, probably due to transitory and non-subsidized unemployment periods for individuals in the sample or new entrants in the employed population. Even in the presence of these outliers, I do not adopt the usual practice of truncating the sample. If these observations actually correspond to new entrants or reentrants, and we suppose that entrants are likely to be in the low tail of the conditional distribution, quantile regression would be a robust method if the sample is not truncated. Under this assumption the outliers would represent observations in the low tail of the conditional distribution for which the value of the dependent variable has been reduced. As noted above, conditional quantiles located above the correct values of the outlying observations will not be affected. So, in the worst case, it is probable that unemployment and hours differentials have a significant influence only on low quantiles estimates. Nevertheless, if the effect of these distortions is similar in both periods, our basic findings would not be affected.

Note that any conclusion based on the results should be applied

[7]For simplicity, in what follows I will refer to the dependent variable as labor income. However, individuals in the selected subsample are less likely to have been affected by unemployment, so changes in labor income for the entire population may be very different from what is shown here due to unemployment effects. Note that the behavior of potential labor income will be very similar to that of wages.

mainly to workers that have not been affected by unemployment during the relevant year. Those who have been affected by unemployment have low probability of being in the selected sample, and their presence should have little influence on the results under the above mentioned assumptions.

4. Empirical results

The basic results of this paper are reported in Table 1. This table contains MD estimates of mean and quantile regressions of log labor income on educational dummies, age and age squared.[8] This places nominal effects on the constant and gives an approximate percent interpretation to the coefficients. *Primary* represents primary school as last grade completed, *Secondary* and *University* have the same meaning for high school and college respectively. Illiterates and individuals that did not complete primary school are left in the constant.

The sample is partitioned in 192 schooling-age cells. Then, cells with fewer than 15 observations are dropped, leaving 141 cells with 9596 observations for 1980/81 and 134 cells with 6803 observations for 1990/91. The linearity hypothesis is rejected by overidentification restriction tests. For example, the Sargan test statistic[9] takes the value of 231.51 for the mean in 1980/81, which is a very unlikely high value under the null hypothesis distribution, a $\chi^2_{[135]}$. This leads us to use the MD method proposed by Chamberlain (1994) for the misspecified case. Standard errors are given by the order statistics

[8] Most of the studies for the US use potential experience as a covariate instead of age. Potential experience is defined as min{age−years of schooling−6, age −18} and is used as a proxy for working experience. I chose age for two reasons. First, the EPF does not contain information about years of schooling; they can only be approximated using the last grade completed. The second reason is that the high incidence of unemployment in Spain makes potential experience less suitable for Spain than for the US.

[9] The Sargan test for linearity of the conditional mean was implemented in the following way. Define $m = (E[y \mid x = x^1], .., E[y \mid x = x^J])'$ and - $=$ $diag(Var(y \mid x = x^1)/\phi_1, .., Var(y \mid x = x^J)/\phi_J)$. It can be proven that under linearity $\widehat{\chi}^2 \equiv \min_\beta\{n\,(\widehat{m} - G'\beta)'\,\widehat{-}^{-1}\,(\widehat{m} - G'\beta)\} \xrightarrow{d} \chi^2_{[J-k]}$, where $\widehat{m}$ and $\widehat{-}$ are the sample counterparts of m and - . Then, we reject $H_0 : m = G'\beta$ if $\widehat{\chi}^2 >$ quantile$_{1-\alpha}(\chi^2_{[J-k]})$. Tests for linearity of conditional quantiles can be constructed in a similar way by replacing m with Q_τ and - with Λ_{Q_τ}.

Table 1

method for an approximate 0.95 confidence interval.[10] In addition to the mean and the median, quantiles 0.10, 0.25, 0.75 and 0.90 are considered.

The most interesting feature of the results is the sharp decrease in returns to schooling during the 1980's. The reduction was specially large for secondary schooling in such a way that university graduates improved their position against secondary school graduates.

In 1980/81, coefficients related to primary and secondary schooling show a decreasing pattern as we go from low quantiles to higher ones. Returns to university education have a U-shaped behavior with a higher effect on extreme quantiles. Incremental returns[11] are roughly constant across quantiles for secondary schooling and larger in high quantiles for university graduates. In 1990/91 the U-shaped pattern has generalized to all educational levels; incremental returns are increasing across quantiles for secondary schooling and constant for college graduates.

More synthetic information is reported in Table 2, which shows changes in returns to schooling during the 1980's for different age groups and for the entire sample. These changes are defined as the difference in the respective schooling coefficients between the two periods. Take the entire sample, that is the columns with entry 18-65 for the age group: for primary and secondary schooling the larger decreases take place at the middle quantiles. University graduates face the biggest reductions at the high quantiles.

In order to describe more precisely the way in which different groups of the population have been affected by those changes, I partitioned the sample into three age groups (18-35, 36-50, and 51-65) and estimated the coefficients for each subsample. In this way we can capture, to a certain extent, possible interactions between schooling and age effects. Results are shown in the correspondent columns of Table 2.

On the mean, young and elderly workers have suffered the sharpest declines in returns to schooling. However, the conditional distribution

[10] The appendix describes the order statistics method in detail.

[11] By incremental returns I just refer to the difference between the coefficient related to an educational level and the immediately lower one.

Table 2

of each group has been affected in a very different way. Young workers face the largest drops in the coefficients related to the low tail of the distribution for all schooling levels. For secondary schooling the reduction is specially large. Primary and secondary schooling returns fall significantly at all quantiles but the 0.90. University returns decline in a significant way only at the 0.10 quantile.

For the largest group, that is, middle-age workers, the most important changes in returns to schooling take place at the median but they are moderate in comparison with the other groups. Again, secondary school graduates are the most affected by those changes.

Elderly workers do not suffer significant changes in the coefficients related to the low tail of the distribution. The larger reductions are reached by secondary schooling at the middle quantiles and particularly by university schooling at the middle and the high quantiles.

To sum up, the most striking results are: (a) a great decline in the coefficients associated with the low tail of the distribution for young workers in all educational levels, and (b) a drop in the returns to university degrees at the middle and the high quantiles for elderly workers.

Our results may be influenced by composition effects. This may happen, for example, if the proportion of university graduates with less valued degrees rises in the population. Composition effects can also arise if unobservable skills within educational groups change between the two periods.

In order to minimize composition effects, I partitioned the sample into cohorts given by the 18-30, 31-50 and 51-65 age groups for 1980/81 and the respective ones 10 years later, that is the 28-40, 41-60 and 61-65 age groups for 1990/91. Attempts to use the MD method were unsuccessful, as dropping cells with fewer than 15 observations eliminated entire educational levels for certain age groups. For this reason the method proposed by Koenker and Bassett (1978) was used in estimation.

Table 3 reports the results. Returns to schooling fall significantly for the three cohorts at all educational levels for some quantiles except for university graduates belonging to the youngest cohort. Individ-

Table 3

uals belonging to the eldest cohort suffer significant reductions in secondary schooling and university returns at the 0.50 and the 0.75 quantiles. Concerning those aged 31-50 in 1980/81, the changes are moderate and take place at the middle of the distribution.

The major disadvantage of using cohorts is that, if age interacts with each educational level in a different way, changes in estimates could include increases due to experience and other contaminating effects. In any case, an identification problem arises from the identity that links age, time and cohort. This prevents us from isolating the effect of each variable in a linear model.

In spite of the descriptive aim of this study, some speculative explanations of the decline in returns to schooling can be outlined. If we think that the demand of skilled labor has risen in recent years, as happens to be commonly accepted, explanations should focus on the supply side. The rise and diversification of educational opportunities is a fact in Spain. This, along with a steady increase of income, a drop in schooling costs, and an increment in the number of cohorts affected by compulsory schooling laws, has raised the skilled labor supply. If the demand was not able to absorb the excess of skilled labor, a simple supply-demand reasoning would explain the reduction in returns to schooling. This view is supported by some results from studies based on the Spanish Labor Force Survey (EPA). During the 1980's the proportion of secondary school graduates rose dramatically in the population, the proportion of illiterates and primary school graduates fell and the proportion of university graduates rose moderately. The biggest group in the unemployed population was formed by illiterates and primary school graduates in 1979 and by secondary school graduates in 1988.[12]

The huge reductions in the returns to schooling for young workers at low quantiles may be due to the above mentioned composition effects. This group has enjoyed the greatest reduction in the costs of schooling, so it is likely to contain a bigger proportion of individuals who are less skilled or who have less valued degrees within each ed-

[12]Revenga (1991) obtains these results for 1979-88. Educational groups are not homogeneous with the respective ones of this study but differences refer to small subgroups.

ucational group. Those individuals would push downwards the low tail of the conditional distribution of labor income for young workers.

Changes in the coefficients within a cohort are more likely to reflect labor demand shifts. If this interpretation is correct, changes in returns to schooling for the youngest cohort may be caused by an increase in the demand of highly skilled labor for this cohort. The eldest cohort may have suffered a reduction in the demand of skilled labor due to depreciation effects.

Age effects are more difficult to interpret. These effects depend on two terms, the coefficients related to the linear term are positive, those related to the quadratic term are negative and much smaller in absolute value. This means that the marginal effect of age is positive for young workers and negative for elderly ones. Table 4 evaluates the joint effect of these two terms for ages 25, 40 and 55 using the coefficients of Table 1. On the whole, age effects are bigger at the extreme quantiles and they rise during the 1980's, especially for elderly workers.

TABLE 4
Age Effects

	Age					
	25		40		55	
	80/81	90/91	80/81	90/91	80/81	90/91
Mean	0.93	1.13	1.13	1.43	1.06	1.44
Quantile 0.10	1.13	1.65	1.35	2.07	1.22	2.06
Quantile 0.25	0.72	1.03	0.87	1.31	0.80	1.33
Quantile 0.50	0.86	0.92	1.04	1.15	0.99	1.16
Quantile 0.75	0.97	1.09	1.19	1.38	1.14	1.38
Quantile 0.90	1.02	1.28	1.26	1.62	1.22	1.64

Note: This table evaluates the sum of the two age terms in the labor income equation, at ages 25, 40 and 55, using the coefficients of Table 1.

Another interesting issue concerning labor income is within groups dispersion. Figure 1 graphs the interquartile range of real labor income against the median for schooling-age groups with 15 observations or more.[13] A huge decrease in within groups dispersion, measured by the interquartile range, can be observed. Because of the sample selection criteria adopted, this reduction is likely to reflect a decrease in wage dispersion in the subsample used for estimation. This may seem surprising in a period of liberalization of the labor market, as we would expect wages to reflect productivity differentials more accurately. However, there are at least three possible explanations for this reduction in within groups dispersion: a) During the 1980's many trade barriers were dismantled and formerly shielded sectors had to face up to competition. This led firms with low productivity to bankruptcy. If wage differentials among firms and sectors reflected to a certain extent productivity differentials, job losses would have affected less paid workers, reducing the dispersion in the remaining occupied population. b) The decrease in dispersion within groups could be also the result of a rise in the progressiveness of taxes due to political decisions or caused by inflation. c) A spread of the number of firms and sectors subject to collective bargaining could also explain part of these changes. Collective bargaining tends to increase rigidities and prevents individual wages from adjusting to productivity. These reasons, among others, could account for the decline in within groups dispersion. Unfortunately, the nature of the data used in this study does not allow us to assess the relative importance of each explanation.

5. Conclusions

In this paper I have applied quantile regression techniques to the study of changes in Spanish labor income structure. This analysis shows a great decline in the returns to schooling during the 1980's, mainly for young and elderly workers. However, the pattern of changes in conditional distributions differs substantially between these two groups. Changes affect the low tail of the distribution for young workers and the high tail for elderly ones. I propose a sim-

[13] The interquartile range is defined as the difference between the quantile 0.75 and the quantile 0.25. This statistic is commonly accepted as a good measure of dispersion in the presence of outliers. For 1990/91, labor income was deflated by the Private Consumption Deflator.

Figure 1

ple supply-demand interpretation to explain these changes, although composition effects are also likely to be present. Further research using data from the Labor Force Survey (EPA) should be carried out to achieve a better understanding of the results.

In addition, a reduction in within groups dispersion is observed. This is interpreted as the result of the dismantling of trade barriers, a rise in the progressiveness of taxation and a spread of collective bargaining, although other interpretations are also possible.

These results apply mainly to permanent employees and may not be extended to the entire population. The study of the changes in labor income structure for the entire population should include an analysis of the changes in the probability of being employed for each group of the population. This strategy would allow us to identify the source of variation of labor income in each case. This constitutes an interesting topic for future research.

Appendix. Estimating the asymptotic variance of sample quantiles

Following Chamberlain (1994), the order statistics method is employed to estimate the asymptotic variance of sample quantiles. This method consists of equating two confidence intervals, one given by a binomial distribution and the other given by a normal approximation.

Denote $y^j_{(r)}$ the r-th order statistic of $\{y_i \mid x_i = x^j\}$, and assume that y has finite density at Q^j_τ given $x = x^j$, then

$$Pr\left(y^j_{(b)} \leq Q^j_\tau \leq y^j_{(t)}\right) = Pr\left(b \leq v_j < t\right),$$

where $v_j \sim B(n_j, \tau)$, $B(\cdot,\cdot)$ denotes a binomial variable, and b and t are some positive integers such that $1 \leq b \leq t \leq n_j$. As $B(n_j, \tau) \xrightarrow{d} N(n_j\tau, n_j\tau(1-\tau))$ then

$$Pr\left(y^j_{(b)} \leq Q^j_\tau \leq y^j_{(t)}\right) \simeq \Phi\left(\frac{t - n_j\tau}{\sqrt{n_j\tau(1-\tau)}}\right) - \Phi\left(\frac{b - n_j\tau}{\sqrt{n_j\tau(1-\tau)}}\right).$$

On the other hand, we know that $\sqrt{n_j}\left(\hat{Q}^j_\tau - Q^j_\tau\right) \xrightarrow{d} N\left(0, \sigma^2_{\tau_j}\right)$. We can construct an approximate confidence interval, that is

$$Pr\left(\hat{Q}^j_\tau - \frac{\sigma_{\tau_j}}{\sqrt{n_j}} z_{1-\frac{\alpha}{2}} \leq Q^j_\tau \leq \hat{Q}^j_\tau + \frac{\sigma_{\tau_j}}{\sqrt{n_j}} z_{1-\frac{\alpha}{2}}\right) \simeq 1 - \alpha,$$

where $z_{1-\frac{\alpha}{2}}$ is the $\left(1-\frac{\alpha}{2}\right)$-th quantile of the standard normal distribution. Then, we choose t_j and b_j such that $\left[y^j_{(b_j)}, y^j_{(t_j)}\right]$ defines an approximate $(1-\alpha)$ confidence interval around Q^j_τ, that is

$$\Phi\left(\frac{t_j - n_j\tau}{\sqrt{n_j\tau\,(1-\tau)}}\right) - \Phi\left(\frac{b_j - n_j\tau}{\sqrt{n_j\tau\,(1-\tau)}}\right) \simeq 1-\alpha$$

and

$$\frac{t_j - n_j\tau}{\sqrt{n_j\tau\,(1-\tau)}} \simeq -\frac{b_j - n_j\tau}{\sqrt{n_j\tau\,(1-\tau)}}\ .$$

This implies that

$$t_j = \min\left\{n_j, round\left(n_j\tau + z_{1-\frac{\alpha}{2}}\sqrt{n_j\tau\,(1-\tau)}\right)\right\}$$

and

$$b_j = \max\left\{1, round\left(n_j\tau - z_{1-\frac{\alpha}{2}}\sqrt{n_j\tau\,(1-\tau)}\right)\right\}.$$

Where $round(\lambda)$ represents λ rounded to the nearest integer. Then, $\hat{\sigma}_{\tau_j}$ is defined to match both confidence intervals, that is

$$\hat{\sigma}_{\tau_j} = \frac{\sqrt{n_j}\left(y^j_{(t_j)} - y^j_{(b_j)}\right)}{2\,z_{1-\frac{\alpha}{2}}}.$$

References

Abadie, A. (1995): "Changes in Spanish labor income structure during the 1980's: A quantile regression approach", Centro de Estudios Monetarios y Financieros. Madrid. Working paper 9521.

Bertola G. and A. Ichino (1995): "Wage inequality and unemployment: US vs Europe", National Bureau of Economic Research. Working paper 1186.

Bound, J. and G. Johnson (1992): "Changes in the structure of wages during the 1980's: An evaluation of alternative explanations", *American Economic Review* 82 , pp. 371-392.

Buchinsky, M. (1994): "Changes in the US wage structure 1963-1987: Application of quantile regression", *Econometrica* 62, pp. 405-458.

Buchinsky, M. (1995): "Quantile regression, Box-Cox transformation model, and the US wage structure, 1963-1987", *Journal of Econometrics* 65, pp. 109-154.

Chamberlain, G. (1994): "Quantile regression, censoring, and the structure of wages ", *Advances in Econometrics, Sixth World Congress* (Volume 1); C. A. Sims (ed.), Cambridge University Press.

Grilliches, Z. (1977): "Estimating the returns to schooling: Some econometric problems", *Econometrica* 45, pp. 1-22.

Juhn, C., K. Murphy and B. Pierce (1993): "Wage inequality and the rise in return to skill", *Journal of Political Economy* 101, pp. 410-422.

Katz, L. and K. Murphy (1992): "Changes in the relative wages, 1963-87: Supply and demand factors", *The Quarterly Journal of Economics* 107, pp. 35-78.

Koenker, R. and G. Bassett (1978): "Regression quantiles", *Econometrica* 46 pp. 33-50.

Koenker, R. and G. Bassett (1982): "Robust test for heteroscedasticity based on regression quantiles", *Econometrica* 50, pp. 43-61.

Manski, C. F. (1988): "Analog estimation methods in econometrics", New York. Chapman and Hall.

Powell, J. (1984): "Least absolute deviation estimation for the censored regression model", *Journal of Econometrics* 25, pp. 303-325.

Powell, J. (1986): "Censored regression quantiles", *Journal of Econometrics* 32, pp. 143-155.

Revenga, A. (1991): "La liberalización económica y la distribución de la renta: La experiencia española", *Moneda y Crédito* 193, pp. 179-223.

Resumen

El objetivo de este trabajo es identificar los cambios habidos en la estructura de la renta laboral en España durante los años ochenta. Con este propósito se utilizan regresiones de cuantiles que nos permiten describir detalladamente la distribución condicional de la renta laboral. De esta forma, podemos estudiar los cambios que se han dado no solo en la media, sino también en cada una de las colas de dicha distribución. Los resultados muestran una caída en los rendimientos de la educación y en la dispersión intra-grupos.

TABLE 1
Labor Income Equation, Minimum Distance Estimates
(Dependent variable: Log Labor income)

	Mean		Quantile 0.10		Quantile 0.25		Quantile 0.50		Quantile 0.75		Quantile 0.90	
variable	1980/81	1990/91	1980/81	1990/91	1980/81	1990/91	1980/81	1990/91	1980/81	1990/91	1980/81	1990/91
constant	11.824487	13.337107	11.102254	12.619350	11.866590	13.208006	11.929676	13.491900	12.021380	13.559676	12.151501	13.636262
	(162.74)	(378.15)	(45.12)	(89.23)	(105.31)	(181.65)	(143.67)	(426.58)	(138.18)	(376.24)	(101.51)	(253.59)
primary	0.322861	0.235247	0.409300	0.395338	0.311406	0.237873	0.304592	0.170365	0.292063	0.216726	0.288830	0.239118
	(28.86)	(13.09)	(10.22)	(4.44)	(18.12)	(7.63)	(24.69)	(11.87)	(21.55)	(12.27)	(16.28)	(6.25)
secondary	0.659408	0.509212	0.729946	0.608590	0.626788	0.494346	0.606546	0.440548	0.593887	0.478507	0.606238	0.531341
	(38.74)	(24.47)	(16.20)	(5.79)	(21.29)	(14.39)	(30.14)	(25.67)	(24.07)	(21.29)	(19.78)	(11.35)
university	0.875662	0.799307	0.865537	0.909160	0.804446	0.762022	0.809177	0.721597	0.842436	0.751057	0.942623	0.807849
	(47.35)	(37.92)	(16.34)	(9.53)	(27.32)	(20.05)	(34.61)	(41.11)	(30.51)	(29.26)	(29.86)	(17.20)
age	0.052158	0.060821	0.064267	0.089706	0.041066	0.055613	0.047871	0.049636	0.054048	0.059108	0.056216	0.069362
	(15.14)	(14.55)	(5.46)	(5.66)	(7.84)	(6.65)	(12.04)	(13.15)	(13.22)	(13.10)	(9.51)	(10.09)
age squared	-0.000596	-0.000630	-0.000765	-0.000951	-0.000483	-0.000572	-0.000545	-0.000520	-0.000604	-0.000617	-0.000617	-0.000720
	(-14.87)	(-13.37)	(-5.44)	(-5.27)	(-7.92)	(-6.20)	(-11.74)	(-12.31)	(-12.77)	(-11.84)	(-8.75)	(-9.12)

Note: t-ratios in parentheses.